Title

Information Theoretic and Statistical Models for Spatial Transportation Networks: Total Mixing Entropy on Optimal Fluid Flow Networks and Time Dependent Stochastic Block Models

Author

Mentus, Cassidy

Publication Date

2019

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA
Los Angeles

Information Theoretic and Statistical Models for Spatial Transportation Networks: Total Mixing Entropy on Optimal Fluid Flow Networks and Time Dependent Stochastic Block Models

A dissertation submitted in partial satisfaction
of the requirements for the degree
Doctor of Philosophy in Mathematics

by

Cassidy Mentus

2019

ABSTRACT OF THE DISSERTATION

Information Theoretic and Statistical Models for Spatial Transportation Networks: Total Mixing Entropy on Optimal Fluid Flow Networks and Time Dependent Stochastic Block Models

by

Cassidy Mentus
Doctor of Philosophy in Mathematics
University of California, Los Angeles, 2019
Professor Marcus Roper, Chair

This thesis contains two studies about models on organized spatial transport networks. The first introduces a new objective function aimed at understanding the ability that networked organisms such as fungi and slime molds to mix and efficiently disperse nuclei and molecular cues via advective currents. The second develops a novel type of stochastic block model to multilayer networks expressing model human transportation. It is a statistically based model that aims to classify parts of the network based on the function they serve for commuters. Our specific application is to bicycle share networks in different urban communities. Although these models are applied to disparate subjects they are connected from a mathematical point of view and illuminate a central theme in general tranpsport networks in two contrasting lights.

The first part of the thesis is inspired by the work of Murray [Mur26], who hypothesized that the geometries of blood flow networks are optimized to minimize the friction of flows through the network for a given total investment in the material that makes up the network. In the spirit of Murray, we hypothesize that biological networks, maximize their performance of objectives that are beneficial to the organism's existence while respecting constraints. We

are inspired by experimental observations that show that the filamentous fungus *Neurospora crassa* mixes nuclei and the slime mold *Physarum polycephalum* mixes signals it receives from its environment on the distribution of food sources [AAP17]. We use the concept of information entropy to describe how advection currents within the network carry information. We construct a probability space for the signals passing through the network and write down a novel objective function, called the total negative mixing entropy, representing each node receiving the most mixed collection of signals. Put another way, to maximize its ability to adapt to stimuli, each node receives the most even distribution of signals from the other nodes within the network.

We then define optimal networks to be ones minimizing a cost function that is the sum of the total negative mixing entropy and of fluid dissipation. A constraint assuming a fixed amount of energy used for network upkeep is assumed. Using original optimization techniques, that we describe in this paper, we numerically calculate optimal networks under different assumptions on the driving force of the flow, the underlying topology and the total material cost function. From our numerical results we identify highlight results about the structure of optimal networks, which we are then able to prove rigorously. The proofs involve constructions and computations that illluminate how energy efficient fluid transport flows are connected to mix signals and the effects that Murray's law has on features such as whether the networks posess loops.

In the second part of the thesis, we define two new types of time-dependent stochastic block models for Bicycle-Sharing Networks. The model assumes that network can be modeled by a random process based on partitioning the possible origins and destinations into blocks. The blocks in our model express the roles that the stations play in relation to the entire network and trips are assumed to be generated by a mixture of time dependent commute patterns occuring between the blocks. The only parameter of our model that is chosen by the practitioner is the number of different blocks K. The block to block commute patterns are represented in a $K \times K \times 24$ array, and the commute patterns are not assumed to be

equal if the order of the pair of blocks is changed to take into account direction of flows. It can be viewed as a degree corrected model in that there are multiplicative terms for each station representing their importance within each block. The commute patterns and degree-correction terms are inferred parameters, optimized using gradient descent. We derive both a continuous and discrete version of this model.

We apply our models to Los Angeles, Manhattan, New York and San Francisco bike-share communities. The results reveal crisp divisions of home and work communities as defined by the preference of commuters to use bikes if we use two blocks. The models also reveal other relevant functional regions, such as parks, leisure commutes, and broad communities representing micro-cosms where riders mostly do not exit. With increasing blocks reveals more roles detecting new functional roles as well as refining roles with less blocks, such as breaking a geographical community into its home-work commute roles. How to choose the number of blocks is touched on, although we do not reach a definitive result with regards to that.

The dissertation of Cassidy Mentus is approved.

Christoper Anderson

Mason Porter

Van Savage

Marcus Roper, Committee Chair

University of California, Los Angeles

2019

To mom and dad

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGMENTS

I would like to thank my graduate advisor, Marcus Roper, for mentoring me as a researcher and academician. With his support and collaboration, I have completed a thesis that I am very proud of. I am very thankful to Jane Carlen, Jaume de Dios Pont, Shyr-Shea Chang, Stephanie Wang and Mason A. Porter, with whom I co-authored the paper that is the second part of my thesis: Role Detection in Bicycle-Sharing Networks Using Multilayer Stochastic Block Models. I consider the education that I have received from UCLA in both pure and applied mathematics to be invaluable, and so I am very thankful to all of my professors. I have had the pleasure of knowing and working with friends throughout my graduate career at UCLA. The many interesting, humorous and encouraging conversations are among my best of memories. I am especially thankful to Lanya Ali Abdulkarem for being a very close and supportive friend over these years. I am extremely grateful to my parents for their endless love, support and encouragement. I am thankful for NSF grant DMS-1351860 for funding my graduate studies and research.

VITA

2012 Sc.B Mathematics, Brown University

2014 M.S. (Mathematics), UCLA, Los Angeles, California.

CHAPTER 1

Introduction

1.1 Evolutionary and Mathematical Biological Background

The evolutionary forces that the adaptation of organisms to their niches have been likened to the action of an optimization algorithm, searching over a landscape of possible phenotypes in search of optimal traits. The analogy is made explicit by optimization methods such as the genetic algorithm [Hol92], which has been successfully applied to a wide variety of optimization algorithms. The genetic algorithm finds optima by repeatedly adding (mutations) to the proposed optima of an objective function and subjecting them to recombination and artificially selection. Yet does evolution really locate optimal traits? In a widely cited perspective piece [Jac77], Francois Jacob argued that evolution is a tinkerer rather than an engineer -- that without a blueprint of the landscape, evolution could produce only literative changes in the traits of an organism, and that most of these changes will be neutral or even deleterious to its fitness.

How much of an organism's traits can be attributed to the optimization of its fitness? Charles Murray was the first scientist to apply ideas of optimization to the stsudy of the dissipation and geometry of naturally occuring biological vascular networks and most specifically of blood flow networks [Mur26]. He posited that a fluid transport network must perform its role of delivering nutrients or oxygen while taking the least amount of energy to maintain the network. Murray assumed that the energy required to maintain a network was proportional to the volume of the network. The radii of vessels within the network are therefore controlled by tradeoffs between minimizing dissipation costs (which favor large

vessels) and maintenance costs (which favor small vessels). This optimization principle has been applied to estimate how the cost of blood flow scale with organism size [WBE97] and to model the geometry of vessels [She81].

Although there is some data to support the idea that among blood vessel networks, some of the larger vessels specifcically, obey Murray's law, or that they are geometrically organized to minimize dissipation [Zam77, ZWL83, Luz73]. In [Dur07] it was proven that networks with constant boundary flows must be trees. This has brought up whether certain vascular biological structures such as plant leaves are truly minimally dissipative [KSM10] or whether constant flow boundary conditions truly valid [BM07, Bar14, CR18, Cor10, KSM10] since many plant leaves are formed of dense hierarchical loops [KM12]. Both testing other boundary conditions such as fluctuating flow [KSM10, Cor10] and other functions than dissipation such as including the possibility of damage occuring in the network, and maximizing the network's robustness to that damage [KSM10] yield networks with loops. Still, in all of these studies, minimizing dissipation was one of many objectives of the optimized networks.

In our paper we study fluid transport networks optimized for multiple objectives which stem from the assumption that they are beneficial for the organisms survival and reproduction. Our inspiration is drawn from network forming organisms with indetermine sizes and geometries, including fungi and slime molds. Based on the observation of mixing of nuclei within mycelial fluid transport networks of the species *Neurospora crassa* [RSH13] and the transference of information in the form of molecules and particles in *Plasmodium polycephalum*, we construct a new type of entropy on flows.

As a simplified assumption, there is only one flow, which we may compute given boundary conditions on the flow and the radii and lengths of conduits in the network. The fluid networks are represented as graphs we address the movement of signals between nodes (also reffered to as spatial locations). We also assume that new signals may potentially occur at any node in the flow network and that networks are organized to ensure every location

receives a diverse distribution of all upstream signals. For example, in a fungal mycelium network, nuclei flow freely by advection caused by osmotically maintained pressure gradients that drive the nuclei from the interior of the network (where nuclei are produced by division) to the growing periphery of the network, where new network edges (hyphae) are continually created by a combination of hyphal tip extension and branching [RS19, Lew11]. Two types of signals may need to be dispersed within the network: genetically diverse nucleotypes introduced within the network by mutations, fusions with other networks, or translocating from other parts of the same organism, and cues from the environment. To disperse the effects (both deleterioius and advantageous) of new nucleotypes, it may be useful for every position in the network to sample a mixed distribution of nucleotypes arriving from different positions up-stream to it.

Fungi also have distinctive life histories, in that any nuclei within the network can be packaged into spores, called conidia, that can then be dispersed off to make progeny networks. The spores can be made anywhere within the network, and it is theorized [RD19] that they should represent as much of the diversity of the parent fungus as possible. It is also theorized that fungi and slime mold react to nearby stimuli such as a nearby food source by secreting signalling molecules into their fluid networks. The information from the stimuli is transferred by flows and there is evidence [Puk11] that the signals need to be dispersed as widely as possible throughout the network. In other words, the flow should transfer the most mixed bundle of signals to every down-stream portion of the network. We discuss in further depth the work on *Neurospora crassa* and *Physarum polycephalum* that have inspired these concepts.

1.2 Mixing flows in *Neurospora crassa* and *Physarum polycephalum*

Networked fungi and slime mold share structural and functional similarities in that the entire organism comprises a fluid-filled vascular network. The organism forages for food continuous

growth at the periphery of the network, which is fed by fluid flows from the interior of the network. The fitness of the network is likely to depend upon the efficiency with which it uses its energy resources. It is assumed a-priori that network organisms are moulded through evolutionary pressures, and energy efficiency leads to a higher rate of survival of an individual to pass on genes (i.e. higher fitness), and there is experimental evidence that *P. polycephalum* networks organize to keep energy loss as low as possible[NKN04, THA00a]. Although, in the case of slime-moldes, there is evidence that this is the case due to certain sufficiently pruned experimental examples demonstrating Murrays law on parent and daughter vessel radii.

In [RSH13] Roper et al argue using experimental evidence and mathematical models that *N. crassa* mycelium are in part optimized so that nuclei generated within the network and that are flowing to the growing periphery of the network have the maximum probability of being delivered to different hyphal tips. This is conceptually equivalent to the contents of the cytoplasm at different up stream locations being mixed by advection caused by the flow. The experiment is done keeping track of only two variants of nuclei, which are originally not well mixed, but as the fungus grow, and the different nucleotypes travel up stream, they become mixed. Fungi with variants of nuclei with green flourescent and red fluourescent tags are bred. Red fluorescent labeled nuclei are introduced into wild type and a mutant, referred to as soft, that does not posess the capability for hyphal fusion. In both type of organisms, the conidia are observed to have poportions of red and green nuclei whcih a more uniform over the set of conidia than would be predicted by models which do not include mixing. Mixing is quantified via heterozygosity– the probability that two randomly chosen nuclei have different nucleotypes (i.e. one is red fluorescent and the other is green fluorescent). The mixing is roughly expressed as the extent to which conidia posess sample distributions of nuclei which are similar to the total population distribution of types. It is found that the wild type exceeds the mixing which would be the product of random flows, and in the tree-like case, mixing is greater than predictions using a biologically-feasible random branching model). In fact, the mixing observed in the tree-like mutant nearly matches the branching

pattern which optimizes heterozygosity in hyphal tips.

In [RSH13] optimization of mixing is accompanied by a strict hierarchy of branch orders as one passes from the largest hyphae feeding the tips to the tips themselves. This pattern of branching is not consistent with the network minimizing dissipation. Therefore, it can be concluded that dissipation is not the only evolutionary pressure that shapes the Neurospora mycelial network. Indeed, as noted in [RET11], previous studies suggest the ability of fungi to marshal internal genetic diversity from the multiple nucleotypes that may be present within the organisms is part of their extraordinary ecological success [Jin52, Cat96, RSH13]. [RSH13] supports this story by relating these data to direct measurements of the flow networks that provide this diversity.

The second network organism which calls us to this research topic is *P. polycephalum* slime mold. Slime molds are single celled amoeboid network organism. Like fungi, this cell can contain many nuclei in a single connected cytoplasm. It is able to move and change its morphology to react to stimuli to grow toward nutrrient sources and away from light. Remarkably, despite having no central system of organization, slime mold growth is able to solve complex network-making problems: Slime molds have been shown to be able to find the shortest path in a maze [THA00b, NYT01], find efficient networks connecting food sources [NKN04, TTS10], and anticipate time-periodic temperature changes[NYU00]. These abilities have been applied in surprising ways such as placing food sources in the formation of a map of train stations in a region in Japan containing Tokyo to design an efficient transportation network [TTS10].

A slime mold is able to store information about stimuli and transfer it to different parts of its network. It has been suggested that information is held and transfered through the fluid in the network [AAP17]. A possible mechanism for the transfer of information is advection or morphogens via the flows in the cytoplasm . Cytoplasmic flows are caused by waves of contraction that occur on the order of 100s and can have a wave front that cuts across the entire network [AAP13]. Advection of information caused by the contraction waves has been

modeled in a similar way to peristalsis in a tube: the contraction moves as a front pushing the the fluid in one direction and then, since the network is closed, it flows backwards towards the original position. Alim et al use physical models for dispersion of molecules and experimental measurements of the contraction pattern to describe the passage of information and the feedback of this information upon the flows within the network. For such a network to adapt to signals initiated and propagated away from many different locations we again expect the mixing entropy of the flow to be an important network constraint.

1.3 Numerical Optimization of Biological Fluid Transport Networks

A growing body of research tackles the question of how biological transportation networks should be organized to optimize one or more target functions, that are thought to represent quantities important to real organisms. Biological networks are believed have evolved to minimize. As mentioned above the networks connecting mixed sources and sinks while optimizing dissipation are trees [Dur07]. As pointed out in [KSM10, Cor10, KM12] biological fluid transport networks are rarely perfectly tree-like, instead they have loops. The main example used in these sources is dicotyledon leaf venation, but other two dimensional biological networks contain characteristic loops such as retinal tissue and, the source of inspiration for our model, *Neurospora crassa* mycelium. In [KSM10] two new models are introduced: They allow the computation of dissipation of networks with changing sinks and broken bonds in the transport network, to take into account controlled changes in sink pressure by the plant (referred to as stomatic patchiness) and breaking of the network, due to loss of individual conduits due to herbivory or embolism. In [Cor10] uncorrelated changing fluctuating source magnitudes are considered and Cai and Hu consider networks with adaptive vessel diameter in response to fluctuating sinks [HC13]. All of these modifications of the network target function cause optimal networks to form loops [KSM10, Cor10, HC13].

Another line of study emphasizing the divergence of experimentally mapped transportation

networks from minimal dissipation models is that of S.S. Chang and M. Roper on blood flow in animals, in particular the zebra fish. Not only does the embryonic zebrafish contain many loops, but a tapering of vessel conductances from the tail of the fish to its head visibly leads to an eleven fold increase in dissipation compared to a fish in which conductances do not taper. In other words, even if loops are assumed, the vessels show no evidence of minimizing dissipation. In [CR18] they consider general networks where not only are boundary flows specified (they refer to as Neumann conditions), but also some nodes have fixed pressure, to more realistically model zebrafish blood flows and animal blood flows, noting that extra constraints on pressures lead to minimally dissipative networks containing loops. They present results about the structure of minimal dissipation networks under this general set of boundary conditions. In [CTB17] they directly analyze the structure of the zebrafish trunk network, and show that the occlusion of small blood vessels causes feed-back effects which prevent short circuiting of the vascular network, but also contribute to the uniformity of transport flows. Uniform flows are likely necessary to ensure uniform oxygen perfusion throughout the neighboring tissue. Similarly to our work, they consider optimal networks obeying a building cost constraint inspired by Murray's law because it describes the power for upkeep of the network as they optimize for uniformity of flow in [CR17, CR19].

We cite the paper [CR17] that introduces the efficient Lagrange multiplier based method for computing the gradient of our objective function, but they also consider constraints combining dissipation and upkeep of the network , as this is inline with Murray's formulation for power consumption [Mur26] of biological fluid transport networks. On the other hand, we build a minimization procedure that searches through many possible local optima by applying a gradient based method in tandem with a topology changing step. We draw inspiration for this method from the convolution and simulated annealing method described in the supplementary information of [KSM10], but we find that directly entering domains of conductance networks with known flow topologies is more efficient than a random search over perturbations of the conductances.

1.4 Contributions of this thesis

We develope a new function on fluid flows using the concept of information entropy to describe how advective currents within biological transportation networks carry information. This is the total negative mixing entropy NME, and it measures the extent to which each node receives the most mixed collection of signals. Put another way, to maximize its ability to adapt to stimuli, each node must receive signals stemming from responses from a collection of spatial regions assuming the least amount of prior information. We also define the negative total sending entropy, the entropy of the distribution of receivers that a signal from each node can reach. We prove that these quantities are related by reversing the signs of the boundary flows.

We then define optimal networks to be ones minimizing a cost function that is the sum of the total negative mixing entropy and dissipation terms, NME$+c\times$dissipaion for positive constants c. Using original optimization techniques, that we describe in this paper, we simulate optimal networks under different assumptions on the driving force of the flow, the underlying topology and the total material cost function. We analyize the results from the different perspectives such as their overall size as well as the presence and number of loops.

The results produce smany interesting trends that we record. Out of these we select several to expand upon with in-depth mathematically rigorous analysis. We develop precise statements that capture observations on the numerical results and rigorously prove them. The theorems can be divided into two categories: statesments about the mixing-dissipation cost function and its graph and the structure of networks optimizing mixing. The theorems we prove about the mixing dissipation cost function are a basic approximation for the graphs and the nature of the optima as we modify our material cost constraint. In second category of theorems we prove that on a network with N nodes, the optimum of the NME is a path graph. An observation from our numerical results is that often the optima are path networks with a loopy subnetwork occuring near to the source. We state this theorem precisely and

prove it as well.

CHAPTER 2

Definitions and Mathematical Background

2.1 Mixing entropy on flow networks

Consider a flow on network with N nodes enumerated $1, 2, \ldots, N$. The flow on the edge (i, j) that joins two of these nodes is denoted by q_{ij}. We define two types of information entropy on these flows. The first is a measure of the accumulation of signals at every node in the network and the second represents the dispersal of signals throughout the network. The first type of entropy, we identify as the total mixing entropy and the second type as the total sender entropy. The terminology is chosen to represent the entropy of the mixing of signals arriving at any node and the entropy of the possible destinations that a signal starting at a node can be sent to. As we will see in this section, these entropies are adjoint functions: the total mixing entropy of a network is the total sender entropy of the same network except with the sources and sinks of the network swapped, so that the flows on each edge are perfectly reversed. We now describe the intuitive framework from which these entropies are derived.

We model particles being pushed by the fluid flow network by a Markov chain model where as the particle reaches a node it the probability it will flow down a given outgoing edge is proportional to the strength of the flow along that edge. That is, if the t-th node visited by the particle is x_t, then the particle performs a random walk on the network, in which the the conditional probabilities or Markov kernel are given by

$$T_{ij} \quad = \quad P(x_{t+1} = j | x_t = i) = \frac{q_{ij}}{\sum_{k \in n(i):\, q_{ik} > 0} q_{ik} + |Q_i| \mathbf{1}_{Q_i < 0}}.$$

Hence the flow of random walkers from i to j is simply proportional to the total flow along the edge (i,j). If $Q_i < 0$ then $\sum_j T_{ij} < 1$ but $\sum_j T_{ij} + \frac{|Q_i|}{\sum_{k\in n(i):\, q_{ik}>0} q_{ik} + |Q_i|\mathbf{1}_{Q_i<0}} = 1$. This is because the outflow at node i models flow of random walkers exiting the network at that node, and we define a state in the Markov chain representing a particle leaving the network. We call this state the "exit" and we say $P(x_{t+1} = \text{exit} | x_t = i) = \frac{|Q_i|\mathbf{1}_{Q_i<0}}{\sum_{k\in n(i):\, q_{ik}>0} q_{ik} + |Q_i|\mathbf{1}_{Q_i<0}}$.

Our concept of mixing is inspired by previous studies of mixing in fluid-carrying biological networks, such as the filamentous fungus *Neurospora crassa.* In [RSH13] experimental results were compared against mathematical models to show that flows guided through the network by a gentle pressure gradient are capable of mixing different nucleotypes present within the organism. Entropy is a commonly applied measure of how uniformly mixed a distribution is, which is why we immediately chose it to describe the mixing hyphal flows. We assume that an event that introduces a new important nucleotype can happen at any place in the network. A new nucleotype can be formed by mutation or hyphae fusing with another network which posesses a different genotype. Alternatively we may consider the network that we are modeling as representing some fraction of the total network, which may consist of many sources and sinks -- too many to be modeled computationally. Diverse nuclei can be introduced into the modeled fraction of the network by flowing into from the larger fungal network. Our model considers only localized flow patterns within one part of the network, and we consider nuclei flowing in to this part of the network from its other parts. Inflows of new nuclei are allowed at any node within the network. For simplicity, we do not associate nuclei introduced in this way with physical inflows. Although substantial flows of nuclei would carry cytoplasm as well as nuclei and would have to be identified as inflows, nuclei can also be moved around within the network by molecular motors. Nuclei entering the network this way are not carried with cytoplasm and do not need to be modeled by adding inflows.

Let us represent a location in an *N. crassa* mycelium by x. Assume that a nucleus with a new nucleotype or a cytoplasmic cue is introduced at this point. By our model, the nucleus

or cue takes a uni-directional random walk through the network, down the pressure gradient, visiting a random path of locations. In this paper we refer to the nucleus and its influence on a location in the network as a signal. Since the nucleus or cytoplasmic cue was introduced at point x we say that it is the signal from x. If it arrives at another point y we refer to it as the signal from x arriving at y.

By this measure, each site in the network can send signals to other sites in the network, while any point in the network will be receiving signals from other sites within the network. Since we cannot tell ahead of time which nodes will provide the useful signals, we assume that it is best if every node has the most uniform sample of signals possible within biological and physical constraints. Similarly, although in [RSH13] genetic diversity was considered only at growing hyphal tips, we can not tell ahead of time which sites need to have the largest diversity. We therefore calculate the diversity and every node within the network. Networks also operate under additional constraints: fluid transported within the network dissipates energy due to friction, while the number and sizes of edges within the network reflect the amount of energy that must be expended to build, and then to maintain the network. Metabolic investment in transport or in network building creates tradeoffs with growth or reproduction, and network building organisms have been posited to build networks that optimize dissipation for a given material cost [TTS10, BHD07]. In our investigation, we therefore consider the cost of dissipation and the amount of material in the network in addition to the total mixing within the network.

Since we express the probability distribution of paths of signals as a Markov chain, we are able to compute the probability

$$P_{ij} \quad = \quad P(x_t = j \text{ for some } t \geq 0 | x_0 = i).$$

For node i let f_i be the total flow through node i, i.e. $f_i = \sum_{j \in n(i): q_{ij} > 0} q_{ij} + |Q_i| \mathbf{1}_{Q_i > 0}$. We use this to define a measure on the flows by defining the amount of fluid flowing from i to j

to be $\tilde{q}_{ij} = P_{ij} f_i$. We refer to this as the flow from i to j. Now consider, for node i, the flows entering node i originating from upstream nodes j, $\tilde{q}_{ij}$. We now make the assumption that new signals arise at a node j in proportion to the amount of flow through j . An physical assumption that would imply this is that the comonents of the signal are concentrated uniformly throughout the cytoplasm. A node will also originate its own signal in proportion to the amount of fluid flowing through it. It then follows that the relative proportions of signals from the various upstream nodes j are the same as the relative proportions of $\tilde{q}_{ji}$. Now we are able to define the mixing entropy at node i. In matrix form the total set of flows within the network is written as:

$$\tilde{\boldsymbol{q}} = \begin{bmatrix} \tilde{q}_{11} & \tilde{q}_{12} & \cdots & \tilde{q}_{1N} \\ \tilde{q}_{11} & \tilde{q}_{22} & \cdots & \tilde{q}_{2N} \\ \vdots & \vdots & \ddots & \vdots \\ \tilde{q}_{1N} & \tilde{q}_{2N} & \cdots & \tilde{q}_{NN} \end{bmatrix} \tag{2.1.1}$$

We model the diversity of signals that are received at each node via the entropy of the distribution of signals arriving at i. To write the entropy we must we must normalize the quantities $\tilde{q}_{ji}$ via division by $\sum_{j:\tilde{q}_{ji}>0} \tilde{q}_{ji}$. We define the probability distribution on up-stream nodes of i

$$\mathcal{P}_i(j) \quad = \quad \frac{\tilde{q}_{ji}}{\sum_{j:\tilde{q}_{ji}>0} \tilde{q}_{ji}}.$$

In matrix form this is written

$$\begin{bmatrix} \mathcal{P}_1(1) & \mathcal{P}_2(1) & \cdots & \mathcal{P}_N(1) \\ \mathcal{P}_1(2) & \mathcal{P}_2(2) & \cdots & \mathcal{P}_N(2) \\ \vdots & \vdots & \ddots & \vdots \\ \mathcal{P}_1(N) & \mathcal{P}_2(N) & \cdots & \mathcal{P}_1(N) \end{bmatrix}$$

and is obtained by taking the matrix in equation 2.1.1, dividing every entry by the sum of the entries in the same row. Then we may define the **local mixing entropy at node** i as $H(\mathcal{P}_i) = -\sum_j \mathcal{P}_i(j) \log(\mathcal{P}_i(j)) = -\sum_{j:\tilde{q}_{ji}>0} \frac{\tilde{q}_{ji}}{\sum_{j:\tilde{q}_{ki}>0} \tilde{q}_{ki}} \log\left(\frac{\tilde{q}_{ji}}{\sum_{j:\tilde{q}_{ki}>0} \tilde{q}_{ki}}\right)$ where H is the Shannon information entropy applied to finite probability distributions. Since we are interested in the tradeoffs between mixing and dissipation due to friction within the network (a quantity that networks may minimize), we cast both optimization of mixing of dissipation as minimization problems, namely we consider the negative of the local mixing entropy, the **negative local mixing entropy at node** i

$$\mathrm{NME}_i = \sum_{j:\tilde{q}_{ji}>0} \frac{\tilde{q}_{ji}}{\sum_{j:\tilde{q}_{ki}>0} \tilde{q}_{ki}} \log\left(\frac{\tilde{q}_{ji}}{\sum_{j:\tilde{q}_{ki}>0} \tilde{q}_{ki}}\right).$$

We consider the total flow through i as a measure of the "importance" of the node. In our model, the diversity of signals is more important at high traffic nodes than at low traffic nodes. This principle is useful mathematically, since it ensures that rearrangements of very low conductance edges don't greatly affect the overall mixing associated with a network. At the same time, this weighting reflects the relative biological importance of nodes within the network -- a node with high flow supplies a greater volume of cytoplasm to the rest of the network, so it is more important that all of the signals (whether cues or nucleotypes) are present at the node. We weight the negative local mixing entropies by the importance of the nodes and sum them. We get

$$\mathrm{NME} = \sum_i f_i \mathrm{NME}_i.$$

In a similar vein as [TT93] we can think of this as a conditional entropy. Consider the distribution of choosing a node at random with probability in proportion to its total flow and then conditioned on our choice of node i we chose a signal at random via the distribution $\mathcal{P}_i$.

2.1.1 Sending entropy on flows

It may also be important for a fluid network organisms to spread out signals as much as possible with their flows. We can think of this as a type of dispersion: i.e. it represents the diversity of places that a new signal originating at a node within the network will eventually reach. We can use the formalism developed above to now consider the entropy of the distribution of nodes that a signal originating at a node i will reach before exiting the network. Now, instead of taking the mass distribution of incoming flows and normalizing them to a probability distribution, we use the out-going flows. That is we define

$$\mathcal{P}'_i(j) \quad = \quad \frac{\tilde{q}_{ij}}{\sum_{j:\tilde{q}_{ij}>0} \tilde{q}_{ij}}.$$

This is equivalent to diving the entries in 2.1.1 by the sum over their rows.

We now define the **local sending entropy at node** i to be the shannon information entropy

$$H(\mathcal{P}'_i) \quad = \quad -\sum_j \mathcal{P}'_i(j) \log(\mathcal{P}'_i(j)) = -\sum_j \frac{\tilde{q}_{ij}}{\sum_{j:\tilde{q}_{ij}>0} \tilde{q}_{ij}} \log\left(\frac{\tilde{q}_{ij}}{\sum_{j:\tilde{q}_{ij}>0} \tilde{q}_{ij}}\right).$$

Now in a similar fashion to the negative local mixing entropy we define the **negative local sending entropy at node** i to be

$$\mathrm{NSE}_i \quad = \quad \sum_j \frac{\tilde{q}_{ij}}{\sum_{j:\tilde{q}_{ij}>0} \tilde{q}_{ij}} \log\left(\frac{\tilde{q}_{ij}}{\sum_{j:\tilde{q}_{ij}>0} \tilde{q}_{ij}}\right).$$

We also define the **negative total sending entropy** of the entire flow network q_{ij} to be

$$\mathrm{NSE} \quad = \quad \sum_i f_i \mathrm{NSE}_i.$$

The sending entropy can be considered being adjoint to the mixing entropy. Notice the

similarities in their calculation: we use the adjoint of the matrix of $\tilde{q}_{ij}$, $\tilde{q}_{ji}$ to calculate $\mathcal{P}'_i(j)$. We will show that the NSE is equal to the NME on a flow network obtained by reversing the direction of all of the flows.

Theorem 2.1. *Let q_{ij} be a flow network compatible with boundary flows Q_i. Let q'_{ij} and Q'_i be the flow network and boundary flows obtained from q_{ij} and Q_i by reversing the flows, i.e. $q'_{ij} = -q_{ij}$ and $Q'_i = -Q_i$. Then NSE(q_{ij}) = NME(q'_{ij}) .*

Proof. First, to prove this theorem we note that the node strengths for the flow network $f_i = \sum_{j:q_{ij}>0} q_{ij} + |Q_i|\mathbf{1}_{Q_i<0}$ are the same as for the network with the reversed flows because the total flow at i is $\sum_{j:q'_{ij}>0} q'_{ij} + |Q'_i|\mathbf{1}_{Q'_i<0}$. By the way we defined reversed flows and by Kirckhoff's 1st law

$$\begin{aligned}\sum_{j:q_{ij}>0} q_{ij} + |Q_i|\mathbf{1}_{Q_i<0} - \sum_{j:q'_{ij}>0} q'_{ij} - |Q'_i|\mathbf{1}_{Q'_i<0} &= \sum_{j:q_{ij}>0} q_{ij} - Q_i\mathbf{1}_{Q_i<0} + \sum_{j:q_{ij}<0} q_{ij} - Q_i\mathbf{1}_{Q_i>0} \\ &= \sum_{j\in n(i)} q_{ij} - Q_i = 0.\end{aligned}$$

Thus $\sum_{j:q'_{ij}>0} q'_{ij} + |Q'_i|\mathbf{1}_{Q'_i<0} = \sum_{j:q_{ij}>0} q_{ij} + |Q_i|\mathbf{1}_{Q_i<0}$ and we refer to the total flow of node i for both flow networks as f_i.

In order to prove this theorem it suffices to show that $\tilde{q}_{ij} = \tilde{q}'_{ji}$ where $\tilde{q}'_{ij}$ is defined the same way as $\tilde{q}_{ij}$ for the flow network q'_{ij}. Furthermore, this can be reduced to showing that for all signal paths x_t $t = 0, 1, \ldots, T$ that follow positive flows the probability given by the Markov chain T_{ij} the probability of the path multiplied by f_{x_0} is equal to the probability of the reversed path $x'_t = x_{T-t}$ with the Markov chain $T'_{ij} = \frac{q'_{ij}}{f_i}$ multiplied by f_{x_T}. If this statement holds true then $\tilde{q}_{x_0x_T} = f_{x_0}P_{x_0x_T}$ will be equal to $\tilde{q}'_{x_Tx_0} = f_{x_T}P'_{x_Tx_0}$ where, again P'_{ij} is calculated in the same way that P_{ij} is on q_{ij} except it is for the flows q'_{ij}.

Since $|q_{ij}| = |q'_{ij}|$ it follows that $T_{ij}f_i = T'_{ji}f_j$ and so $T'_{ji} = \frac{f_i}{f_j}T_{ij}$. We multiply the probability of the path x_t given by the Markov chain with kernel T_{ij} by the node strength $f_{x_0}$to obtain $f_{x_0}\prod_{t=0}^{T-1} T_{x_tx_{t+1}}$. We use the telescoping product form of $f_{x_0} = f_{x_T}\prod_{t=0}^{T}\frac{f_{x_t}}{f_{x_{t+1}}}$

to obtain

$$\begin{aligned} f_{x_0}\prod_{t=0}^{T-1}T_{x_tx_{t+1}} &= f_{x_T}\prod_{t=0}^{T}\frac{f_{x_t}}{f_{x_{t+1}}}T_{x_tx_{t+1}} \\ &= f_{x_T}\prod_{t=0}^{T}T'_{x_{t+1}x_t} \\ &= f_{x_T}\prod_{t=0}^{T}T'_{x'_tx'_{t+1}}. \end{aligned}$$

For any nodes i and j, using this formula, if we sum over the probability of all possible paths we obtain $\tilde{q}_{ij} = f_iP_{ij} = f_jP'_{ji} = \tilde{q}'_{ji}$. Therefore the matrix with i,j entry $\tilde{q}'_{ij}$ the transpose of the matrix with i,j entry $\tilde{q}_{ij}$. Therefore the distributions $\mathcal{P}'_i(j)$ for the flow network q_{ij} is equal to the distributions $\mathcal{P}_i(j)$ for the network q'_{ij}. Therefore $H(\mathcal{P}'_i(j)$ (for network q_{ij})) = $H(\mathcal{P}_i(j)$ (for network q'_{ij})) and so the sending entropy of the original network is the same as the mixing entropy of the reversed network. □

The flow network we focus on in this thesis is the physical flow network (see Definition 2.7): the physical flow network, or flow network is the unique assignment of flows to each edge of the network that is compatible with a given set of boundary flows Q_i and that minimizes a function of flow networks called the dissipation (denoted $\mathcal{D}$). The dissipation is symmetric with regards to reversing flow; That is, $\mathcal{D}(-q_{ij}) = \mathcal{D}(q_{ij})$ for all flow networks q_{ij}. Notice that because of this symmetry and the linearity of the compatibility conditions, the adjoint of the physical flow can be achieved by reversing the sources and sinks in the network; that is, replacing a source with inflow Q_i by a sink with outflow Q_i, and conversely. In most of the networks we will be considering, there is a single source and a single sink, with matching inflow and outflow. Thus the adjoint flow network is simply the same conductance network with source and sink exchanged.

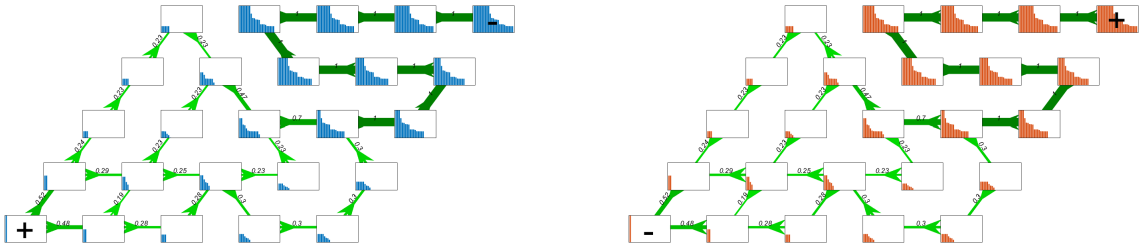


Figure 2.1.1: Two flow networks compatible with a single source (labeled +) and sink (labled −) of magnitude 1. The flows, source and sink of the network on the right are the reverse of those on the left. Depicted with a bar graph at each node i are the unnormalized distributions $\tilde{q}_{ji}$ for computing NME in blue and $\tilde{q}_{ij}$ for computing NSE in orange.

2.2 Fluid Transport Networks

2.2.1 Biophysical background

Our research stems from the general question based on prior work: what are the geometrical or topological features that give rise to the ability found in some network organisms to mix the consituents of their cytoplasm or transfer information. In some biological networks, advection currents in the form of flows driven by inflow and outflows at approximately fixed locations span spatially connected regions containing many vessels. Transportation of of nuclei in *Neurospora Crassa* mycelial networks have been observed to be driven by pressure gradient with fluid flowing towards growing tips [RSH13, HOG12]. Leaves are venous organs that receive fluid from the stem to be released through evaporation on the surface via connections between the venous network and stomatal pores that are distributed throughout the entire network [KSM10, HC13]. *Physarum polycephalum* flows are caused by waves of contractions that span the entire network [NYU00, AAP13, AKF17]. In blood vessels, pressures created by the pumping heat push oxygen carrying blood and glucose carrying plasma through a closed loop formed of arteries and veins[CR18, CR19, CTB17].

We make the simplifying assumption that fluid transport currents are driven by constant boundary flows. We also assume that the total volume contained by the network is constant

so we take the total volume input to be equal to the total output. Another assumption we make is based on the fact that network morphology for these examples is approximately fixed over short time scales. Thus we investigate the effect of fixed network geometries, in which all vessel lengths and radii, and topological features of the fluid transport networks can be fully specified. In this section we describe our mathematical construction using a network with single edge weights representing their conductance. In Section 2.4, we describe how single conductances capture both length and radius parameters. Given a set of boundary flows and a network of pipes with fixed conductance, we can derive an explicit formula (Theorem 2.3) for the flows within the network. This allows us to express functions of the flow on the network in terms of the network's geometry and topology. Functions defined in terms of flows q_{ij}, such as $\mathrm{NME}(q_{ij})$ and $\mathrm{NSE}(q_{ij})$, or that depend both upon conductances and flows such as dissipation, can be all written as functions of conductances alone. For example $\mathrm{NME}(\kappa_{ij}) = \mathrm{NME}(q_{ij})$ where q_{ij} is the physical flow of κ_{ij}. We now mathematically describe our model for fluid transport on networks of vessels.

2.2.2 Mathematical description

We model the fluid transport network in the form of a conductance network. A conductance network is a weighted graph with on a set of nodes $\mathcal{N}$and edges $\mathcal{E}$.

Definition 2.1. For all $i, j \in \mathcal{N}$ such that $(i, j) \in \mathcal{E}$ the weight on the edge (i, j) is κ_{ij}, referred to as the **conductance** on (i, j).

We represent the rate of volume of fluid entering the network and leaving the network at each node by real numbers called boundary flows.

Definition 2.2. The rate of fluid entering or exiting the network at node i is the **boundary flow at node** i denoted Q_i for all $i \in \mathcal{N}$.

Positive Q_i corresponds to fluid entering the network node i with rate Q_i, and negative Q_i correspondes to fluid exiting the network at node i with rate $|Q_i|$.

Definition 2.3. If $i \in \mathcal{N}$ and $Q_i > 0$ then we call node i a **source**. If $i \in \mathcal{N}$ and $Q_i < 0$ we call node i a **sink.**

In this paper we assume the total volume of fluid is fixed and impose this by assuming $\sum_i Q_i = 0$. We assume fluid is transported through the network at a constant rate of volume per time along each edge. The rate of volume transported along each edge is represented by real numbers called flows and they can be positive or negative depending on the direction of transportation.

Definition 2.4. A **flow network** is a weighted directed network on the set of nodes $\mathcal{N}$ and directed edges $[i, j]$ and $[j, i]$ for all $(i, j) \in \mathcal{E}$ with weights q_{ij} such that $q_{ij} = -q_{ji}$. The weight of the directed edge from i to j is the **flow from** i **to** j.

We also require that the total flow entering a node is balanced with the flow exiting every node. For a node i, this assumption considers flows from within the network flowing along edges adjacent to i as well as the flow entering or leaving the network at i.

Definition 2.5. A flow is called **compatible** with regards to the boundary flows Q_i if $\sum_{j \in n(i)} q_{ij} = Q_i$ for all $i \in \mathcal{N}$ where $n(i)$ is the undirected neighbor set of node i. A flow network being compatible with the set of boundary flows is known as Kirchoff's second circuit law.

Because the edge and boundary flows are so balanced and do not vary with an extra time parameter we say that the fluid network is in a steady state. Since we are are working in a physical setting, we assume that the flows along edges will be those that minimize the power used to pump fluid into the sources and out of the sinks at their respective rates. The specific flow we are interested in is what is referred to as the physical flow. We will give an explicit expression for the physical flow in terms of the conductances and the boundary flows. First we define the dissipation and present an equation for power used by a flow through a conductance network.

Definition 2.6. For a conductance network κ_{ij}, the **dissipation** $\mathcal{D}$ by flow q_{ij} is the power used when fluid is transported along the edges with flows q_{ij}. It is defined as a function $\mathcal{D}(q_{ij}) = \sum_{ij} \frac{q_{ij}^2}{\kappa_{ij}}$.

Let $\mathcal{N}$ and $\mathcal{E}$ be the set of nodes and edges of a conductance network with conductances κ_{ij} for each $(i,j) \in \mathcal{E}$. Let Q_i be a set of boundary flows on $\mathcal{N}$ such that $\sum_i Q_i = 0$. A physical flow is then the flow that minimizes the dissipation $\mathcal{D}$ over the set of all compatible flows.

Definition 2.7. For a conductance network κ_{ij} on a set of nodes $\mathcal{N}$ and edges $\mathcal{E}$ and a set of boundary flow Q_i such that $\sum_i Q_i = 0$. A **physical flow** q_{ij} is a flow on edges $\mathcal{E}$ which is compatible with Q_i (i.e. $\sum_{j \in n(i)} q_{ij} + Q_i = 0$ for all $i \in \mathcal{N}$) and minimizes $\mathcal{D}$ over the set of all compatble flows. That is, q_{ij} is a flow network such that $\sum_{j \in n(i)} q_{ij} + Q_i = 0$ for all $i \in \mathcal{N}$ and

$$\sum_{(i,j)} \frac{q_{ij}^2}{\kappa_{ij}} \quad = \quad \min \left\{ \sum_{(i,j)} \frac{f_{ij}^2}{\kappa_{ij}} : f_{ij} \in \mathbb{R}, f_{ij} = -f_{ji} \, \forall (i,j) \in \mathcal{E}, \sum_{j \in n(i)} f_{ij} = Q_i \, \forall i \in \mathcal{N} \right\}.$$

Although we have defined the flows that would arise in a physical situation when boundary flows are applied to a conductance network, we have yet to rigorously prove their existence. It turns out that not only do they exist, but they are unique according to what is known in physics as Thomson's principle. We now prove this.

Theorem 2.2. *(Thomson's principle) Let κ_{ij} be a set of conductances on a network with nodes $\mathcal{N}$ and edges $\mathcal{E}$ and let Q_i be a set of boundary flows on the nodes $\mathcal{N}$. If the network is in a steady state (i.e. $\sum_i Q_i = 0$) then there exists a unique set of flows q_{ij} on the edges $\mathcal{E}$ with $q_{ij} = -q_{ji}$ for all $(i,j) \in \mathcal{E}$ obeying compatible with the boundary flows Q_i (i.e. $\sum_j q_{ij} + Q_i = 0$) for all $i \in \mathcal{N}$*

Proof. Thomson's principle is equivalent to showing that there is a unique set of flows in the affine sub-space $S = \left\{ q_{ij} : \sum_j q_{ij} = Q_i \text{ for all } i \in \mathcal{N} \right\}$ that minimizes $\mathcal{D}$ over S. We

first establish that S is non-empty. The collection of sets of boundary flows that sum to zero is linearly spanned by the collection of boundary flows which are 1 at node u, -1 at node v and zero for all other nodes in the network, for all pairs of nodes u, v. Since the conductance network is connected let $x_0 = u, x_1, \ldots, x_t = v$ for some positive integer t be a simple (non-self intersecting) directed path connected u to v and define the flow τ_{uv} to be the flows $q_{x_i x_{i+1}} = 1$, $q_{x_{i+1} x_i} = -1$ for all $1 \leq i < t$ and $q_{ab} = 0$ on all other pairs. Then τ_{uv} is compatible with a set of boundary flows which are 1 at u , -1 at v and 0 elsewhere. For an arbitrary set of boundary flows $Q = \{Q_i\}_{i \in \mathcal{N}}$ such that $\sum_i Q_i = 0$ let a_{uv} $u, v \in \mathcal{N}$ the linear combination of the flows $\delta_u - \delta_v$ (where δ_i is the boundary flow defined to be 1 on i and 0 elsewhere) $\sum_{u,v \in \mathcal{N}} a_{uv}(\delta_u - \delta_v) = Q$. Then $\sum_{u,v \in \mathcal{N}} a_{uv} \tau_{uv}$ is a flow compatible with boundary flows Q_i.

We show $\mathcal{D}$ attains a minimum over the domain of flows compatible with Q_i and that the minimizer is unique. First note that $\sum_{(i,j)} \frac{q_{ij}^2}{\kappa_{ij}}$ is a quadratic form in the flows q_{ij} with postive coefficients, so it is positive definite. If S has dimension n and we parametrize S with coordinates $(x_1, x_2, \ldots, x_n) \mapsto \sum_{i=1}^n x_i v_i + p$ where $\{v_i\}_{i=1}^n$ is a basis for S and p is some point $p \in S$ we have that the restriction of $\sum_{(i,j)} \frac{q_{ij}^2}{\kappa_{ij}}$ to S in terms of coordinates $\{x_i\}_{i=1}^n$ is a positive definite quadratric function. Therefore $\mathcal{D}$ attains a minimum over S , and by positive definiteness of the $\mathcal{D}$ restricted to S the minimizer is unique. □

Thomson's principle enables us to consider a mapping from the conductances to the set of physical flows. We make extensive use of this to express functions defined on the flows or on both the flows and the conductances completely in terms of κ_{ij} in a physics setting. As a first example and of central importance to this research, given a set of boundary flows, we define the dissipation of a conductance network with boundary flows to be $\sum_{(i,j)} \frac{q_{ij}^2}{\kappa_{ij}}$ where q_{ij} are the physical flows of the network.

Definition 2.8. Let κ_{ij} be a conductance network and Q_i be boundary flows on the nodes such that $\sum_i Q_i = 0$. We define the **dissipation of** κ_{ij} , written as $D(\kappa_{ij})$ to be the dissipation of the network with conductances κ_{ij} and the corresponding physical flows for

κ_{ij} with boundary conditions Q_i. This can be written

$$D(\kappa_{ij}) \quad = \quad \min_{\text{flows } q_{ij} \text{ compatible with } Q_i} \sum_{(i,j)} \frac{q_{ij}^2}{\kappa_{ij}}.$$

We also define the NME and the NSE to have alternate definitions when applied to conductance rather than flow networks.

Definition 2.9. Given a set boundary flows Q_i we define the **negative mixing entropy as a function of network conductances** κ_{ij} to be $\text{NME}(\kappa_{ij}) = \text{NME}(q_{ij})$ where q_{ij} are the physical flows for the conductance network κ_{ij} with boundary flows Q_i. Similarly, we define the **negative sending entropy as a function of network conductances** κ_{ij} to be $\text{NSE}(\kappa_{ij}) = \text{NSE}(q_i)$.

Thomson's principle allows us to pull-back functions on flows representing how well a transport network achieves an objective, to functions on conductance networks, which represent vessel radii, length and network topology– the degrees of freedom that the organism building the network has direct control over. We will derive an explicit formula for physical flows given a conductance network κ_{ij} and boundary flows. First we have the definition which will be helpful in making the formula more compact and is used Chapter 5 for designing numerical optimization algorithms.

Definition 2.10. For a conductance network κ_{ij} on nodes $\mathcal{N} = \{1, 2, \ldots N\}$ for some $N \in \mathbb{N}$, we define the **Laplacian of** κ_{ij} $\Delta_{\kappa_{ij}}$, to be the $N \times N$ symmetric matrix with real entries assigned as follows. Let a_{ij} be the i, j-th entry of $\Delta_{\kappa_{ij}}$. For all $i \in \mathcal{N}$

$$a_{ii} \quad = \quad \sum_{j \in n(i)} \kappa_{ij}$$

and for all $(i,j) \in \mathcal{E}$ let

$$\begin{aligned} a_{ij} &= -\kappa_{ij} \\ a_{ji} &= -\kappa_{ji}. \end{aligned}$$

Again, for simplicity, we assume that κ_{ij} is connected. Otherwise, we can solve for the physical flows on each connected component. The theorem stating the equation for the boundary flows has two parts. First we prove the existence of real quantities called the pressures at node i such that the physical flow q_{ij} is in terms of κ_{ij}, p_i and p_j. Then we describe a linear system for which the p_i are a solution.

Theorem 2.3. *Let κ_{ij} be a connected conductance network with nodes $\mathcal{N} = \{1, 2, \ldots, N\}$ for some $N \in \mathbb{N}$ and edges $\mathcal{E}$. Let Q_i be boundary flows for all $i \in \mathcal{N}$ such that $\sum_i Q_i = 0$. Let q_{ij} be the physical flows of this network and boundary flows.*

(a.) Then there exists $p_i \in \mathbb{R}$,called the pressure at node i, such that $q_{ij} = \kappa_{ij}(p_j - p_j)$.

(b.) And $v = \{v_i\}_{i=1}^N \in \mathbb{R}^N$ is such that $q_{ij} = \kappa_{ij}(v_j - v_i)$ for all $(i,j) \in \mathcal{E}$ if and only if it is a non-trivial solution of the system of linear equations

$$\left(\sum_{j \in n(i)} \kappa_{ij} \right) v_i - \sum_{j \in n(i)} \kappa_{ij} v_j = \sum_{j \in n(i)} \kappa_{ij}(v_i - v_j) = Q_i. \tag{2.2.1}$$

Remark 2.1. We can express equation 2.2.1 in matrix form as

$$\Delta_{\kappa_{ij}} p = Q$$

where p is a column vector and Q is the column vector with Q_i as its i-th entry.

Proof. Let κ_{ij}, $\mathcal{N}$, $\mathcal{E}$ and Q_i be as defined in the hypothesis. We first prove part (a), that for a set of flows compatible with Q_i, obeying Thomson's principle is equivalent the existence $p_i \in \mathbb{R}$ for each node i such that $q_{ij} = \kappa_{ij}(p_i - p_j)$ for all $(i,j) \in \mathcal{E}$. Recall,

the dissipation of a any flow given the conductances κ_{ij} is defined $\mathcal{D}(f_{ij}) = \sum_{(i,j)} \frac{f_{ij}^2}{\kappa_{ij}}$. Let $S = \left\{ f_{ij} : \sum_j f_{ij} = Q_i \text{ for all } i \in \mathcal{N} \right\}$, the affine sub-space of compatible flows. We can view $\mathcal{D}$ as a smooth function on the set of flow networks viewed as the real space $\mathbb{R}^{|\mathcal{E}|}$ where we consider one direction for each edge in the network. Then we can also view S as the intersection of N non-trivial constraint affine hyperplanes $H_i = \sum_j f_{ij} = Q_i$ in $\mathbb{R}^{|\mathcal{E}|}$ where if the summand f_{ik} for directed edge $[i,k]$ is the opposite order than that of the identification of edge (i,k) we replace it $-f_{ki}$.

Since $\mathcal{D}$ is smooth, if $\{q_{ij}\}_{(i,j)\in\mathcal{E}} \in S$ is a minimizer of $\mathcal{D}$ over the domain given as the intersection of the constraint hyperplanes $\bigcap_{i=1}^N H_i$ by the method of Lagrange multipliers there exists $\lambda \in \mathbb{R}$ such that q_{ij} is a critical point of $f_{ij} \mapsto \left(\mathcal{D}(f_{ij}) - \sum_l \lambda_k \sum_{l\in n(l)} f_{lk} \right)$. Let $u, v \in \mathcal{N}$ such that $(u,v) \in \mathcal{E}$. Taking the derivative of this function with respect to f_{uv} gives us

$$\frac{\partial}{\partial f_{ij}} \left(\mathcal{D}(f_{ij}) - \sum_l \lambda_k \sum_{l\in n(l)} f_{lk} \right) \quad = \quad \frac{2f_{uv}}{\kappa_{uv}} - \lambda_u + \lambda_v = 0.$$

By Thomson's principle (Theorem 2.2) the physical flow network q_{ij} for κ_{ij} with boundary flows Q_i exists and is the unique minimizer of $\mathcal{D}$ over S. Therefore there exists $\lambda_i \in \mathbb{R}$ for all $i = 1, 2, \ldots, N$ such that $\frac{2q_{ij}}{\kappa_{ij}} - (\lambda_i - \lambda_j) = 0$ for all $i, j \in \mathcal{N}$ and $(i,j) \in \mathcal{E}$. Thus $\kappa_{ij} \left(\frac{\lambda_i}{2} - \frac{\lambda_j}{2} \right) = q_{ij}$. Allowing $p_i = \frac{\lambda_i}{2}$ for all $i = 1, 2, \ldots, N$ we have proven the first part of the theorem. That is, there exists real numbers p_i for all $i \in \mathcal{N}$ such that $\kappa_{ij}(p_i - p_j) = q_{ij}$ for all $i, j \in \mathcal{N}$ with $(i,j) \in \mathcal{E}$.

We now prove part (b) of the theorem. First we show that $\ker \Delta_{\kappa_{ij}}$ has dimension 1 and is spanned by $c \in \mathbb{R}^N$ the vector whose entries are all 1. Plugging c into the left side of 2.2.1 we immediately have $c \in \ker \Delta_{\kappa_{ij}}$. Suppose that $d \in \ker \Delta_{\kappa_{ij}}$ $d \neq \alpha c$ for all $\alpha \neq 0$ with the i-th entry of d being d_i. We can assume all of the entries of d are non-negative and at least one of them is 0. We can make this assumption by re-assigning the value of d to be $d - (\min_{i\in\mathcal{N}} d_i)\, c \in \ker \Delta_{\kappa_{ij}}$. We can find an $k, l \in \mathcal{N}$ such that $l \in n(k)$, and $d_k = 0$

and $d_l \neq 0$. Otherwise the set $\{u : d_u = 0\}$ would be disconnected from $\{u : d_u \neq 0\}$ in the sense that there would be no paths in the conductance network from nodes in the first set to nodes in the second, contradicting the assumption that the conductance network is connected. Applying the Laplacian to d gives and focusing on the k^{th} entry, we obtain:

$$\Delta_{\kappa_{ij}} d = \left(\sum_{j \in n(k)} \kappa_{kj} \right) d_k - \sum_{j \in n(k)} \kappa_{kj} d_j = - \sum_{j \in n(k)} \kappa_{kj} d_j < -\kappa_{kl} d_l < 0$$

contradicting that $d \in \ker \Delta_{\kappa_{ij}}$. We have $\ker \Delta_{\kappa_{ij}}$ has dimension 1 and is spanned by c.

Now let $p = \{p_i\}_{i=1}^N \in \mathbb{R}^{\mathbb{N}}$ be as in the proof of part (a): $\kappa_{ij}(p_i - p_j) = q_{ij}$. Then p is a solution to 2.2.1. Let $v = \{v_i\}_{i=1}^N \in \mathbb{R}^N$ be another solution to $\Delta_{\kappa_{ij}} v = Q$. Then since the kernel of $\Delta_{\kappa_{ij}}$ is spanned by c we have $v = p + tc$ for some $t \in \mathbb{R}^N$. Therefore $v_i = p_i + t$ for all $i = 1, 2, \ldots, N$. Therefore, for all $i, j \in \mathcal{N}$ with $(i, j) \in \mathcal{E}$ we have $\kappa_{ij}(v_j - v_i) = \kappa_{ij}(p_j + t - p_i - t) = \kappa_{ij}(p_j - p_i) = q_{ij}$. □

Theorem 2.3 opens up the possibility of finding q_{ij} by using linear solvers or other methods to solve matrix equation in terms of κ_{ij}. Therefore we use this relation to find q_{ij} in our numerical experiments.

2.3 Murray's law

We now re-cast the results of Charles Murray's study of fluid transport networks optimized to minimize a biologically relevant expression for power usage [Mur26]. We model the conduits along edge (i, j) in the network as a cylinder with length l_{ij} and cross-sectional radius r_{ij} and we assume that the flow is laminar and the fluid itself is Newtonian. The Hagen-Poiseuille equation (also derived as an application to biological networks), states that we can model the conductance κ_{ij} of the vessel connecting i to j by $\kappa_{ij} = k \frac{r_{ij}^4}{l_{ij}}$, where $k > 0$ is a constant that depends only on the viscosity of the fluid [Dur07, CR17, Ach90]. Murray also posited that in addition to the cost of pushing fluid through the network (i.e. the dissipation, defined

above), maintaining the network would incur a cost based on the the fact that the fluid may either contain organelles (if it is protoplasm), or living cells (e.g. blood cells, for a cardiovascular network), the conduit walls may also need to be maintained. The first type of maintenance would lead to energy is used at a rate that we would expect to be proportional to volume of the network and the second that it is a constant times the surface area of the conduits in the network. According to the first of these assumptions, the power use of a fluid transport network is then

$$\sum_{(i,j)} \frac{q_{ij}^2}{\kappa_{ij}} + \alpha_3 \text{vol}(\kappa_{ij}) \quad = \quad \sum_{(i,j)} \frac{l_{ij}}{k r_{ij}^4} q_{ij}^2 + \sum_{(i,j)} \alpha_3 r_{ij}^2 l_{ij} \tag{2.3.1}$$

where $\alpha_3 > 0$ is a constant. The sub-script of 3 signifies that it applies to the volume of the network.

For the networks studied in this paper, in both the setting for numerical experiments as well mathematical analysis as well as, we make the assumption that $l_{ij} = 1$. Note: this assumption different vessel lengths can occur by concatenating several edges in a row. This results in the dissipation being equal to $\sum_{(i,j)} \frac{1}{k r_{ij}^4} q_{ij}^2$. Therefore the conductances along each edge are $\kappa_{ij} = k r_{ij}^4$ and power used by the volume of fluid is $\sum_{(i,j)} \alpha_3 r_{ij}^2 = \frac{\alpha_3}{\sqrt{k}} \sum_{ij} \kappa_{ij}^{\frac{1}{2}}$. Taking $\beta_3 = \frac{\alpha_3}{\sqrt{k}}$ we can write equation 2.3.1 as

$$\sum_{(i,j)} \frac{q_{ij}^2}{\kappa_{ij}} + \alpha_3 \text{vol}(\kappa_{ij}) = \sum_{(i,j)} \frac{q_{ij}^2}{\kappa_{ij}} + \beta_3 \sum_{(i,j)} \kappa_{ij}^{\frac{1}{2}}.$$

In the case where power for up-keep of the network is proportional to total surface area (with constant of proportionality $\alpha_2 > 0$) , by a similar computation we find

$$\sum_{(i,j)} \frac{q_{ij}^2}{\kappa_{ij}} + \alpha_2 \text{area}(\kappa_{ij}) \quad = \quad \sum_{(i,j)} \frac{q_{ij}^2}{\kappa_{ij}} + \beta_2 \sum_{(i,j)} \kappa_{ij}^{\frac{1}{4}}$$

for some constant $\beta_2 > 0$. We both of these laws follow the general form

$$\sum_{(i,j)} \frac{q_{ij}^2}{\kappa_{ij}} + \beta \sum_{(i,j)} \kappa_{ij}^{\gamma}$$

for $\gamma > 0$. Prior theoretical and computational work has illuminated commonalities between optimal networks for that may be more apparent if γ is allowed to take any value $0 < \gamma < 1$ [Dur07, AKF17] and not just $\frac{1}{4}$ or $\frac{1}{2}$. Also there are interesting transitions if γ is allowed to continuously vary [BM07, KSM10].

In this paper we assume that the rate of energy expended for network upkeep can be a multiple of $\sum_{(i,j)} \kappa_{ij}^{\gamma}$ for any $\gamma \in (0, 1)$. The rate of energy lost through friction is taken be dissipation of the physical flow through the conductance network (Definition 2.8). Next we make use of this generalized Murray's law, together with the NME to describe the problem that is the central theme of our paper.

2.4 Main Optimization Problem

In this paper we are interested in conductances and their physical flows which simultaneously optimize the NME, a function solely depending on flows, and the power expended by the network. In section 2.3 we express the power used by the network to be $\sum_{(i,j)} \frac{q_{ij}^2}{\kappa_{ij}} + \beta \sum_{(i,j)} \kappa_{ij}^{\gamma}$ for some $0 \leq \gamma \leq 1$ and $\beta > 0$ (Murray's Law) where the second term $\beta \sum_{(i,j)} \kappa_{ij}^{\gamma}$ is the power needed for the upkeep of the material in the network. We choose to express the upkeep term as a constsraint $\sum_{(i,j)} \kappa_{ij}^{\gamma} = C$ where $C > 0$ is referred to as the **material cost of the network**. The constant β is dropped because it can be lumped into C. This frames the energy allotted to network upkeep as a constraint on the total amount of conductance and together with the inequality constraints $\kappa_{ij} \geq 0$ for all nodes i, j makes the domain compact.

On the other hand, we take dissipation into account by defining the **cost of mixing and dissipation for weight** $c > 0$ (often shortened to mixing-dissipation cost) denoted CMD

(or frequently through this thesis as θ) to be

$$\text{CMD} \quad = \quad \text{NME}(\kappa_{ij}) + cD(\kappa_{ij}).$$

To formalize our optimization problem we need to restrict the space of possible networks so that their connectivity, such as the maximum number of vessels at a junction, is physically and biologically relevant. Let G be an unweighted undirected network with nodes $\mathcal{N}$ and edges $\mathcal{E}$ with the connectivity properties we require optimal conductance networks to posess. We then restrict consideration to conductance networks whose nodes and unweighted edges are subnetworks of G. We refer to G as the **ambient network**. We say that κ_{ij} **is a conductance network on a network** G if the nodes and edges of κ_{ij} form a sub-network of G. We refer to a conductance network κ_{ij} on G that lacks an edge $(u, v) \in \mathcal{E}$ interchangeably with the network that has this edge, but with $\kappa_{uv} = 0$.

We are now ready to summarize the main theme of this part of the thesis: to explore and characterize fluid transport networks (and their mixing-dissipation costs) that solve the following optimization problem for different weights on dissipation.

Let G be a undirected network with nodes $\mathcal{N}$ and edges $\mathcal{E}$ and Q_i for all $i \in \mathcal{N}$ be a boundary flow on G. For a material cost $C > 0$, material cost exponent γ and weight $c > 0$, we are intererested in conductance networks and their physical flows thar are solutions to

$$\begin{aligned} \text{minimize} \quad &: \quad \text{NME}(\kappa_{ij}) + cD(\kappa_{ij}) \\ \text{subject to constraints} \quad &: \quad \kappa_{ij} \geq 0 \\ &\quad \sum_{(i,j)} \kappa_{ij}^{\gamma} = C. \end{aligned} \tag{2.4.1}$$

Since we are interested in the minima for different c, we define the function we refer to as

the **minimal mixing-dissipation** denoted θ as

$$\theta(c) \quad = \quad \min_{\sum_{(i,j)} \kappa_{ij}^{\gamma} = C} \mathrm{NME}(\kappa_{ij}) + cD(\kappa_{ij}).$$

Fact 2.1. *The minimal mixing-dissipation cost is a concave function of c. That is, for* $c_1, c_2 \geq 0$ $\theta(tc_1 + (1-t)c_2) \geq t\theta(c_1) + (1-t)\theta(c_2)$ *for all* $0 \leq t \leq 1$.

Proof. Let κ_1 be a minimizer of $\mathrm{NME}+c_1 D$ and κ_2 a minimizer of $\mathrm{NME}+c_1 D$. Let $0 \leq t \leq 1$ and κ_3 be a minimizer of $\mathrm{NME} + (tc_1 + (1-t)c_2)\, D$. Then

$$\begin{aligned} t\mathrm{NME}(\kappa_1) + tc_1 D(\kappa_1) + (1-t)\mathrm{NME}(\kappa_2) + (1-t)c_2 D(\kappa_2) \quad &\geq \quad t\mathrm{NME}(\kappa_3) + tc_1 D(\kappa_3) + (1-t)\mathrm{NME}(\kappa_3) + \\ &= \quad \mathrm{NME}(\kappa_3) + (tc_1 + (1-t)c_2)\, D(\kappa_3). \end{aligned}$$

□

We now discuss the classes of ambient networks we consider in more depth.

We consider classes of networks that represent equally spaced points in a 2-dimensional plane to model the morphology of biological fluid transport networks that live on flat surfaces, as fungal mycelia and slime mold often do. As was shown in Section 2.3, every edge is assumed to have equal length so the cost of maintaining the network may be expressed by a sum of the conductances raised to a power γ. Due to the simplicity of their parametrization and generation by size the planar ambient networks we consider are square and triangular grids.

Definition 2.11. A **square grid of side length** N is an undirected network with topology equivalent to a $N \times N$ square grid graph. It is isomorphic to a graph whose nodes are the set $\mathcal{N} = \{1, 2, \ldots, N\} \times \{1, 2, \ldots, N\}$ of ordered pairs $\langle i, j \rangle$ $i, j \in \{1, 2, \ldots, N\}$ and edges $(\langle i, j \rangle, \langle i+1, j \rangle)$,$(\langle j, i \rangle, \langle j, i+1 \rangle)$ for all $i = 1, 2, \ldots, N-1$ and $j = 1, 2, \ldots, N$.

Definition 2.12. A **triangular grid of side length** N is an undirected network with the same nodes as the square grid of side length N whose edges contain the edges of the

square grid of side length N. Included in the set of edges are edges that connect the top left corner to the bottom left corner of each square. It is isomorphic to a graph with the set of nodes $\mathcal{N} = \{1, 2, \ldots, N\} \times \{1, 2, \ldots, N\}$ of ordered pairs $\langle i, j \rangle$ and edges $(\langle i, j \rangle, \langle i+1, j \rangle)$,$(\langle j, i \rangle, \langle j, i+1 \rangle)$ for all $i = 1, 2, \ldots, N-1$ and $j = 1, 2, \ldots, N$ (from the square grid) and $(\langle i+1, j \rangle, \langle i, j+1 \rangle)$ for all $1 \leq i, j \leq N-1$ (diagonal edges).

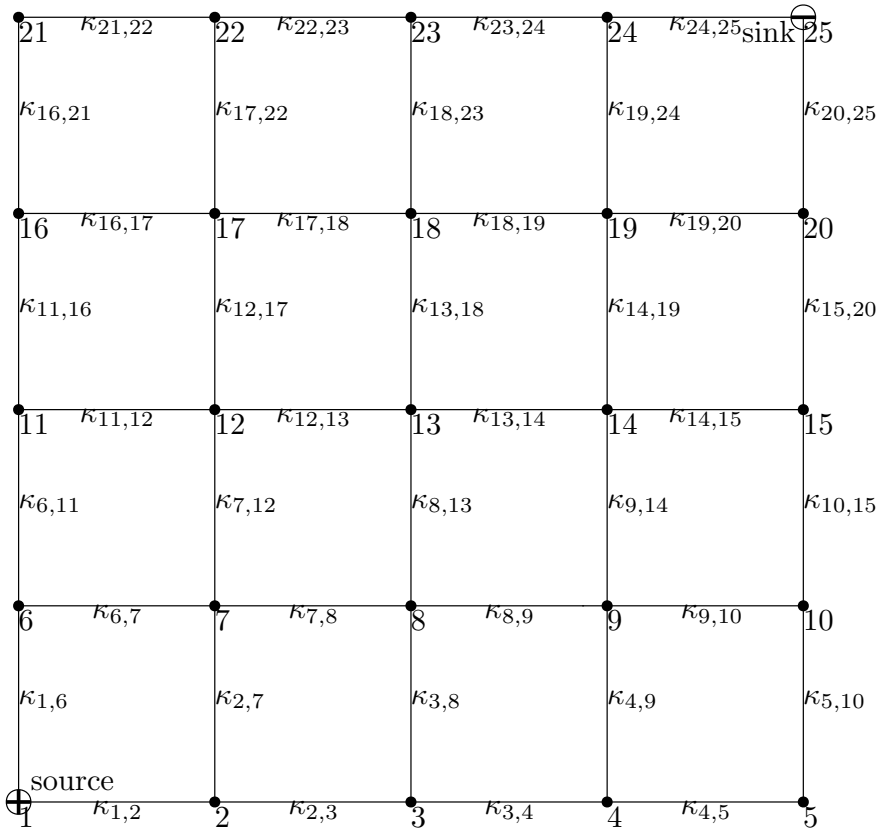


Figure 2.4.1: Example of the square grid with side length 5, showing labeling of vertices.

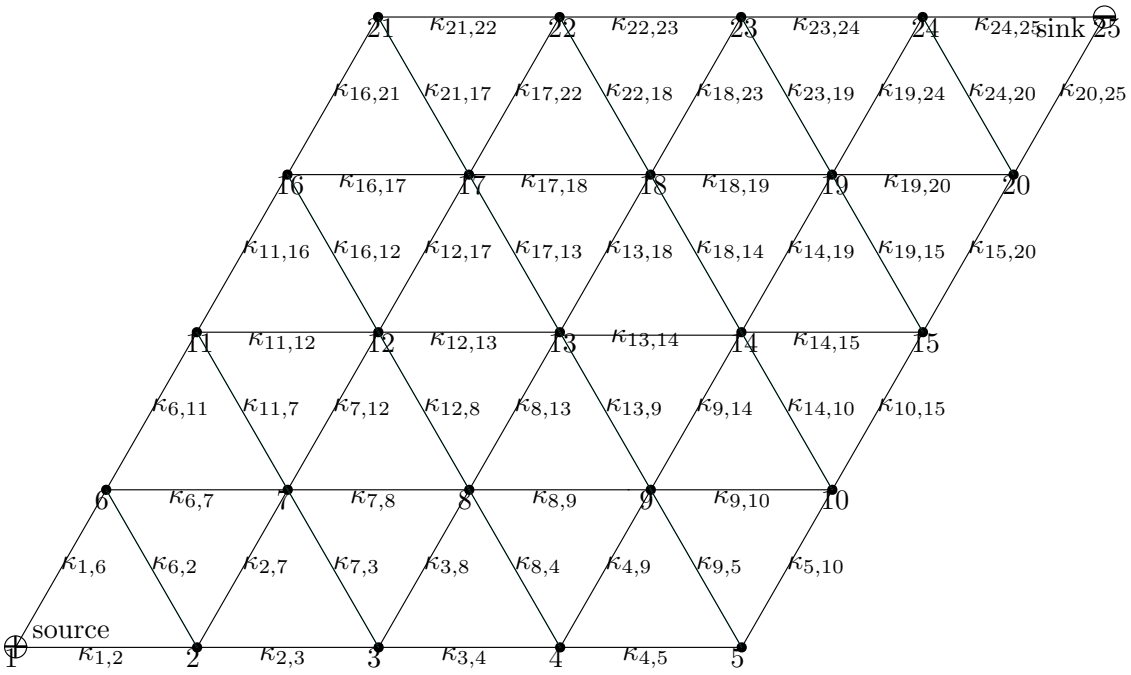


Figure 2.4.2: Example of the triangular grid with side length 5, showing labeling of vertices.

2.5 Basic Network Concepts

Given either a conductance or a flow network, we use the term **loop** to mean a graph cycle in the unweighted undirected graph on the same nodes where there is an edge between two nodes if they are connected in the network. For fluid flow networks, one quantity we are interested in is the **number of independent loops.** For a network embeddable in 2 dimensions, we define the number of independent loops to be the number of cycles whose 2 dimensional interior does not contain any edges. It is more common to define the independent loops using group theory [Hat02], but we choose our definition because it is clear and does not require introducing mathematical concepts that are otherwise isolated from the rest of the paper. We compute this number of independent loops via the formula

$$\#\text{ of independent loops} \;=\; |\mathcal{E}| - |\mathcal{N}| + \#\text{ connected components.} \tag{2.5.1}$$

A type of network that will be immediately relevant in our numerical results is the **path of length** n. For now, we define a path of length n to be any conductance or flow network that contains n nodes, which we can re-label $1, 2, \ldots, n$ with $\{(i, i+1) : i \in \mathbb{Z}, 1 \leq i < n\}$ as its edge set. Many of the approximate mixing-dissipation cost minimizers we will encounter are not paths, but a more general notion of flow network path length will illuminate interesting patterns. Assume that the initial state of a random walker x_0 is a random variable that chooses one of the source nodes with probability in proportion to its stength relative to the other source nodes. Then the average path length is then calculated as $\sum_{i \in \text{sources}} \frac{Q_i}{\sum_{j \in \text{sources}} Q_j} \sum_k P_{ik}$.

CHAPTER 3

Highlights of Numerical Results

3.1 Preliminary result: optimal mixing networks on the 3×3 square grid

In our numerical results section, we computationally search for conductance networks κ_{ij} which approximately minimize NME $+ cD$ for all c in a dense finite mesh of an interval $(0, c_{max}]$ with constraints $\sum_{(i,j)} \kappa_{ij}^{\gamma} = C$ and for different values of γ on both square and triangular grids.

We begin our investigation by taking the ambient network to be a 3×3 square grid with one source at the bottom left corner and one sink placed at the top right corner, both of strength one. Consider the biologically relevant Murray exponent $\gamma = .5$ and total building cost $C = 8$. We choose this example because it is tractable to plain reasoning while it contains a sufficient multitude of sub-networks to provide us with a detailed enough idea of the structure of optimal mixing-dissipation networks at different costs of dissipation. From our results we will conjecture the properties of such networks on grids of arbitrary size. Recall the definition $\theta(c) = \min_{\sum_{(i,j)} \kappa_{ij}^{\gamma}=8} \mathrm{NME}(\kappa_{ij}) + cD(\kappa_{ij})$ for all $c > 0$. We approximate the graph of θ for c on the interval $[0, 2.5]$ by sampling networks minimizing the mixing dissipation cost on 50 evenly spaced points c, $.01 \leq c \leq 2.5$.

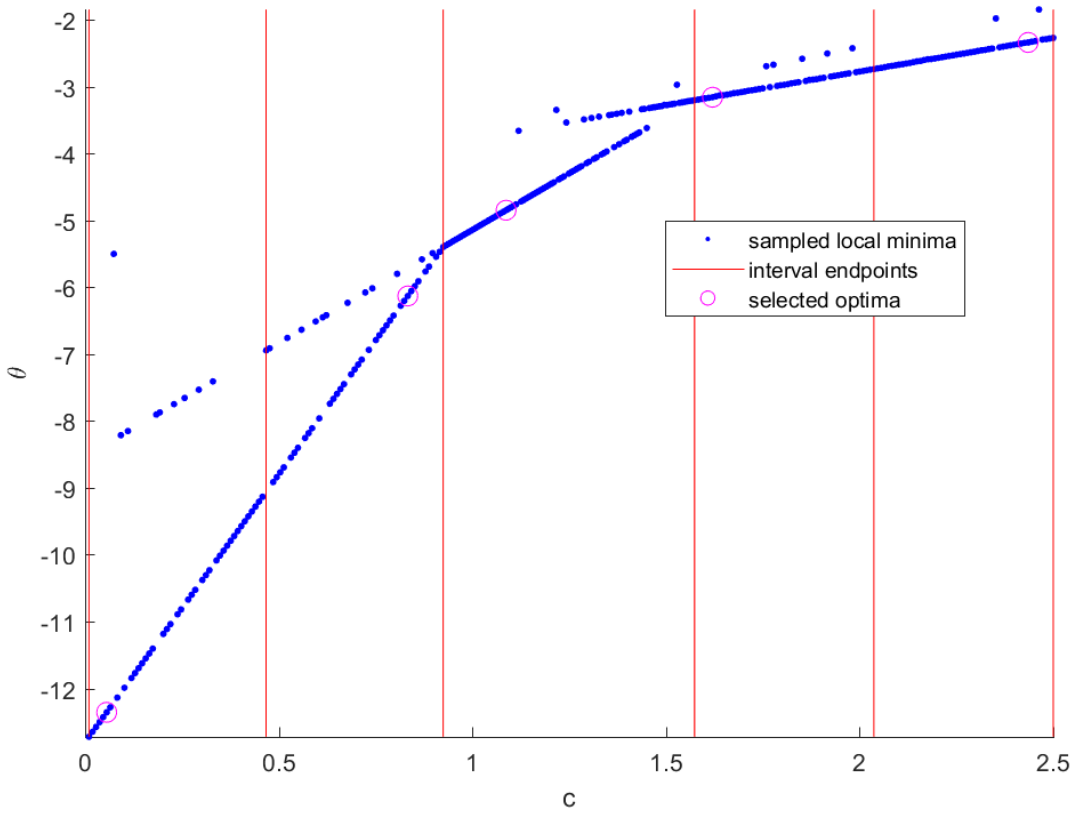


Figure 3.1.1: Numerical optima of NME $+ cD$ for networks minimizing the cost of mixing and dissipation (blue points) over the interval $.01 < c < 2.5$. The vertical red lines represent the end points of subintervals in which $\theta(c)$ is affine in c and the mid-points of the first and third intervals. Magenta points represent the c and NME $+ cD$ globally optimal networks – i.e. networks with minimum distance from the minimal cost curve. (See Chapter 6 Section 6.8.1).

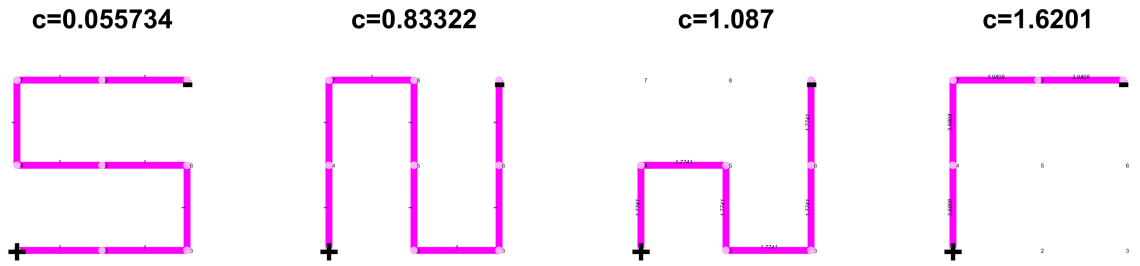


Figure 3.1.2: Optimal mixing-dissipation conductance networks κ_{ij} on the square grid with side length 3 for $.01 \leq c \leq 2.5$ selected according to figure 3.1.1appearing in order of increasing c and appearing left to right .

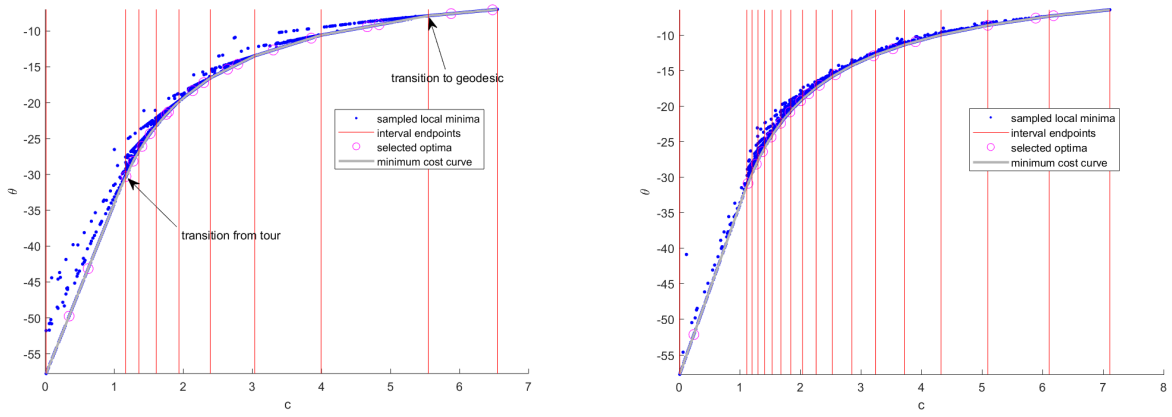
The graph of approximated minima of NME $+ cD$ in figure 3.1.1 leads us to conjecture that the graph of θ is piecewise linear for the ambient network 3×3 grids, maybe even so for larger grids and different values of γ. The optimal networks at the extremes are especially meaningful. Since the left end point is $c = .01$ the mixing-dissipation cost is approximately the total-negative-mixing entropy. We see that the path is the longest possible path that can be embedded in the network in figure 3.1.2. The unweighted graph of the network at the right endpoint $c = 2.5$ is the shortest path by number of nodes between the source and sink in the network and as we show in Chapter 2., this is the dissipation minimizer. In figure 3.1.2 we also choose sample optima on the interior of the three affine segments. Note that with increasing c the length of the path decreases by 2 which is the minimal difference between lengths of paths on a square grid ambient network. Visual inspection of the networks sampled show that they are all paths of lengths 9, 7 and 5. Each of the three piecewise affine parts has an expression as a line $\theta(c) = ac + b$ and it appears that each piece is of the form $c \mapsto \text{NME}(\tau_n) + cD(\tau_n)$.

As we will see in the next section, our numerical results with the same parameters except considering different ambient grids of dimensions 5×5 reveal a similar trend, except there are loops in the conductance networks located at the source, corresponding to bifurcations in the flow networks. We state the non-rigorous conjecture based on these observations: On grid ambient networks with 1 source and 1 sink, the graphs of optimal costs of mixing

and dissipation are *based off of* piecewise affine functions where each line is generated by path networks and the lengths of the generating paths decrease by the minimal path length difference of the ambient grid for increasing c. Here by *based off of* we convey that the graphs are generated piecewise by the functions $c \mapsto \mathrm{NME}(\tau_n) + cD(\tau_n)$ except that in open intervals around the transition points $c_{n,m}$, at which the network transitions from the path τ_n to the path τ_m where n and m are two subsequent path lengths for the given grid, there are optimal conductance networks which contain loops (see Fig. 4.2.5 for an example). Also we conjecture that these transitional networks represent intermediate states between two adjacent paths. These did not appear in our simulations on 3×3 square grids, but transitions between longer networks may have loops. We expand on our conjecture in Section 4.2.2 where we perform the same experiments except with $\gamma = .8$.

3.2 Numerical results for 1 source and 1 sink on ambient grid networks of side-length 5 and biologically relevant $\gamma = .5$

The first in depth numerical study is to optimize the cost of mixing and dissipation on square and triangular grids of side-length 5 with $\gamma = .5$ and $C = 24$.

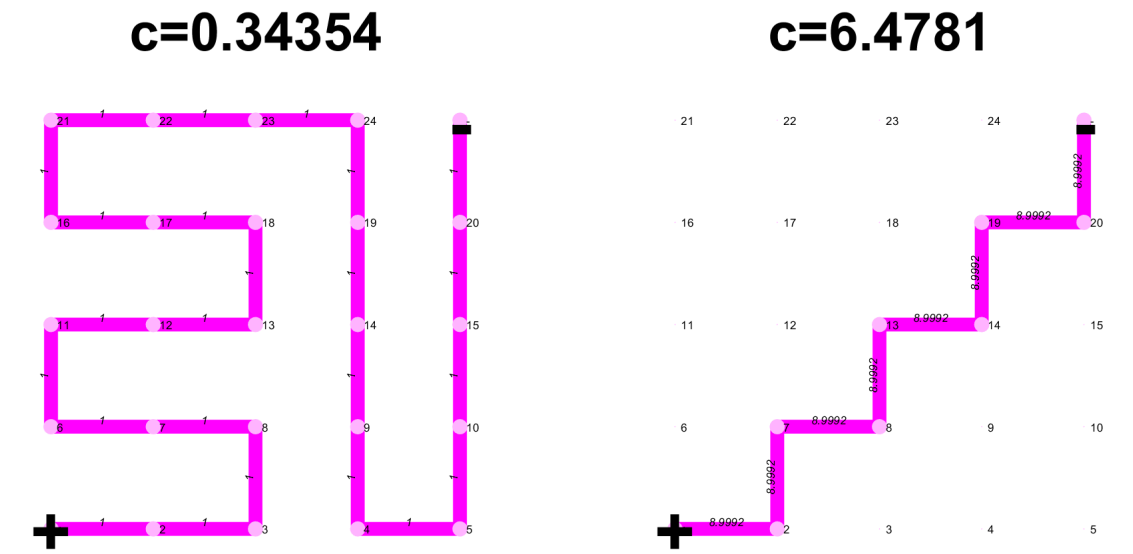


(a) (b)

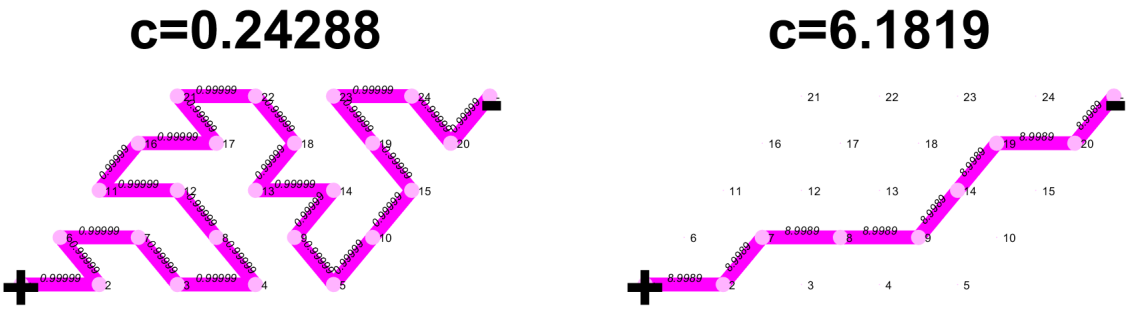
Figure 3.2.1: The plot of sample local optima from optimizing the cost of mixing and dissipation on the (a) square grid and (b) triangular grid. The blue points are the sample local optima, putative global optima are shown as magenta points, the red vertical lines represent the points c_{nm} and the corresponding costs of mixing and dissipation. The grey line is the minimal cost curve of the cost function on the optimal networks (See Chapter 6 Section 6.8 Definition 6.1).

Figures 3.2.1 (a) and (b) both seem to corroborate the conjecture at the end of the previous section, that the cost function is approximately piecewise affine. The piecewise affine structure is more evident in the graph for the square grid because there approximately half as many possible path lengths connecting source in the bottom left corner to the sink in the top right corner for optimal networks on the square grid (9) as there are on the triangular grid (17) and more linear sub-divisions cause the graph to visually appear smooth. In the next chapter we will see that the sample optimal networks provide even more evidence backing up this claim. As in the 3×3 case, for both ambient grids small values of c produce the longest possible path network, the path connecting all of the 25 points, and large c produce the shortest path connecting the source to the sink with 10 points (figure 3.2.2). With a small number of exceptions the optimal networks for both square and triangular ambient grids have the following patterns. For the square grid, as we increase c the optima are paths of lengths decreasing by 2 at a time–i.e. 25,23,... ,11, and lastly 9. For the triangular grid,

path lengths decrease by 1 so the sequential transitions are 25, 24, ..., 10, and 9 occuring at every corner.



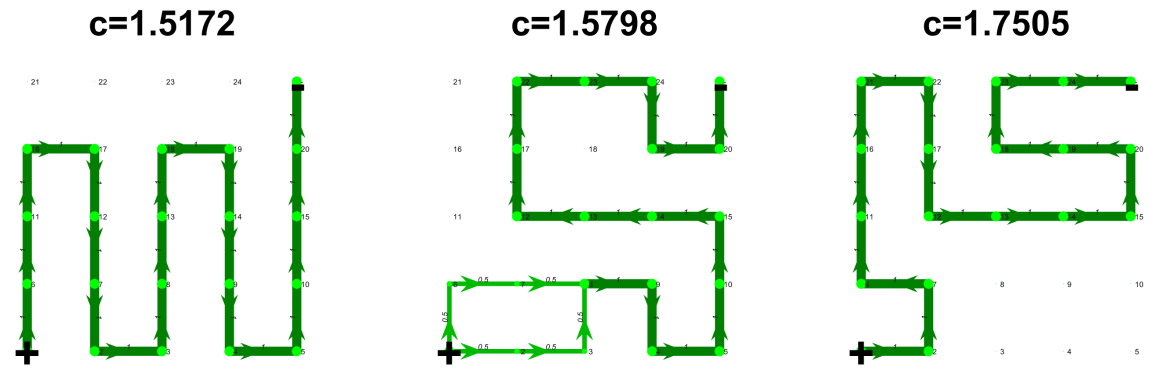
(a)



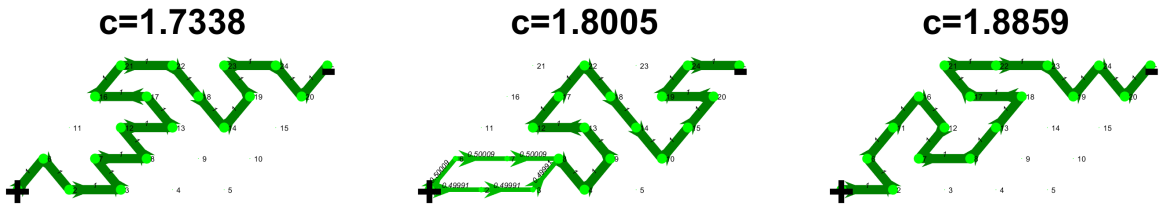
(b)

Figure 3.2.2: Optimal conductance networks chosen from the first interval $[0, c_{25,23}]$ and the last interval $[c_{11,9}, c_{11,9}+1]$ for the square grid (a) and from the first interval $[0, c_{25,24}]$ and the last interval $[c_{10,9}, c_{10,9}+1]$ for the triangular grid (b). The longest path is the optimally mixing network. Both ambient networks contain a path connecting the source to the sink and containing all of the points. The shortest path is the optimally dissipative network.

We find that some of the transitions do not abruptly occur at a point and instead there is an interval where intermediate paths with loops at the source occur. The flow graphs have a bifurcation at the source.



(a)



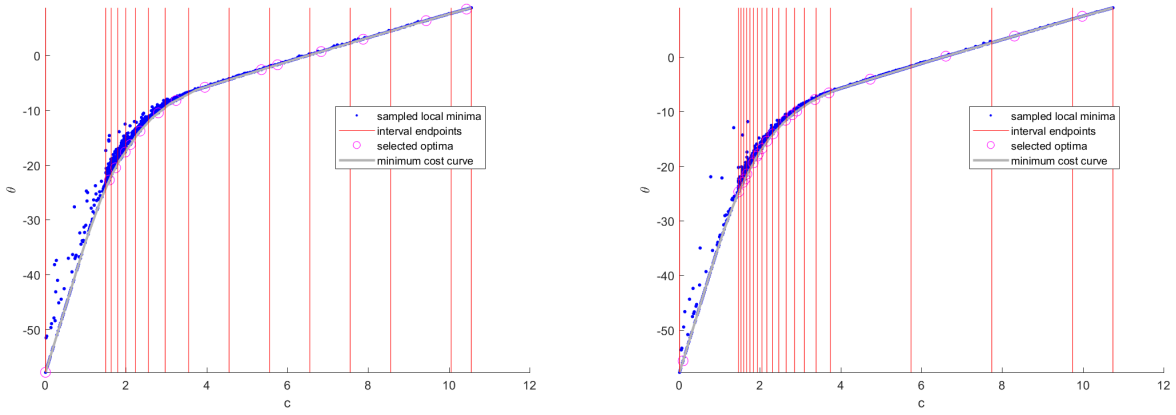
(b)

Figure 3.2.3: Flow networks for putatively globally optimal networks on the 5×5 grid for transitions between path graphs of decrementing lengths for the square ambient grid (a) and the triangular ambient grid (b). For the square grid, the depicted transition is between the paths of length 21 and 19. For the triangular grid the depicted transition is between paths of lengths 19 and 18.

Examples of transitional networks depicted in figure 3.2.3 have the same loop part at the source: an even $\frac{1}{2}$ split of the flow with two nodes on either side, followed by recombination of the flows to feed into a single path that leads to the sink. The geometric arrangement of loop and path parts differ, however. Therefore the changes in optimal mixing-dissipation networks has a more complicated explanation than transitions between paths of gradually decreasing lengths. Still, as we will see in Chapter 4 whose results are highlighted in the next section, path networks of gradually decreasing length form a skeleton of the optimal networks in the sense that even when the optimal networks contain loops, the mean path length from source to sink following flows is decreasing as c increases.

3.3 Numerical results for 1 source and 1 sink on ambient grid networks of side-length 5 and $\gamma = .8$: networks with many loops

We optimize networks obeying the building cost constraint $\sum_{(i,j)} \kappa_{ij}^{.8} = 24$ for the cost of mixing and dissipation for many closely spaced values of c greater than or equal to 0.1 so that the longest path is obtained as an optima, with the upper bound on c. We start with c small enough that we obtain the tour as an optimum and end with c large enough that we obtain the geodesic as an optimum.



(a) (b)

Figure 3.3.1: The plot of sample local optima from optimizing the cost of mixing and dissipation on the (a) square grid and (b) triangular grid. The blue points are the sample local optima, putative global optima are circled in magenta, the red lines represent the points c_{nm} and the corresponding costs of mixing and dissipation. The grey line is the minimal cost function

Note, these graphs do not reveal a piece-wise affine function with many pieces for networks of different lengths and transitions with loops. Instead, our algorithm locates only two path networks, the tour on 25 nodes and the geodesic between the source and sink (Fig. 3.2.2). As c increases, the optimal networks transition through many interesting geometries with loops. All of these globally optimal networks share the structural commonality of being formed from a subnetwork containing loops, connected to a path, with the subnetwork with

loops being located directly at the source and connected through one vertex to the path that travels to the source. In other words, the flow fans out in a bifurcation at the source and is collected at a node and sent through a bottleneck to the sink.

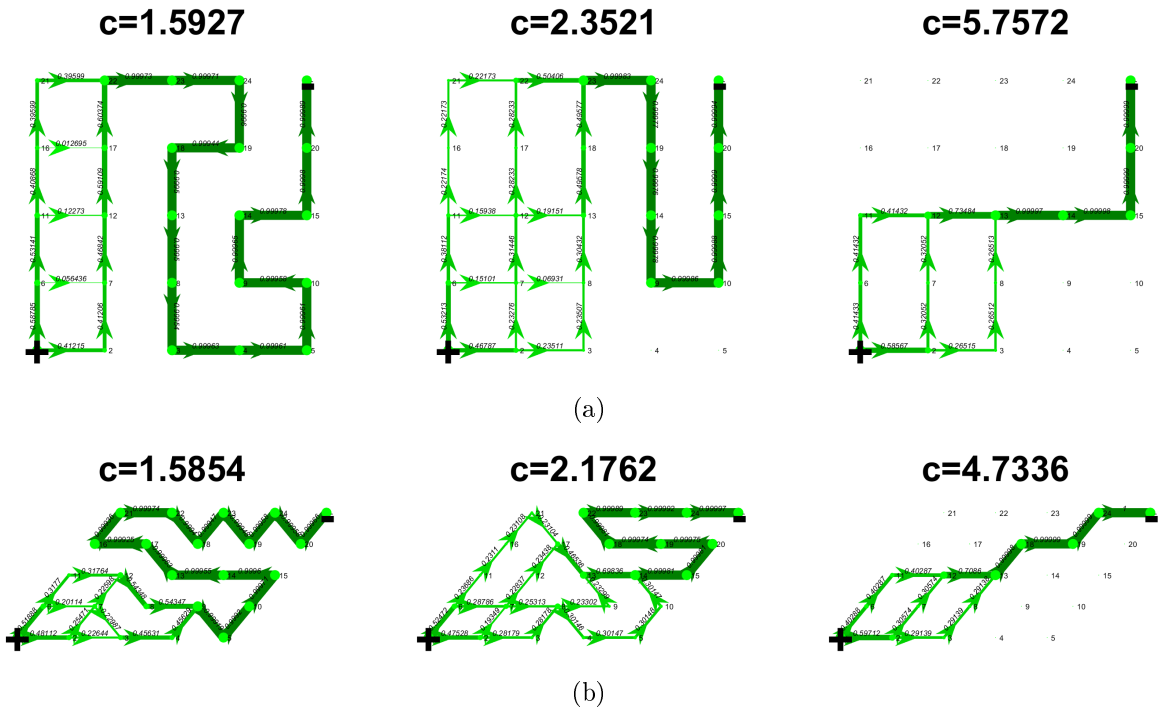


Figure 3.3.2: Three optimal networks with $\gamma = .8$ on the ambient square grid (a) and for triangular grid (b) for c increasing from left to right.

The three networks for both square and triangular ambient grids in figure 3.3.2 outline two general trends in the geometry of the globally optimal path for increasing c; a general decrease in the lengths of paths from source to sink, and an approximately uni-modal rise and then fall in the number of independent loops (Fig. 3.3.3).

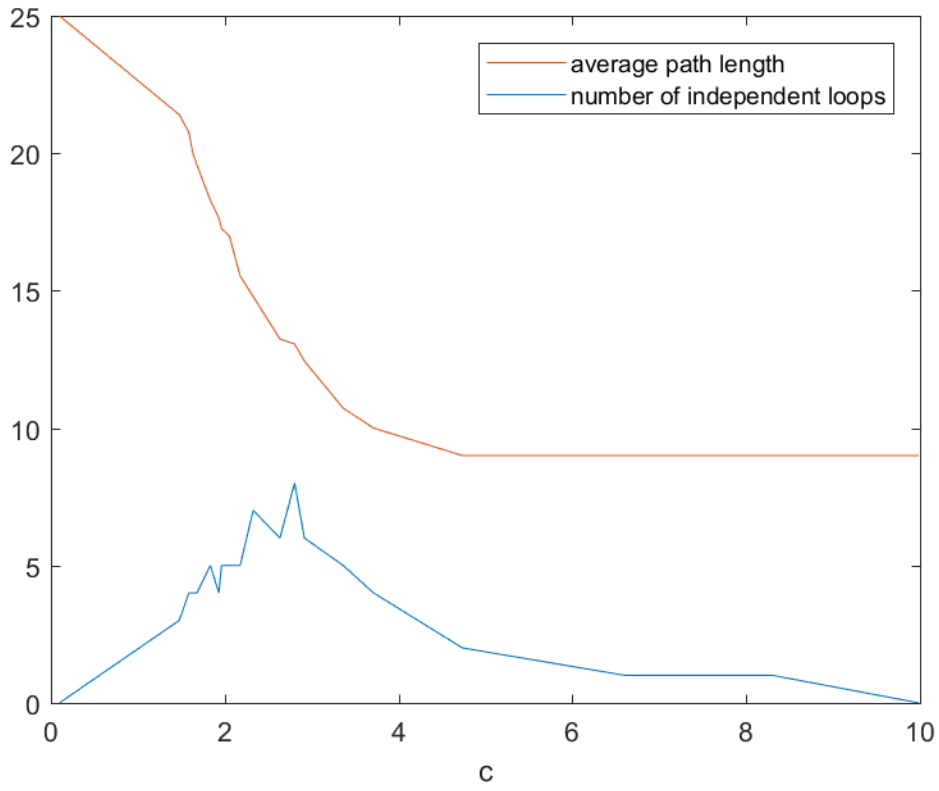


Figure 3.3.3: Plot of the average path length and the number of independent loops for optimal networks on the triangular grid with $\gamma = .8$.

CHAPTER 4

Numerical Results

4.1 Visualization of networks

We use three methods to visualize fluid flow networksgraphs with the weight (thickness) of an edge being proportional to its conductance (κ_{ij}). We visualize the conductance networks as magenta. We visualize the flow networks green graphs with edge thicknesses proportional to by q_{ij}. We also visualize the distribution of signals at each node i by plotting a bar graph of the un-normalized $\tilde{q}_{ji}$ at each node i where the bars are sorted in order of decreasing magnitude. We make a bar graph of the flows $\tilde{q}_{ji}$, $j = 1, 2, ..., N$, rather than of the probabilities that make up $\mathcal{P}_i$. Our bar graphs must therefore be normalized in order to calculate the probabilities within the mixing entropy. But presenting the unnormalized flows has the advantage of also communicating the total flow passing through i because $\tilde{q}_{ii} = f_i$.

In each depiction the sources are labled with "+" and the sinks are labeled with "−".

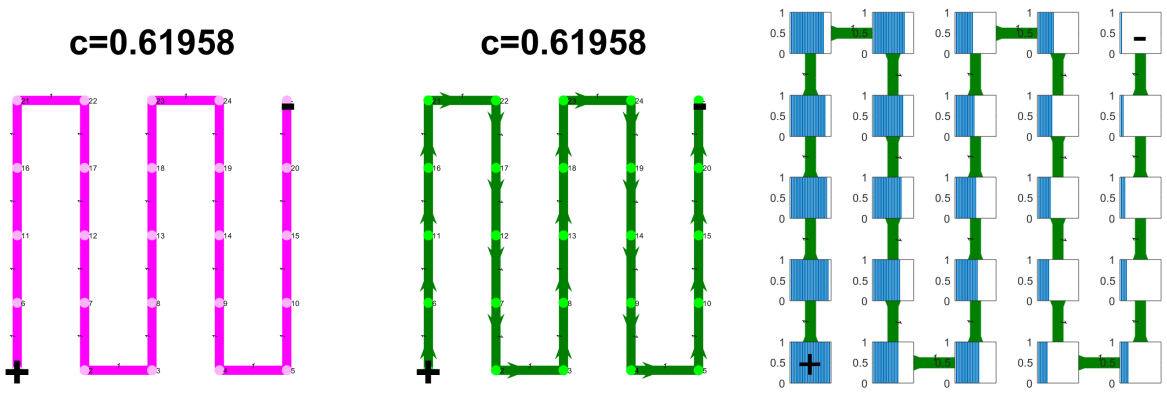


Figure 4.1.1: Example of a conductance network (left), its physical flow network (center) and flows with distribution of $\tilde{q}_{ij}$ on the right.

4.2 One source and one sink at the bottom left and top right corner of square and triangular ambient grids side length 5 with $\gamma = .5$ and $C = 24$.

4.2.1 Square ambient grid

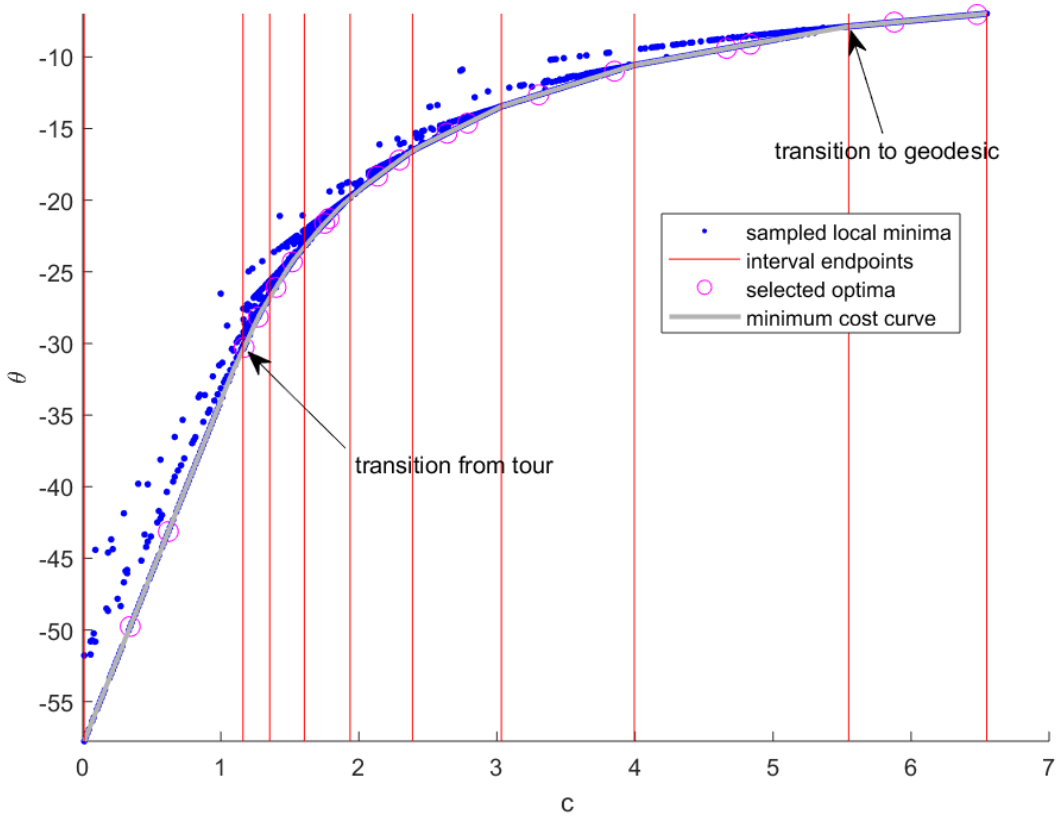


Figure 4.2.1: Numerical optima of NME$+cD$ for networks minimizing the cost of mixing and dissipation (blue points). In grey is the minimal cost curve. The vertical red lines represent the end points of subintervals in which $\theta(c)$ is affine in c. We refer to these intervals as **search intervals.** Magenta points give the c and NME $+ cD$ globally optimal networks – i.e. networks with minimum distance from the minimal cost curve. Two globally optimal networks are chosen for each search interva.. The samples are for a densely spaced set of c between .01 and 6.5481. See Chapter 6 Section ? for technical details. The minimal cost curve is piecewise linear. Transition points where mixing-dissipation cost minimizers transition from the tour and transition to the geodesic are indicated with arrows. These were computed using the optimal networks in Fig. 4.2.6 by taking the first transition point to be the maximum c where the average path length is 25 and the minimum c where the average path length is 9.

We plot the results of repeating our numerical experiment for thousands of times in figure 4.2.1. We observe that the minimal cost curves contains sharp non-differentiable transitions. The numerical optimization method's rendition of these optimal networks is interesting: As a transition is approached from the left with smaller values of c, the numerical optimization irregularly converges in some places to the true optimal network, and in others to the networks that will appear after the transition. We divide the interval of sampled c into 9 search intervals. For each of the 9 search intervals we divided it in half by length to create 2 subintervals. In both of these subintervals, we choose the numerically obtained optimal network whose value of $\mathrm{NME}(\kappa_{ij})+cD(\kappa_{ij})$ is closest to the minimal cost curve. This is what we mean by choosing 2 globally optimal networks in each search interval. We repeated this for each search interval, obtaining 18 globally optimal networks. These sampled networks are shown as the magenta points in Figure 0.2.1, and their conductances and flows are shown in figures 4.2.2 and 4.2.3.

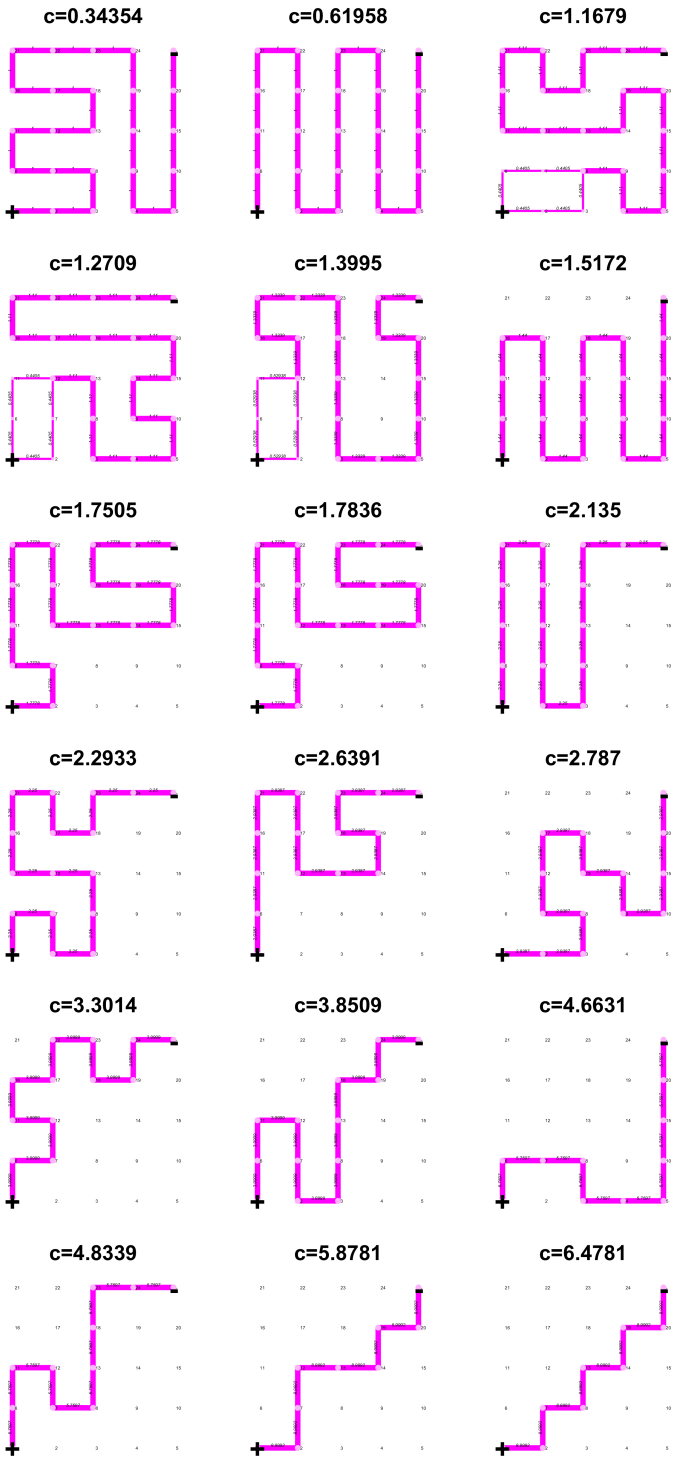


Figure 4.2.2: Selected optimal conductance networks by increasing c from right to left and top to bottom Networks correspond to the magenta points in Figure 4.2.1.

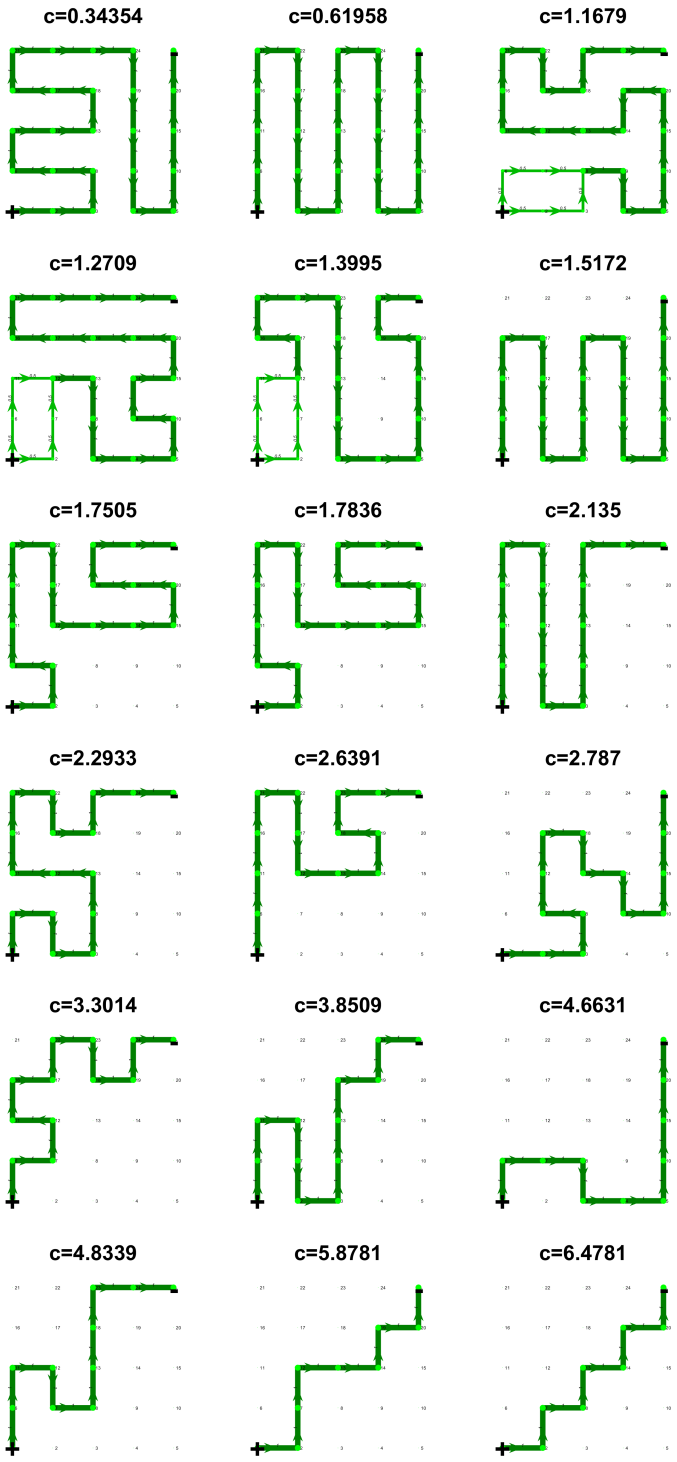


Figure 4.2.3: Selected optimal flow networks by increasing c from right to left and top to bottom.

The network chosen within the first interval is the path network τ_{25} that visits every node in the ambient network. In this first interval $\theta(c)$ is a linear function of c, reflecting the fact that although a single network is selected over the entire interval, the cost of this network is linear in c: i.e. it is equal to $c \mapsto \mathrm{NME}(\kappa_{ij}) + cD(\kappa_{ij})$. In Chapter 5 we will give a rigorous proof that in the limit as $c \to 0$, the optimal network is any path that visits all nodes. Our numerical results suggest that this is also true at small but non-zero values of c. To test this conjecture we look at 5 numerically generated optima for $c = .01$ in figure 4.2.4.

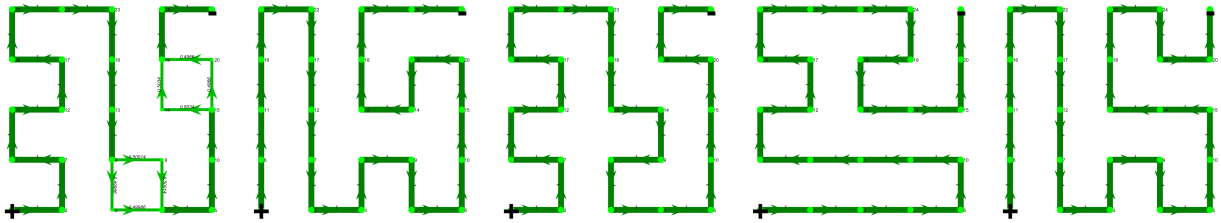


Figure 4.2.4: Results of the optimization for $c = .01$ repeated 5 independent times from random initial conditions. The flow network pictured first on the left has a NME of -51.9897. The four remaining networks each have an NME of -58.0036. The cost function NME + cD on each network gives us -51.7772 for the first network and -57.7636 for the other 4.

The largest value of c that we assay numerically is 6.48, at this value (in fact over the entire subinterval [.01, 6.5481]) the numerically computed optimal network is also path: τ_9, which is the shortest path (geodesic) linking source and sink nodes. This is τ_9. In the limit as $c \to \infty$, optimizing theta is equivalent to minimizing dissipation, and it is known that the unique optimizer here is the geodesic (see [CR17] and our proof in Chapter 5). Our numerical results show that the geodesic is in fact the optimizer even at finite values of c, in fact for $c \geq 3.9955$.

The transition to the geodesic (shown by the black arrow in Fig. 4.2.1 [e.g.]) is representative of a set of transitions that are seen as c is varied, so we will describe in in some detail. The network switches sharply to the geodesic topology at $c = 3.9955$. We infer that thereis not a continuously varying collection of networks with different topologies approaching the geodesic. In the graph of the minimal value of theta (Fig. 4.2.1) this sharp

transition is signalled by a non-differentiable final corner.

We now turn our attention to intermediate values of c, here the optimal $\theta(c)$ curve appears to be piecewise affine, that is over finite subintervals of c values, the optimization algorithm consistently finds the same network (as c varies $\theta(c)$ then varies linearly). What are these optimal networks? From Fig. 4.2.3 we see that the optimal networks are, for the most part, path networks. As we go from one c subinterval to the next one, starting with the smallest value of c, the length of the path in the optimal network decreases by 2. Every possible path length shows up as the optimal network for some range of c values. However, not all optimal networks are paths: The only non-path graphs identified by the numerical optimization are networks 3, 4 and 5, which, respectively, are located as optimal networks when c is in the subintervals $[1.1601, 1.3544]$ and $[1.3544, 1.6047]$. These non-path flows are located between the path τ_{25} and the path τ_{21}. In both of these networks there is a single loop: flows bifurcate symmetrically at the source, the divided flows then each pass along a path of 3 edges, until they meet and form a single bottleneck path heading towards the sink. Intriguingly, the average path length of these networks, as defined in Chapter 2 section **??**, are 23 and 21 respectively. Once the critical value $c = 1.655$ is exceeded, the optimal network reverts again to a path (τ_{21}). We therefore have a sequence of networks with arithmetically decreasing average path-lengths, with the path length decreasing stepwise by 2 as c is increased, starting with a path of length 25 and ending at 9. We investigate how well this trend holds by plotting average path-length as it changes with c .

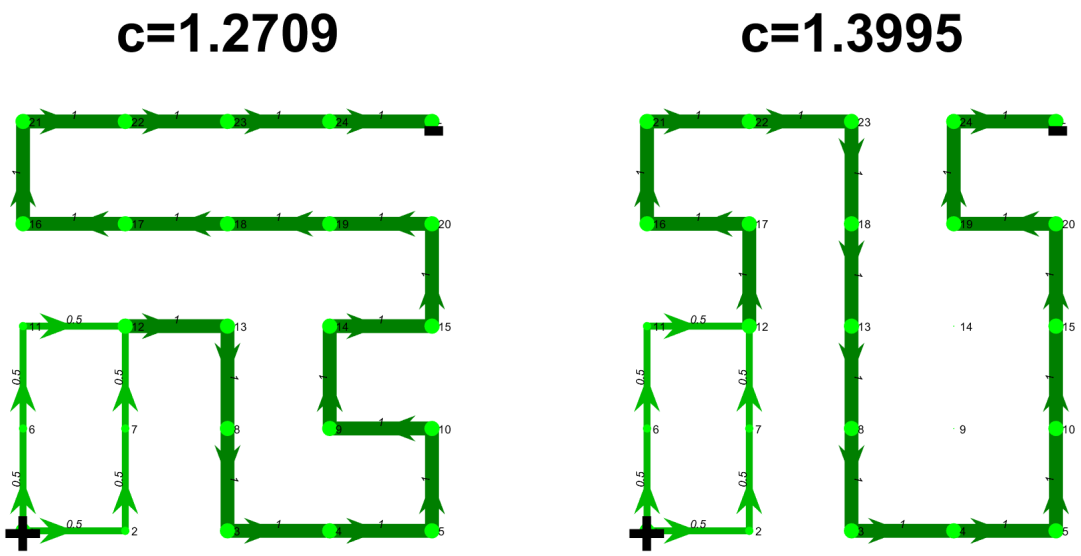


Figure 4.2.5: Networks 4 and 5 from figure 4.2.3

4.2.1.1 Average path length and loops

In figure 4.2.6a. we see a decreasing trend in the average path length for every sample plotted with blue points. Upon refining the sample by only choosing networks which are close to the minimal cost curve, plotted as red circles, a step function with decrements of 2 between each step is revealed. Interestingly, average path lengths of 23 are present in this graph, while our sample of 20 networks did not reveal a path of this length, although there is a network with a loop of average path length 23 (figure 4.2.5). It is remarkable that the experiments suggest that networks optimizing the cost of mixing and dissipation follow a pattern of arithmetically decreasing average path length as the weight is increased. It also seems that path networks are the main networks generating θ.

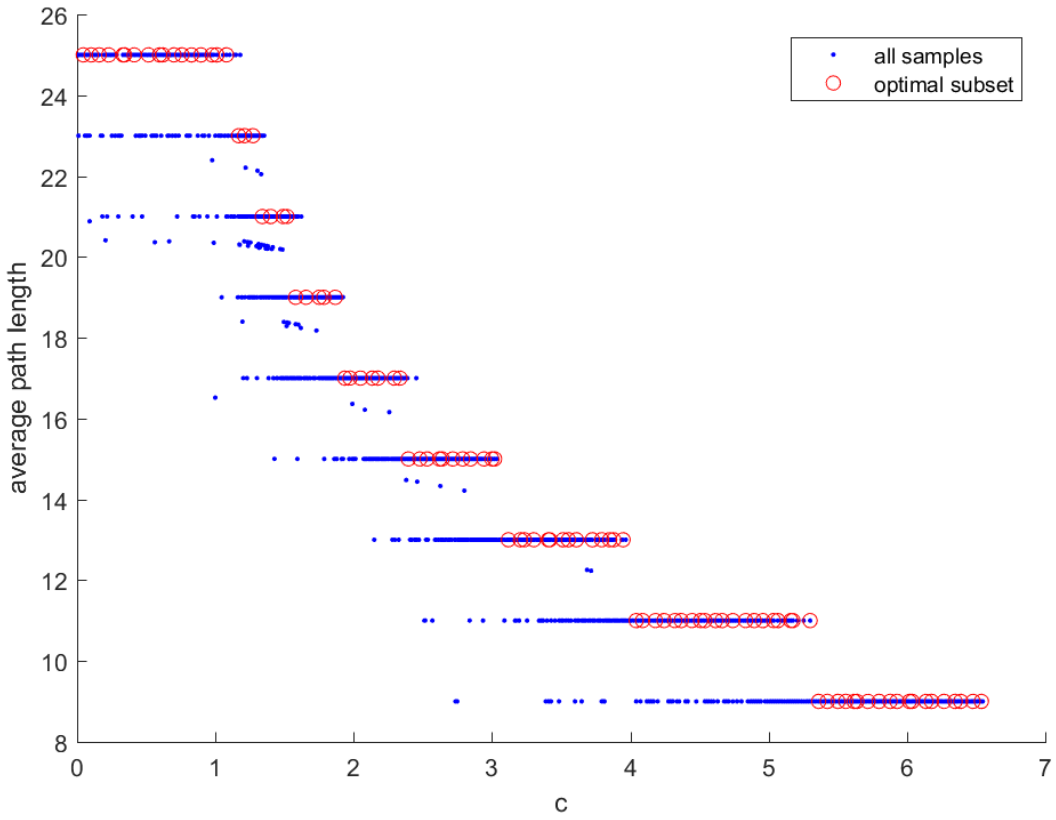


Figure 4.2.6: The average path length for networks optimized at every c plotted as blue points in (a). Each blue point represents a local optimum identified by our algorithm.The red points represent the average path length for 100 globally chosen optimal networks. Globally optimal networkschosen by splitting the entire domain of c into 100 sub-intervals of equal size and choosing the network in each sub-interval for which the difference between $\mathrm{NME}(\kappa_{ij}) + cD(\kappa_{ij})$ and the minimal cost curve is at a minimum.

The exceptions from this rule at path lengths 23 and 21 lead us to wonder about the role loops play. The network of average path length 21 is followed by a path network with 21 nodes, but we have not observed a path of length 23. To investigate the role of loops, we compute the number of independent loops in each conductance network using equation 2.5.1.

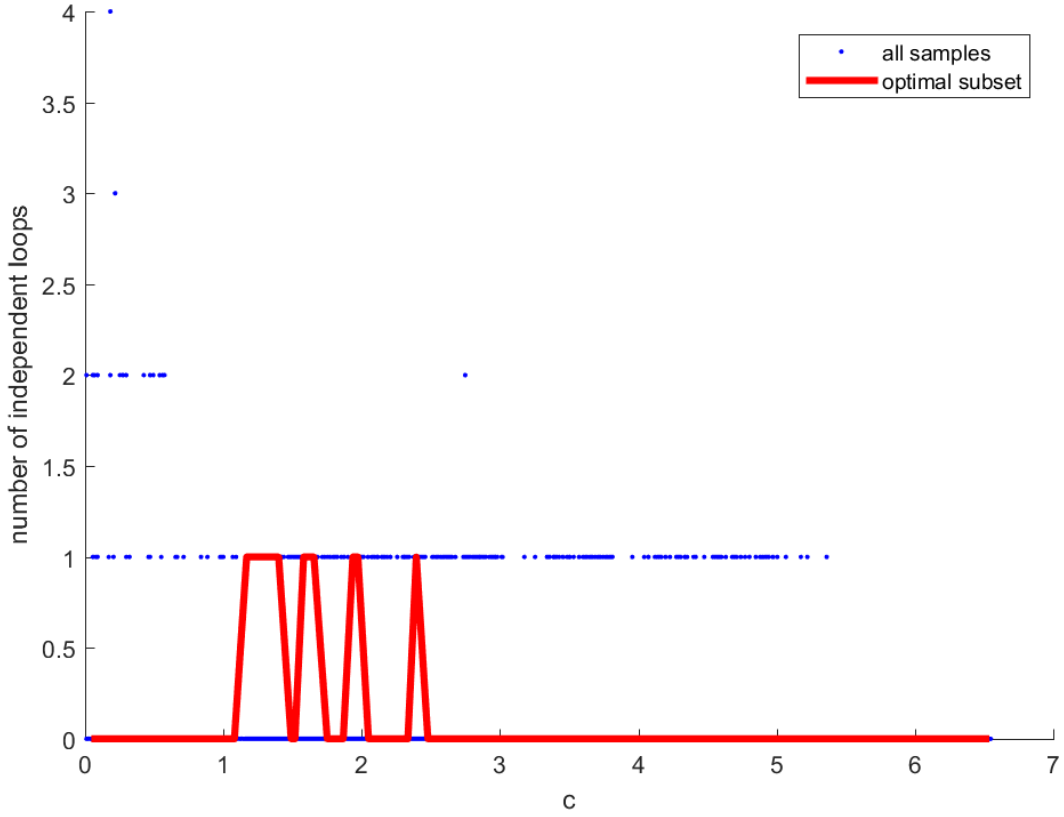


Figure 4.2.7: The number of independent loops for every value of c we sampled is plotted as the blue points. We compute the Betti-number for the same 100 networks as in figure 4.2.6 and connect them with a red line.

Plotting the number of independent loops for each sample shows that there is a maximum of 3 independent loops in any locally optimal network identified by our algorithm. Among networks in our collection of 100 globally optimal networks no more than 1 independent loop is seen. For the most part, it indicates that networks that minimize mixing-dissipation cost are paths. The networks with one loop appear to occur in 4 intervals of decreasing width and are separated by intervals where networks are paths. Figure 4.2.7 shows the complete sequence of globally optimal networks that we see as c is swept the entire interval of values that produce loops. In the figure, we remove topologically equivalent networks (e.g. we show only one representative path of each length produced by our algorithm). At the point where

the paths first join, the probabilities $\mathcal{P}_i(j)$ are equal to $\frac{1}{2}$ from each node on the two sides of the loop and equal to 1 from the source node and from the node itself. The mixing entropy for this node is not far from that for a path containing the same number of nodes.

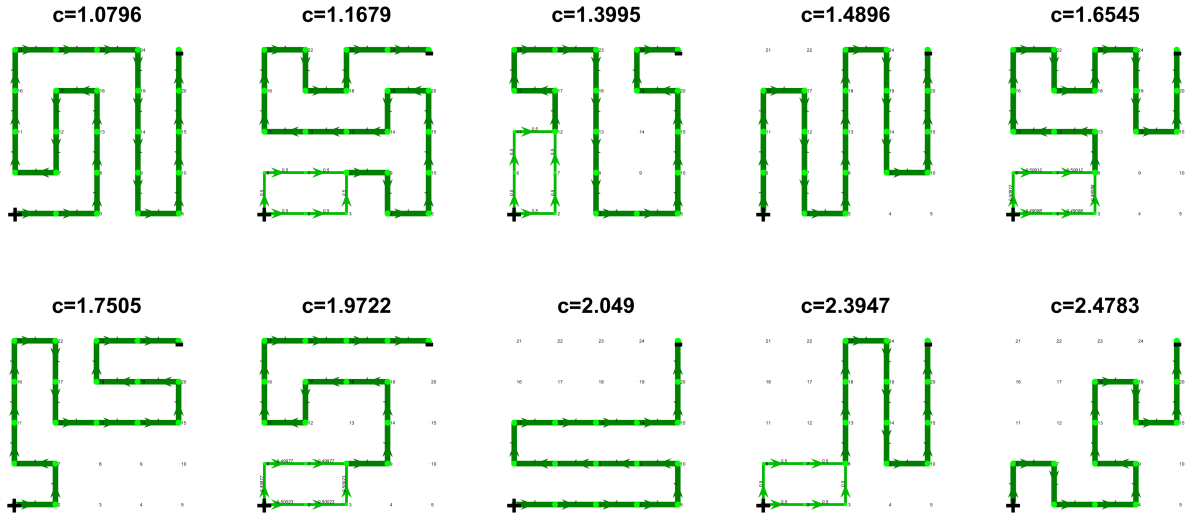
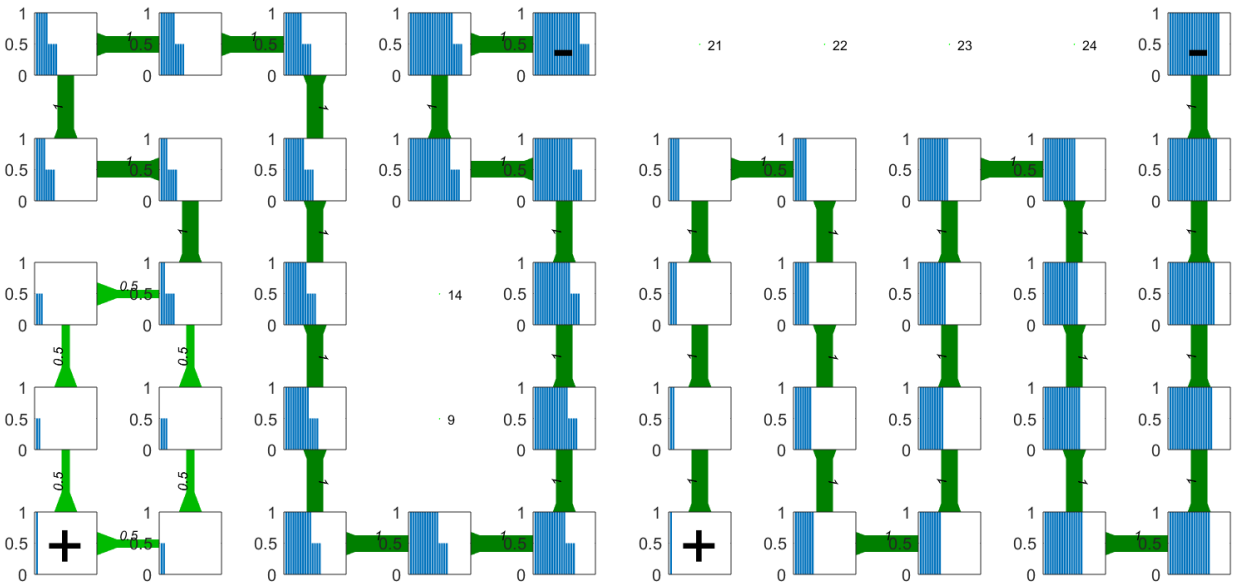


Figure 4.2.8: Networks chosen from the set of100 networks (see figure 4.2.7) representing each connected interval where the Betti-number is constant (either 0 or 1) ordered by increasing c. The third and fourth networks are the two different topologies from the first interval where there is 1 loop. The average path lengths of the networks in the order they appear are are 25,23,21,21,19,19,17,17,15 and 15. The networks for $c \geq 1.168$ in this selection are path graphs of arithmetically decreasing length and have already been discussed.

From the 100 networks sampled in figure 4.2.7, we did not find a path of length 23. For each of the average path lengths 21, 19, 17 and 15, there are two globally optimal networks, one with a loop and one which is a path ($c = 1.3995, 1.6545, 1.9722, 2.3947$ for loopy networks and $c = 1.4896, 1.7505, 2.049, 2.4783$ for path networks). We can express this observation as the optimal networks are for the most part paths, with networks with one loop occuring occasionally between the transition of two consecutive path lengths. The exception for paths of length 23 could be hinting at a different pattern of behavior if we increased the size of the grid. Perhaps paths no-longer are the main optimal networks as there is more room to add loops. A path of length 23 has only 2 nodes to add to a loop. So far there is strong evidence that τ_{25} is the NME optimum on the 5×5 grid, but if we were to consider a 6×6 grid, there

would be $36 - 25 = 11$ nodes that could be appended to the path of length 25 to form loops and 13 that could be appended to a path of length 23. As the path length is much smaller than the number of available points, networks with loops might be preferred as optima.



(a) $c = 1.3995$,NME =?,D =? (b) $c = 1.5172$,NME =?,D =?

Figure 4.2.9: Flow networks 5 and 6 from Fig 4.2.3 with a bar plot of $\tilde{q}_{ji}$ for each node i. The bars are sorted in decreasing order to aid visual comparison of distributions. The distributions plotted are not the normalized $\mathcal{P}_i(j)$ in order to convey both node strength f_i and distribution shape. The maximum bar height is equal to f_i. Both networks have an average path length of 21.

Notice that the only non-path networks identified by our algorithm have a very stereotyped morphology, consisting of a loop formed by a bifurcation at the source that runs through two paths covering three edges on both sides before they converge into a single path (Fig. 4.2.9). The two sides of each loop therefore each have a uniform distribution of species within the nodes of the loop. The four nodes along the split flows have comined local mixing entropies equal to the 2-nd and 3-rd nodes from the source in the path network because they have the same distributions and are multiplied by total flows $f_i = \frac{1}{2}$ that sum to 1 across the split. For the flow network with the loop, every node downstream from the loop has the

same distribution as in the path, except two signals are divided uniformly into 4 signals. The extra signals contribute a gain to the total mixing entropy.

Why are networks with the loop at the source favored? We will explain this feature of the optimal network rigorously in Chapter 5 Section **??**, but discuss the tradeoffs briefly here: The networks have the same average path length as if one side of the bifurcation was removed and the flow was re-directed down the other side. Anywhere the loop is, the combined local negative mixing entropy NME_i of nodes at the same distance from the bifurcation is the same as one node of the same distance from the source. Therefore, it does not seem that the bifurucation helps to mix the signals from up-stream nodes. On the other hand, it does not increase the path length while introducing a new set of signals, which is good from the point of view of reducing dissipation, and down-stream from the loop will receive more signals than in a path network with the same path length and so the mixing entropies of these nodes will be increased.

4.2.2 Triangular ambient grid

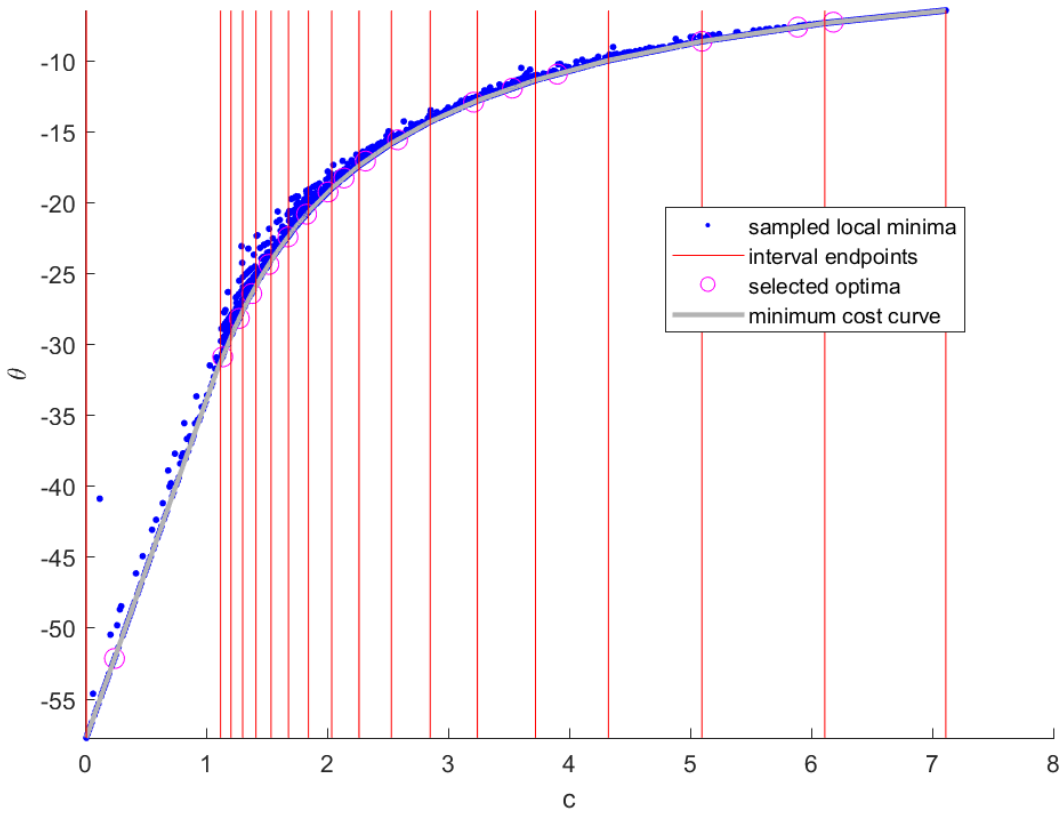


Figure 4.2.10: Numerical optima of NME+cD for networks minimizing the cost of mixing and dissipation (blue points). In grey is the minimal cost curve. The vertical red lines represent the end points of subintervals in which $\theta(c)$ is affine in c. We refer to these intervals as **search intervals** Magenta points give the c and NME $+ cD$ globally optimal networks – i.e. networks with minimum distance from the minimal cost curve

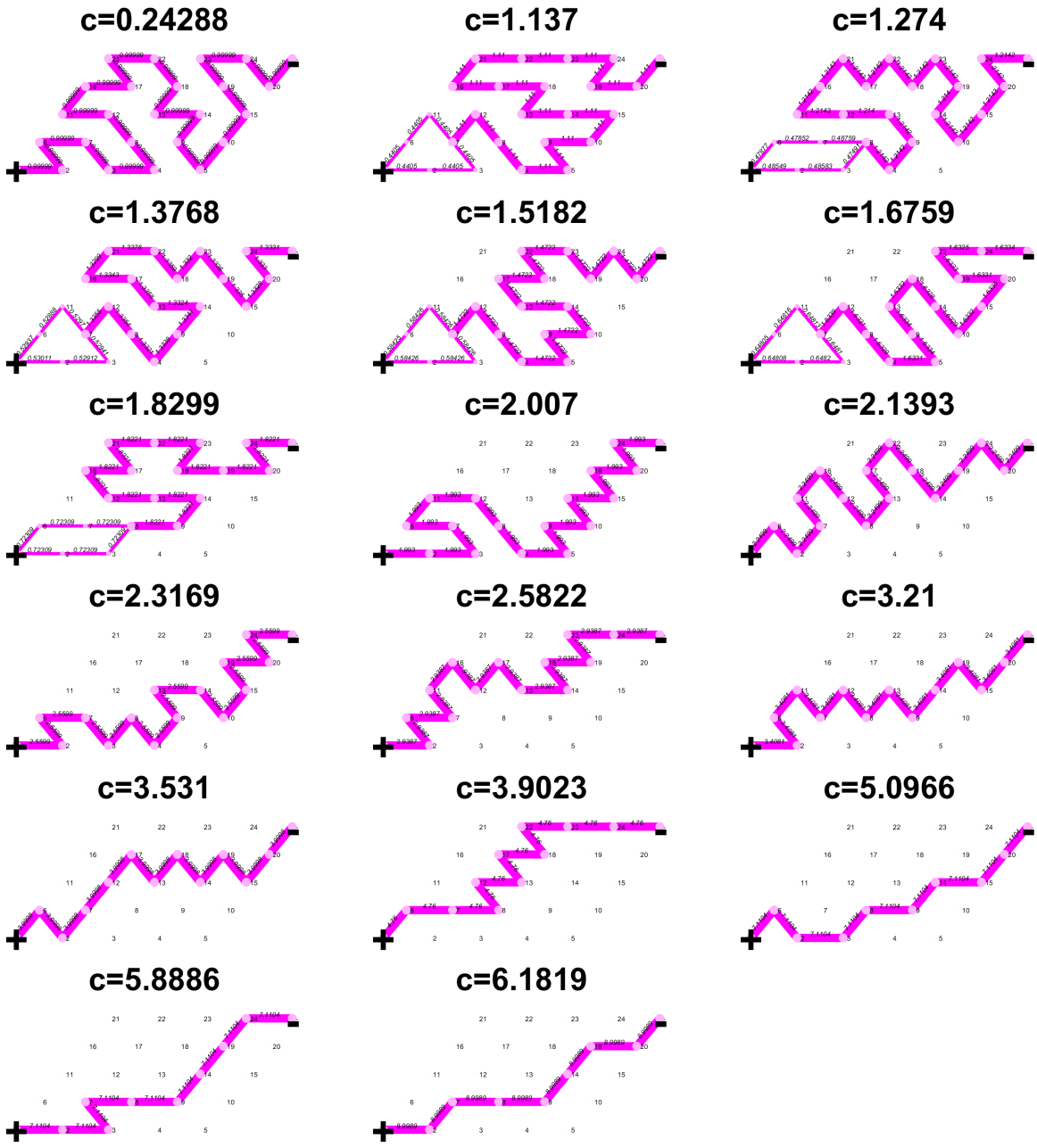


Figure 4.2.11: Selected optimal conductance networks by increasing c from right to left and top to bottom Networks correspond to the magenta points in Figure 4.2.10.

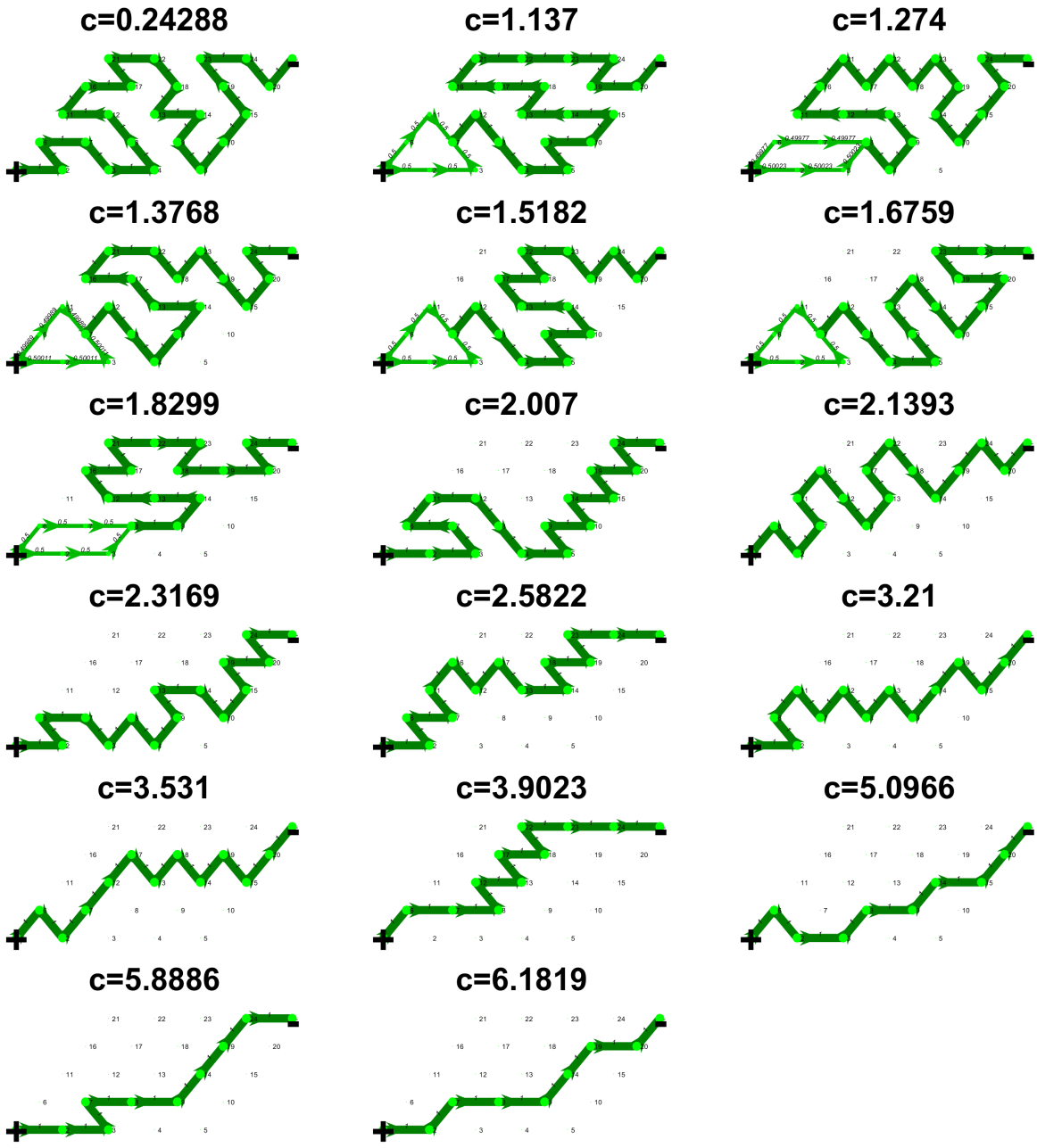


Figure 4.2.12: Selected optimal conductance networks by increasing c from right to left and top to bottom Networks correspond to the magenta points in Figure 4.2.10.

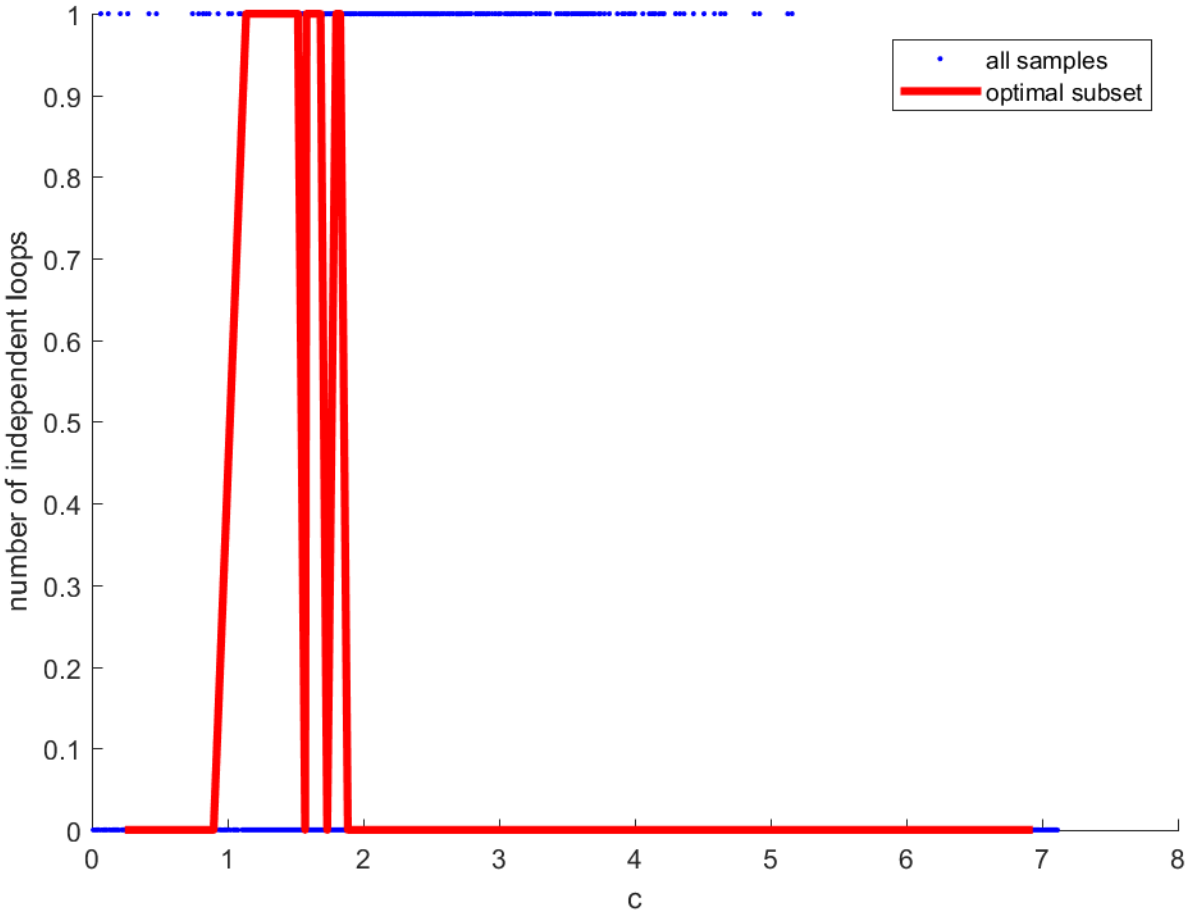


Figure 4.2.13: Number of independent loops for every network sampled. Each search interval is divided into 4 equal parts and the optimal networks are chosen within each search interval and the red line connects corresponding points.

The pattern of independent loops is similar to the square case. Specifically, for $c \geq 1.886$ all of the optimal networks are paths from source to sink, whose length decreases stepwise in increments of 1. Similarly, for $c \leq .8972$, the optimal network is the path τ_{25}, which visits every node in the ambient network. Each globally optimal network remains globally optimal over a finite range of values of c. Over this range the function $\theta(c)$ is linear in c. The argmin network transitions sharply from one network geometry to another at transition values of m (i.e. the minimum curve of $\theta(c)$ is piecewise linear). Networks with loops show

up much more extensively in the square network case. We visualize the Betti number in Fig. 4.2.13. All of the local optimal networks have one or few loops, so the Betti number serves primarily as an indicator variable for the number of loops in the network. We see that loopy networks are selected for a wider range of values of c, and in fact the first morphology that the network adopts after τ_{25}, when c is increased contains a loop.

We visualize representative networks from each of the subintervals in which loops were seen in Fig. 4.2.14. Similarly to networks defined on a square ambient grid, the mean path length from source to sink decreases as c is increased.

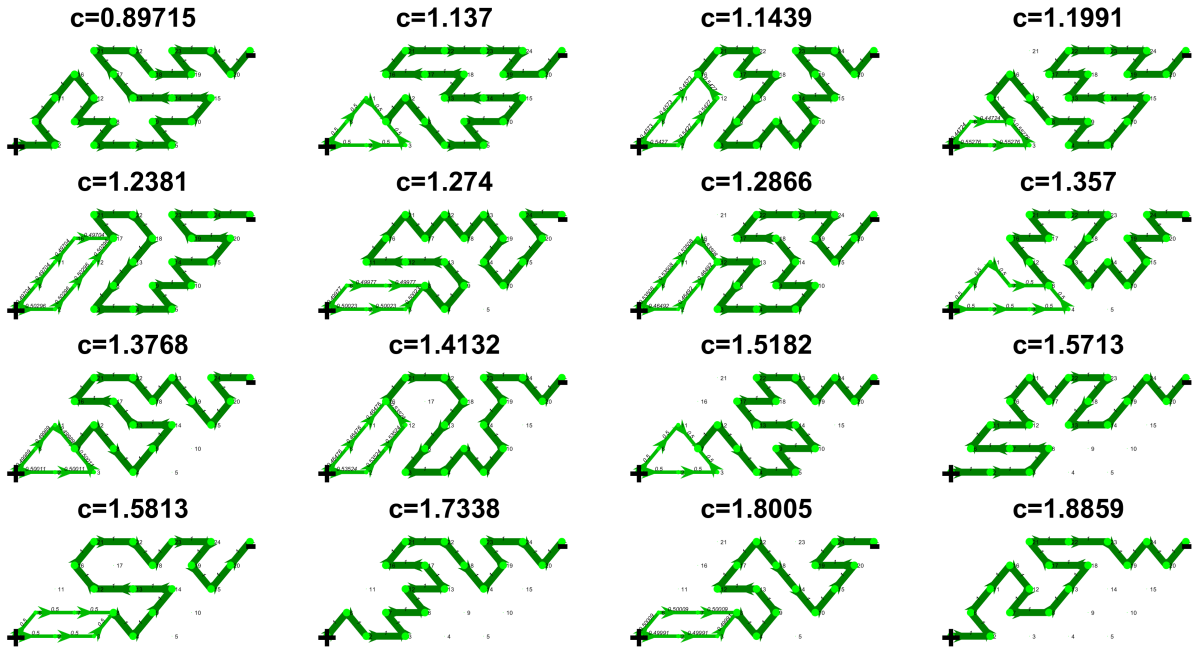


Figure 4.2.14: Networks representing disticinct topologies in the plot of independent loops in figure 4.2.13.

The loopy networks have similar morphologies to the square ambient grid. The loop occurs at the source, and symmetrically divides flow along two equal length paths, before recombining into a single path. However, unlike the square ambient grid, the number of edges on the two sides of the loop is not fixed at 3 - we see loops with sides that are anywhere from two to four edges long. In four cases (e.g. $c = 1.1439, 1.1991, 1.2886, 1.4132$) the number of

edges is not matched on either side of the loop.

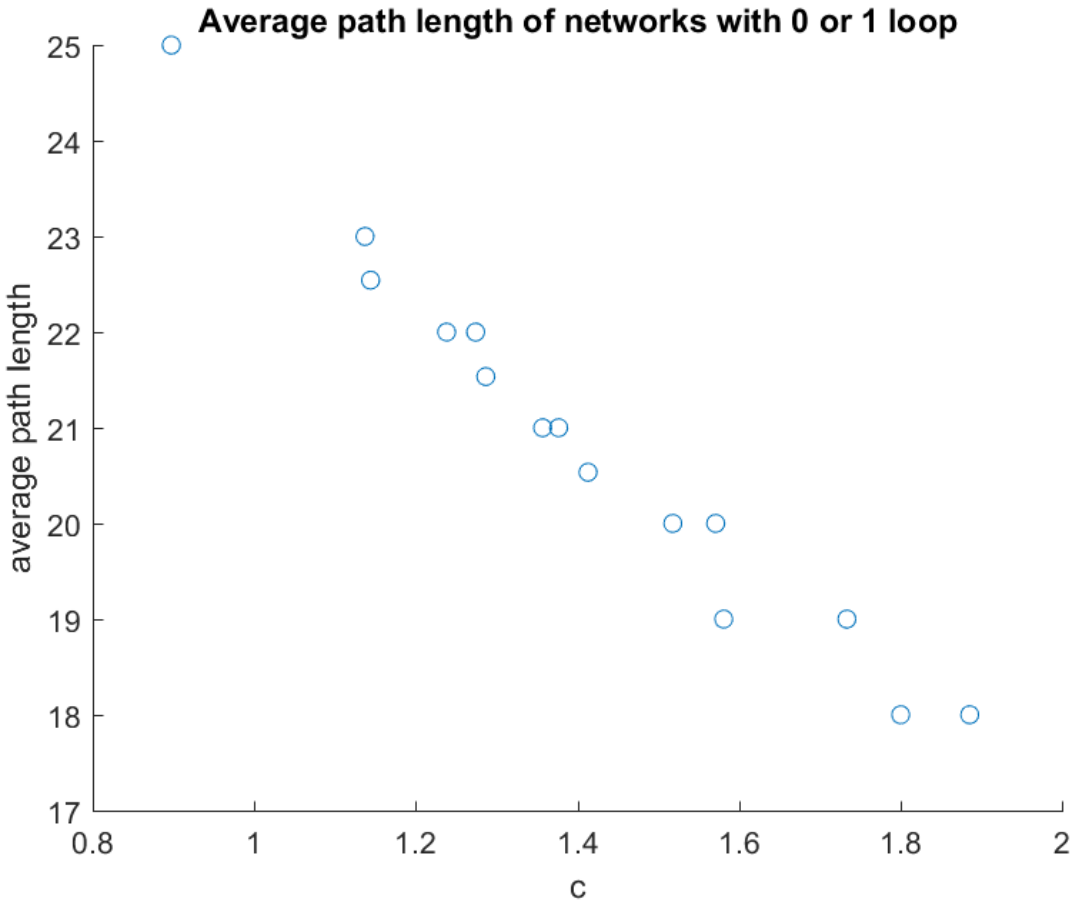


Figure 4.2.15: The average path lengths of networks in figure 4.2.14 as a function of dissipation cost c. Average path lengths trend downwards as c is increased. But unlike the square ambient grid, the path length does not decrease as an arithmetic sequence. There are also multiple networks that have close to identical average path lengths, that appear successively as optima as c is increased. Flow networks are shown for the filled square optima in Fig. 4.2.16. .

The geometric complexity of looping for networks on a triangular ambient grid is also associated with a more complex dependence of mean path length on c among the globally optimal networks. In particular, as c is increased, although mean path length has a decreasing trend, it does not decrease in fixed increments (see Fig. 4.2.15), but rather decreases in steps of 2, 1 and 0.5 in step transitions. Additionally, geometric changes in the network can occur

without any meaningful change in mean path length. For example, FIg. 4.2.16 shows the sequence of globally optimal networks obtained when c is increased from 1.1439 to 1.2381. All three networks (shown as the [e.g.] filled squares in Fig.4.2.15 have mean path lengths close to 22.5, but the loop and path portions of each network have different lengths.

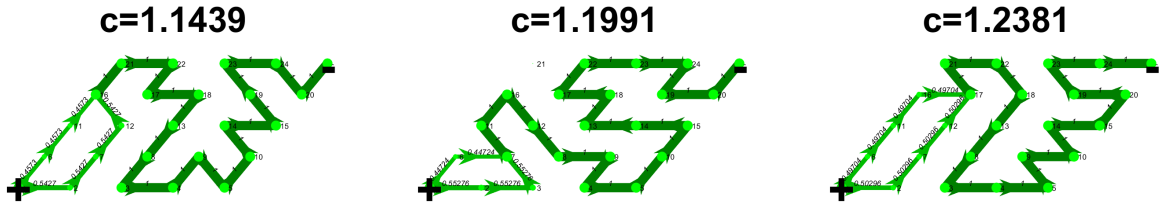


Figure 4.2.16: The middle network is highlighted in orange in figure 4.2.15. The average path lengths of the first two networks are not integer valued. They are 22.5427 and 22.5528 in the order they appear. They have very similar path lengths for how different their topology is. The last network has an average path length of 22. In a triangular grid, the choices of topology can have finer details, and that may lead to some uneven split bifurvations being favorable.

4.3 One source one sink boundary flows with $\gamma = .8$ and $C = 24$

The parameter γ controls the degree of convexity of the cost function; more specifically the material penalty associated with dividing a single conduit into two smaller conduits with the same total conductance. If $\gamma = 1$ then there is no material penalty associated with dividing a conduit. If $\gamma > 1$ then two conduits have lower material cost than a single conduit of the same conductance. Accordingly, networks that minimize dissipation undergo a phase transition from simply connected to loopy geometries if gamma is increased through 1 [BM07]. Under conditions that favor loops (such as when networks must supply sinks with fluctuating strengths, [Cor10]), the number of loops in minimial dissipation networks increases as gamma increased. We wondered whether increasing gamma would have the same effect on networks optimizing mixing, and accordingly we repeated our analysis with $\gamma = 0.8$.

4.3.1 Square ambient grid

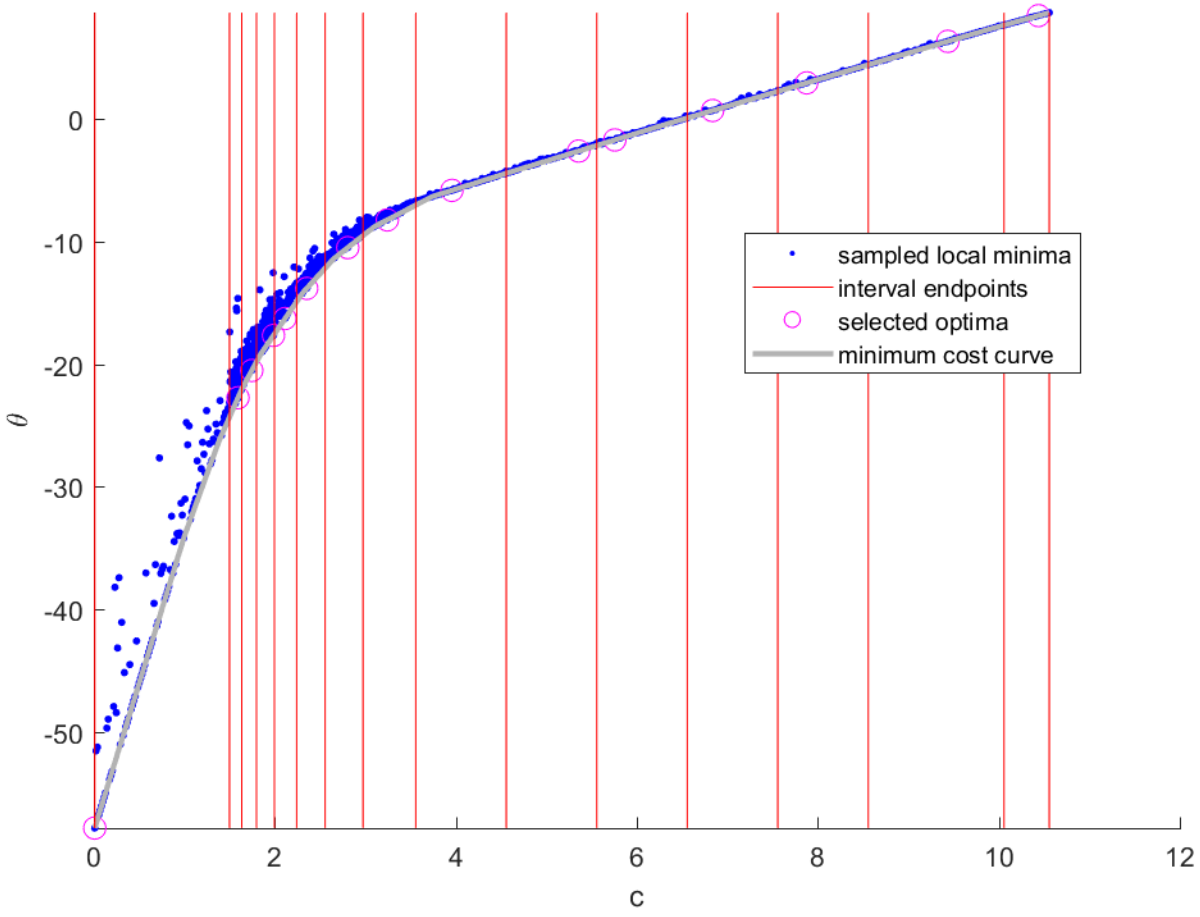


Figure 4.3.1: Scatter plot of NME $+ cD$ for numerically obtained networks that are local optima of the cost of mixing and dissipation (blue points). Grey curve is the minimal cost curve. The vertical red lines represent the end points of subintervals over which we choose globally optimal networks with which have the smallest difference between the mixing-dissipation cost and the minimal cost curve (globally optimal networks; see Chapter 6). We refer to these intervals as **search intervals** c and NME $+ cD$ values for globally optimal networks are given by magenta points.

We plot the mixing-dissipation costs for a collection of networks defined on an ambient grid, for c values that densely cover the interval $c \geq .01$ and $c \leq 10.557$ in Figure 4.3.1. Just as in our $\gamma = 0.5$ simulations, at large and small values of c, the minimal cost curve becomes a

straight line, minimal cost curve appears to be a transition between two affine functions, the first corresponding to a path that visits all nodes the for c small and the geodesic for large c. The interval of transition between these two optimal networks does not have an obvious piecewise linear form as it did in the case where $\gamma = .5$. Visually, the curve seems to almost be smooth. This may be the effect of many different possible transitional networks between the tour and the geodesic.

Figure 4.3.2: Selected optimal conductance networks by increasing c from right to left and top to bottom.

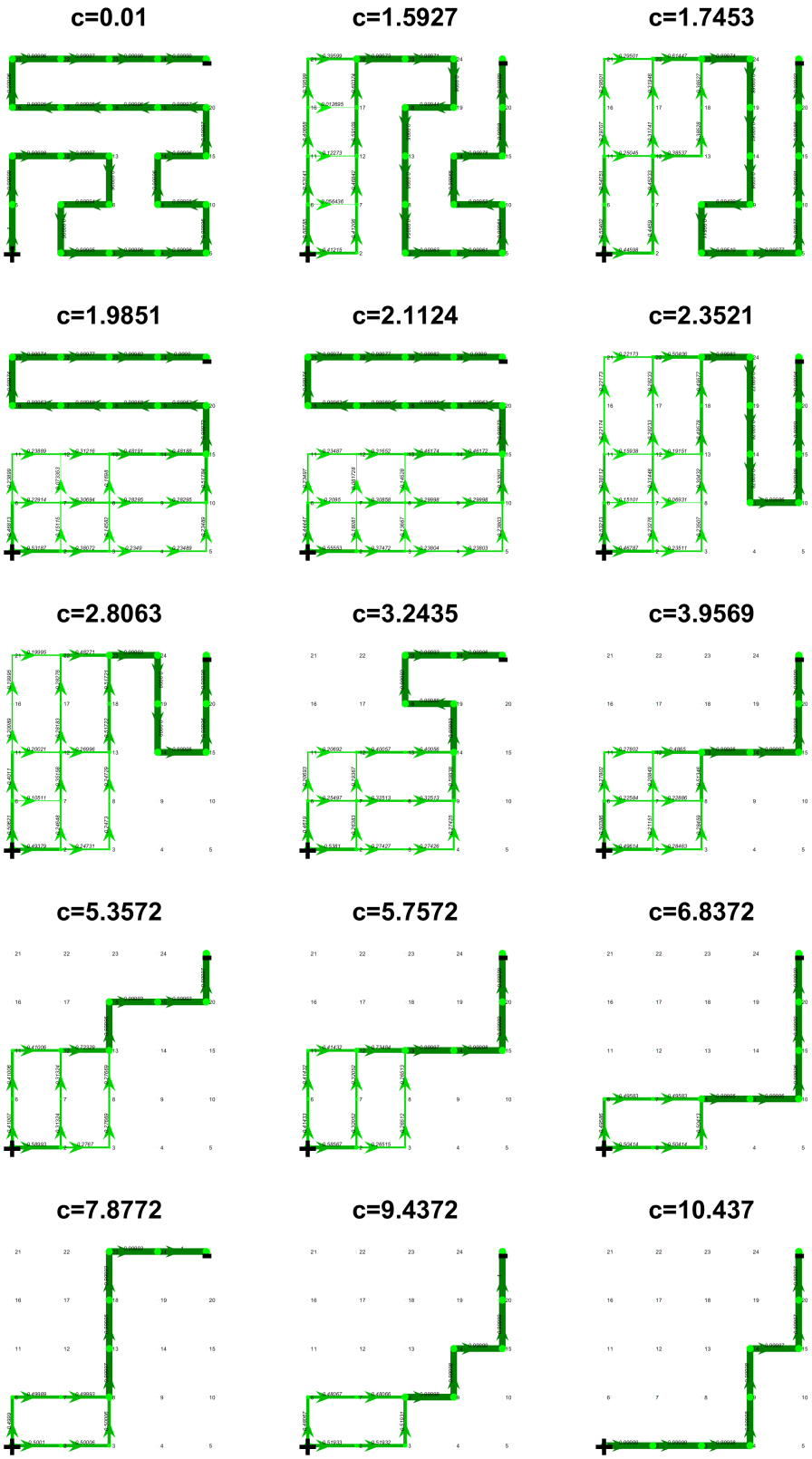


Figure 4.3.3: Selected optimal flow networks by increasing c from right to left and top to bottom.

In figure 4.3.3 the flow networks of the globally optimal selected networks follow patterns that are reminiscent of the case for $\gamma = .5$ but characterizing them completely is beyond the scope of this work. We note, however that each network that has loops, locates them at the source in the following sense. The network bifurcates at the source, and then flows follow reticulated paths through a loop portion. The loops then recombine at a single point (which the call the bottleneck), whereupon all flows pass along a single path to the sink. We call the sub-network consisting of the first bifurcation of the flow of size 1 and ending with a final meet to a flow of size 1 as the loopy sub-network. Even though the connectivity is more complicated than in the $\gamma = 0.5$ case, the loopy sub-network starts at the source, just as the loops with 2 points on either side of the bifurcation started at the source (figure 4.2.5).

The first network at $c = .01$ is the tour. It then appears that for the 2-nd through the 5-th network, the number of points in the network does not change, but the size of the loopy sub-network grows. For networks 6 and 7, the bottleneck decreases by length 2 and for networks 7, 8 and 9 both the bottleneck decreases by 1 and the loopy-subnetwork shrinks in size as well. Between network 9 and 10 we see a loss of the horizontal flowing edges connecting nodes labeled $7, 8$ and 9 in the loopy sub-network. From the 10^{th} network onwards we see a decrease in the loopy sub-network to have only 1 independent loop. The last network is the geodesic.

As the loopy sub-network grows and the number of nodes is fixed at 25, we see a decrease in the average path length. This is because every path in the loopy sub-network is a geodesic between the source and the sink of the loopy sub-network (the bifurcation at the source and the point where all of the flows first recombine). The decrease in the length of the bottleneck, with the decrease of the path length of the loopy subnetwork also brings about a decrease in the length of every path from source to sink. From this we hypothesize that a trend of decreasing average path length for increasing c occurs.

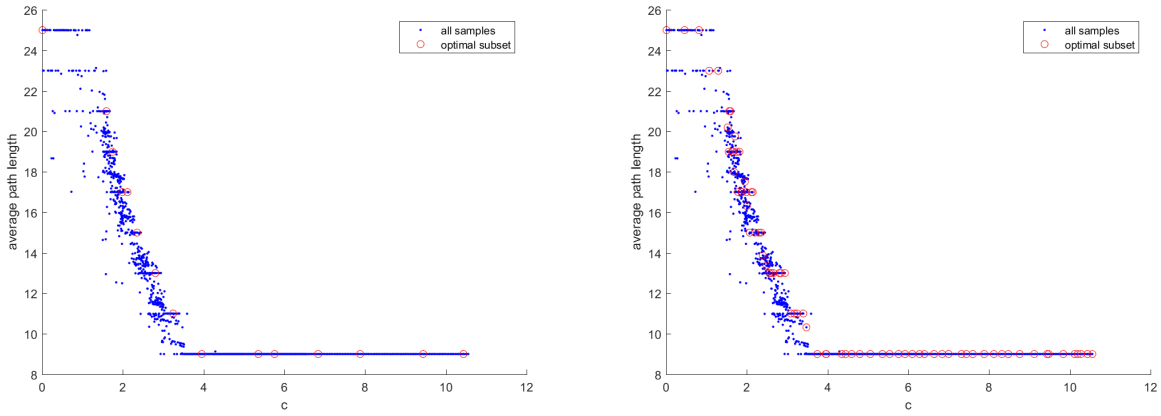


Figure 4.3.4: We plot the average path length of every numerically obtained locally optimal mixing-dissipation network. On the left, globally optimal networks (i.e. the magenta data points from Fig. 4.3.1) are plotted as a red circle. On the right, we increased out sample of optimal networks by creating 5 equal length sub-divisions of each search interval and choosing the network with mixing-dissipation cost closest to the minimum cost curve.

In Fig. 4.3.4, we plot the mean path length from source to sink for all of our locally optimal (blue points) and globally optimal (red points) networks. Locally optimal networks typically have non-integer path lengths, and the mean path length behaves much more like a continuous variable. However, 4.3.4 it appears that the average path length of globally optimal networks is a decreasing step function of c where the steps are spaced by 2. This is reminiscent of the experiment for $\gamma = .5$. To affirm that the step discontinuities in path length are not artifacts of under sampling we re-divided the range of c-values into five times finer subintervals, and plot the globally optimal network in each of these subintervals. The new, more complete set of globally optimal networks still shows a stair-step dependence of path length upon c, again suggesting that the minimal cost curve contains corners at which the optimal network changes discontinuously. Similarly, to the previous case, the numerical optimization continues to find local optima that lie on straight line paths -- i.e. as c increases, it continues to find a now suboptimal network, past the c value at which a shorter network should be selected for.

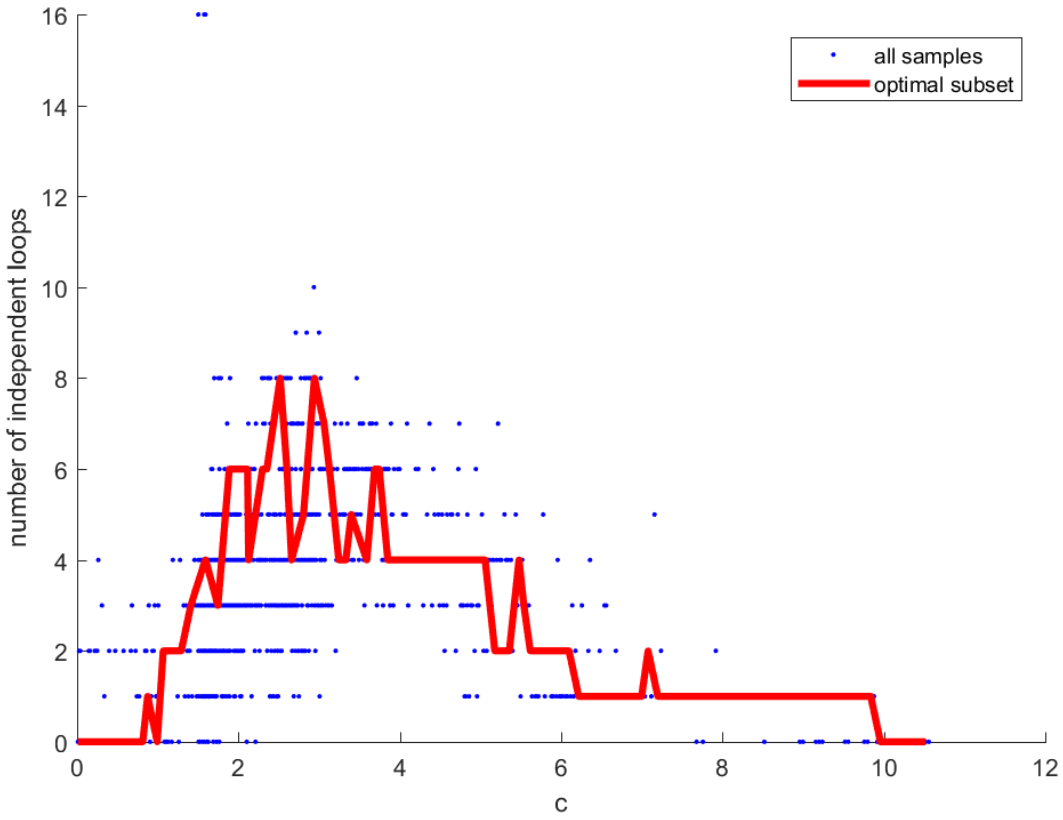
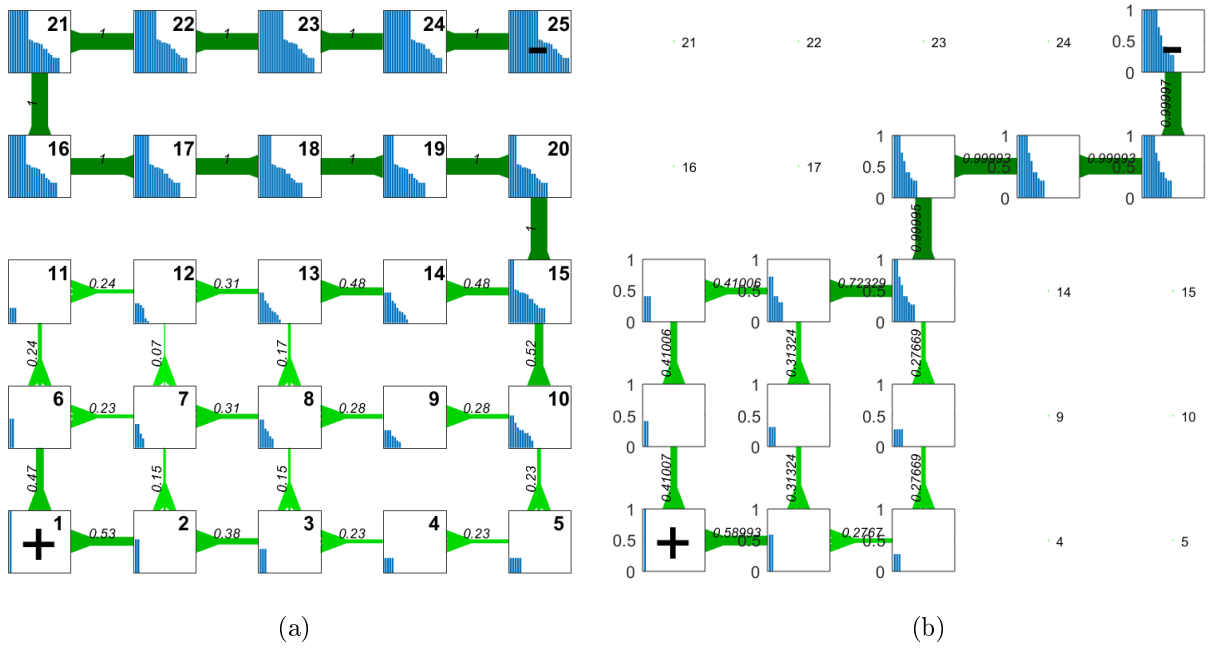


Figure 4.3.5: Plot of the number of independent loops for optimal networks at every c (blue) and the Betti numbers of 100 global optima from equally spaced intervals over the entire domain of c (red). The minimum Betti number is 0, corresponding to path networks, and the maximum globally optimal Betti number is 8. The maximum possible Betti number is 16, the number of independent loops in the ambient grid.

The increased complexity of the loopy subnetwork can be quantified by examining how the Betti number of the optimal networks depends upon c. As c is increased, the Betti-number of networks steeply increases and then slowly tapers off. Our Figure shows that the Betti number stays at 0 (meaning that the optimal network is simply a path), until $c = .8141$. There is a sharp change from the tour to the first network with loops. This is because the 10 first networks that are represented by the horizontal line at 0 with a spike to 1 and return to 0, with exception of the first Betti-number 1, are all tours. At intermediate values of c, the Betti number reaches a much higher value than in the $\gamma = 0.5$ case, and also irregularly

increases and then decreases, reflecting a set of transitions within the loopy subnetwork that involve both gaining and losing loops. The slow decrease shows that eventually there is a trend in simultaneous decreasing average path length and Betti-number.

4.3.2 Signal distributions



(a) (b)

Figure 4.3.6: Visualization of flow networks from the globally optimal networks in Fig. 4.3.3. The bar graphs at each node i depict the distribution of arriving signal strengths $\tilde{q}_{ji}$ with the signals sorted by decreasing magnitude.

In Fig. 4.3.6 we present visualizations of two networks with bar graphs of the arriving signal strengths at each node. Both flow networks have more than one loop. In both networks, the flow from the source has bifurcations at 2 nodes (nodes 1 and 2 in (a) and nodes 1 and 6 in (b)) splitting the total flow into 3 smaller flows. In flow network (a), there are two paths starting at 2 and 3 crossing the split flows. We can see that these paths increase the number of signals arriving at downstream nodes in a way that positively effects mixing. Such cross flows are not present in (b) and so nodes along separate sub-flows only mix signals from their

flow.

4.3.3 Triangular grid

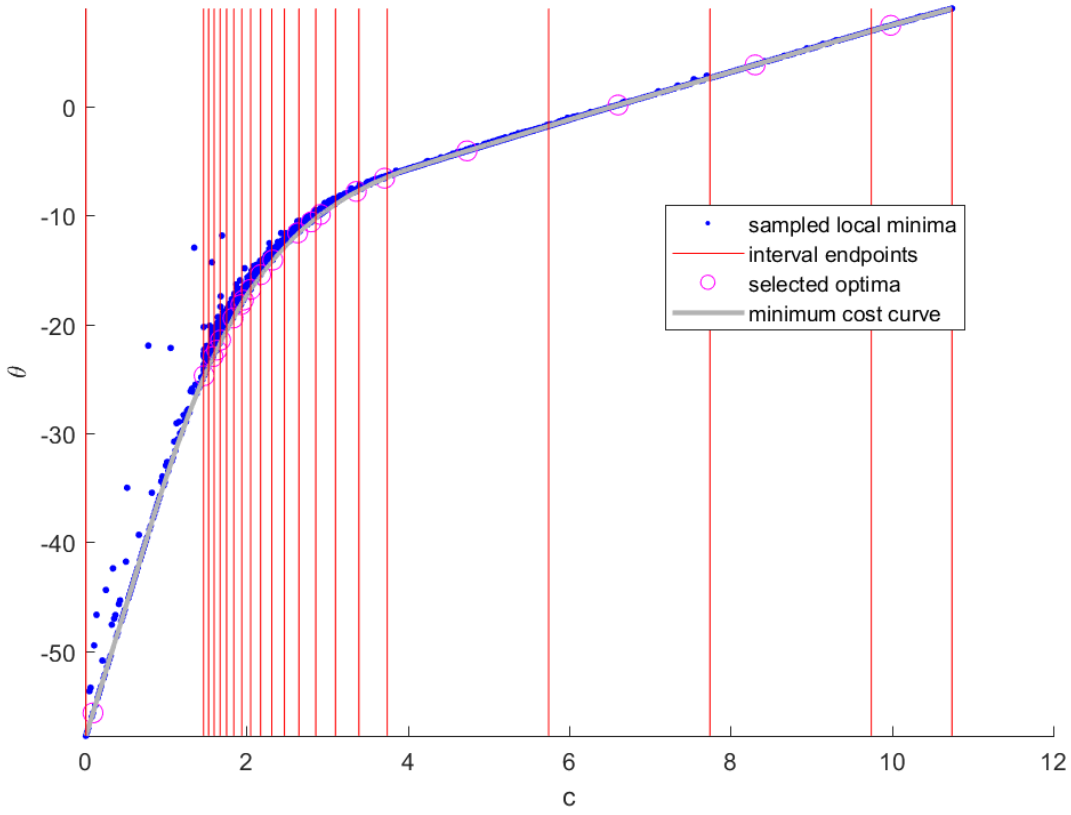


Figure 4.3.7: Scatter plot of the cost function NME + cD for locally optimal networks computed by numerical optimization (blue points). The grey is the minimal cost curve. The vertical red lines represent the end points of intervals over which we choose globally optimal networks, as networks that minimize the distance from the minimal cost curve (See Chapter 6). We refer to these intervals as **search intervals** The c and NME + cD values for globally optimal networks are shown as magenta points.

The ambient triangular grid contains a larger number of edges, and we find a larger number of local and global optima. Indeed, our method for identifying search intervals immediately produces a much larger number of search intervals (17, as compared to 9 for the square grid with the same number of nodes). Except for the first and last subintervals of c, in which

the optimal network is respectively a tour and a geodesic, the minimal cost function curve no longer has a clear piecewise linear structure. We hypothesize that there is continuous interpolation between global optima, rather than, as we previously encountered, a discrete set of global optima between which the optimal network moves by step-wise transitions. To study these transitions we subdivide the total c range more finely – see Chapter 6 for more details as how we pick this exact subdivision.

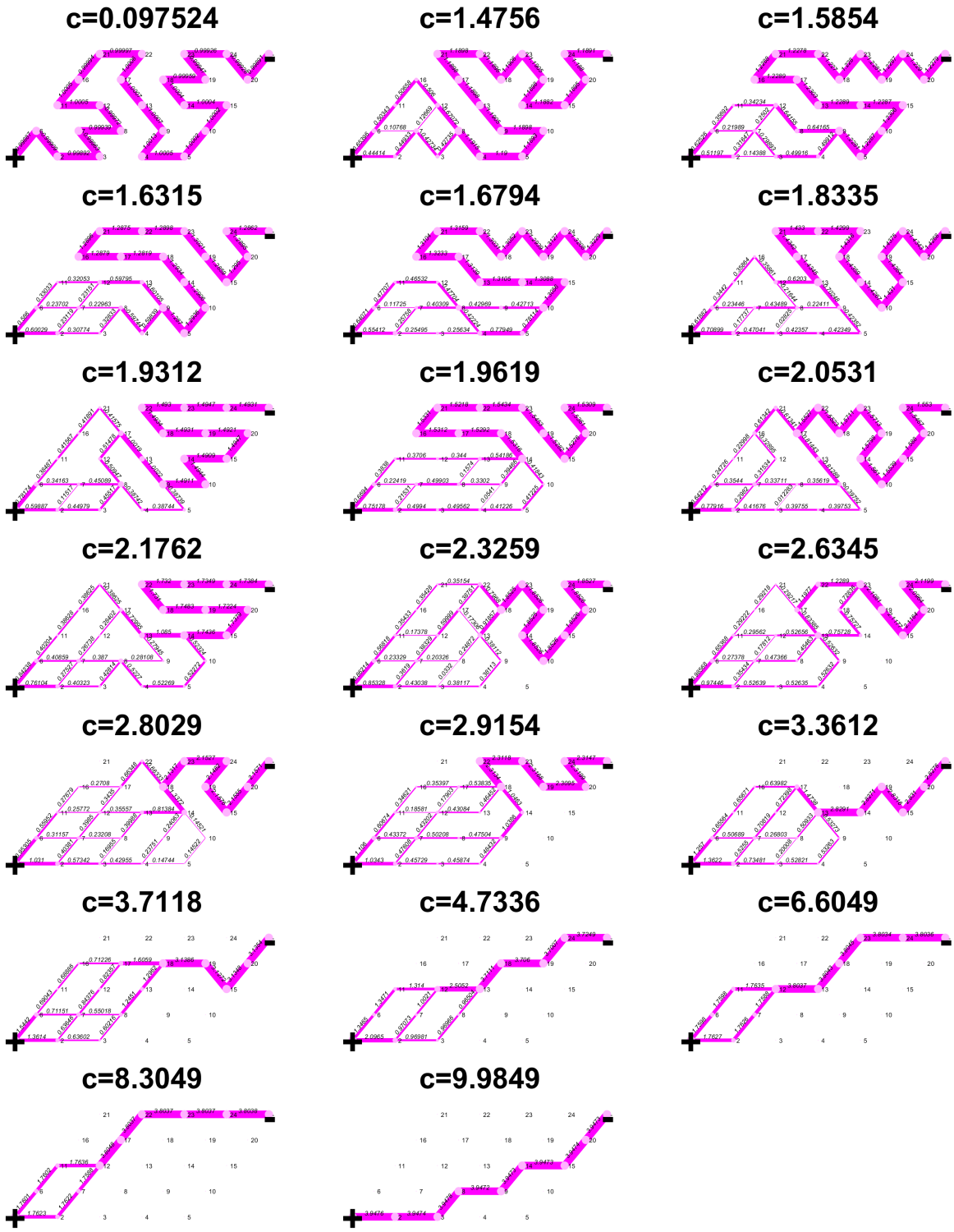


Figure 4.3.8: Globally optimal conductance networks computed when by c is increased.

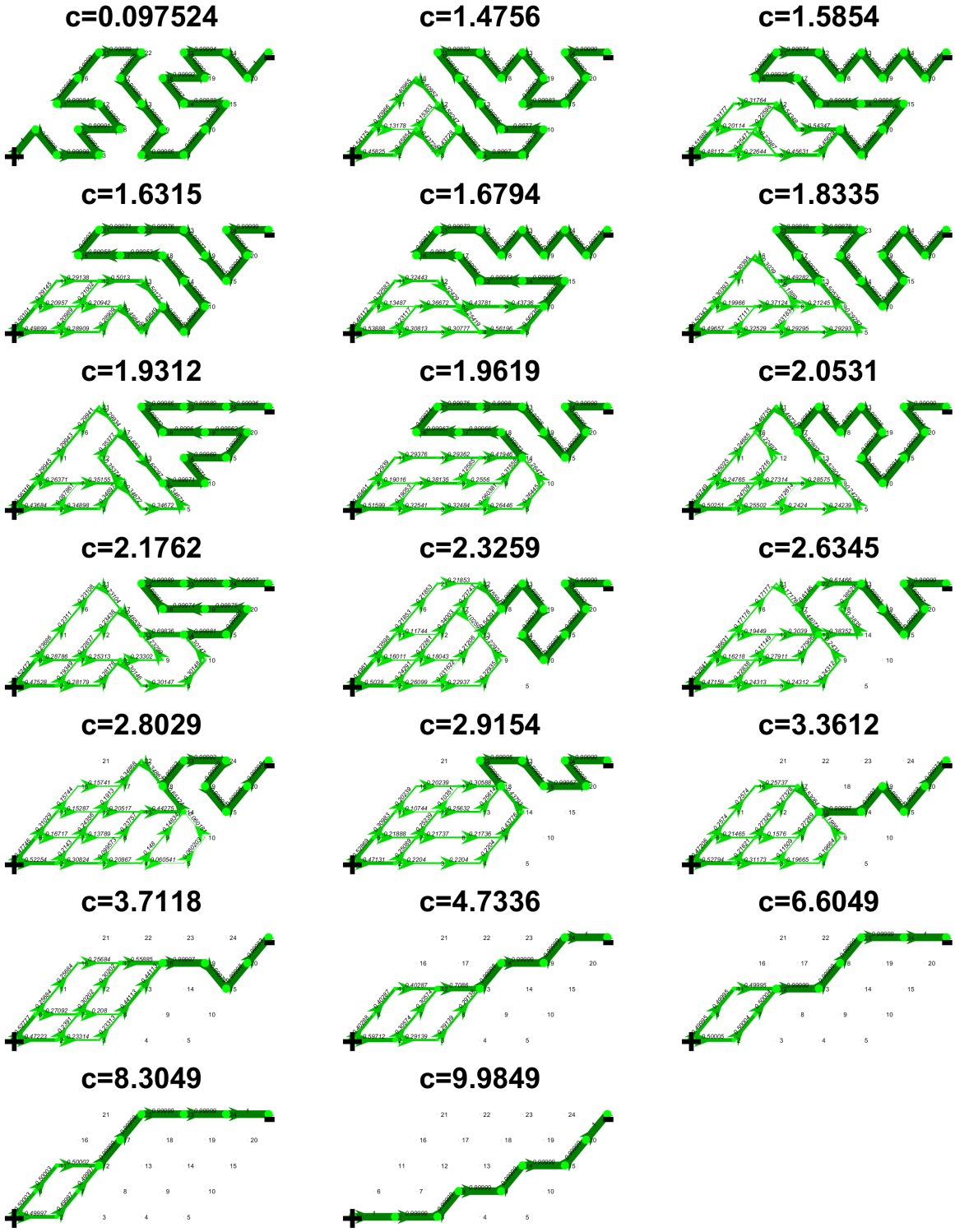


Figure 4.3.9: Globally optimal flow networks computed when by c is increased.

The globally optimal networks on the triangular ambient grid share a similar pattern to those on the square ambient grid (Fig. 4.3.9). Both begin with the tour through all 25 nodes. At a finite value of c loops appear and the number of independent loops grows with increasing c. The structure of the networks is to have the bifurcation at the source and ending with one final recombination that bottlenecks all of the flows into a path to the sink of flow 1. The size of the loopy sub-network increases to maximum and the number of independent loops reaches a maximum as well as c increases. Then both the number of loops and the average path length begin to fall, until the average path length is at the minimal value, 9 whereupon the network starts to transition toward the geodesic by losing loops one at a time.

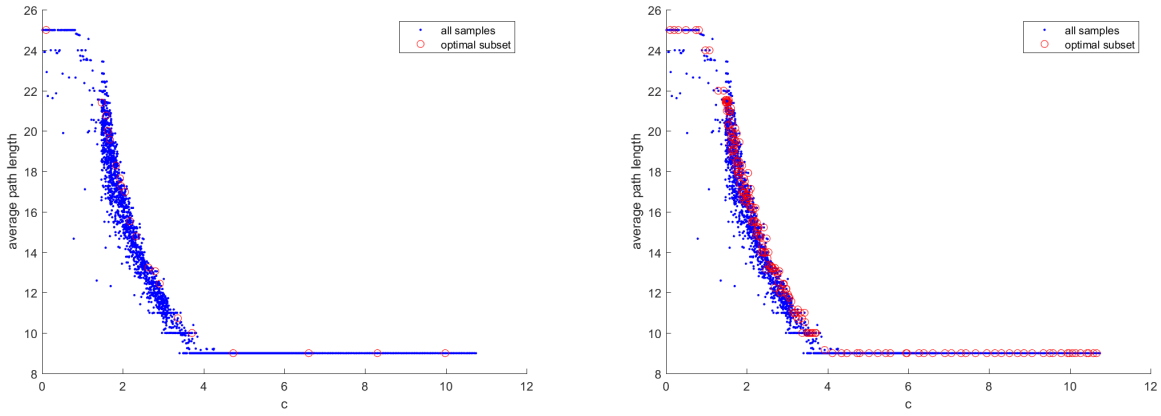


Figure 4.3.10: The average path lengths of minimal networks for every sampled network in blue and each of the globally optimal networks from figure 4.3.7 (red circles). In the right panel, we add globally optimal networks by subdividing each subinterval of c values into 10 equal subintervals.

Plotting the average path lengths of every network reveals the trend of decreasing path length with increasing c (Fig. 4.3.4). There do appear to be terraces in the scatter plot at every possible integer path length. However, the decreases in path length do not occur stepwise except at small values of c (close to the tour) and large values of c (close to the geodesic). To elaborate, the average path lengths appear decrease by 1 or 2 at a time for $c \leq 4.119$, and this structure survives even the addition of more global optima when we

subdivide each of our search intervals into 10 subintervals, increasing the number of global optima by 10. In particular, for c in the interval $[.01, 1.4250]$ including the end-points, the average path length is a non-decreasing step function taking on values 25, 24 and 22. Refining the selection did not result in a network length 23, though we can not rule out that such a network would be discovered if we further subdivided the interval. We interrogate the emergence of non-quantized path lengths by plotting the global optima with 10 subdivisions from Fig. 4.3.10 over the intervals $c = [0.46, 2.20]$ and $c = [1.49, 4.87]$(respectively the left and right panels of Fig. 4.3.11).

Our analysis of networks using loop counting produces very similar results to the square ambient network (Fig. 4.3.12), except the maximum globally optimal Betti number is 9.

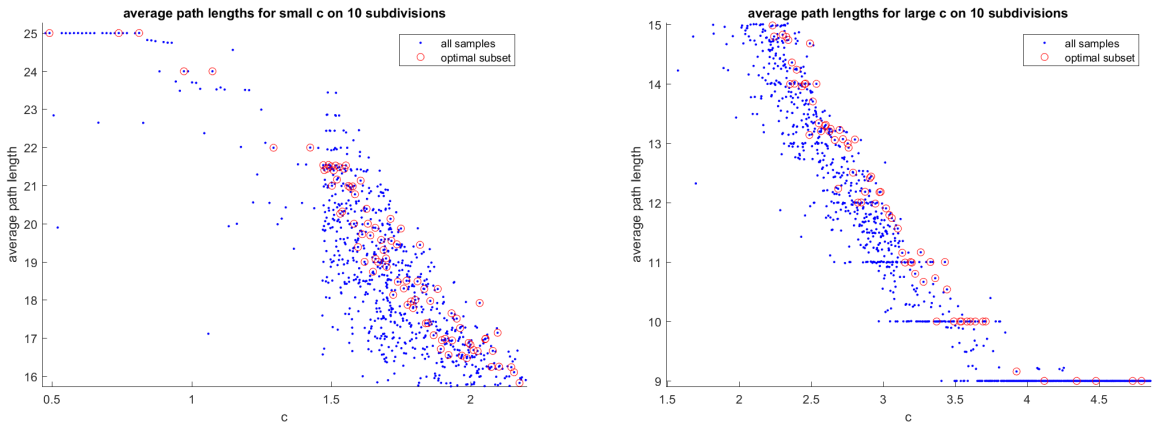


Figure 4.3.11: The average path lengths for small c (left) and large c (right), with 10 global optima per search interval Fig. 4.3.10. Note that globally optimal networks find many different average path lengths, and that the average path length shows high variability among similar values of c. For $c > 2.4$, the globally optimal networks may cluster around average path lengths 14, 13, 12, 11, ..., 9, but in the range $1.5 < c < 2.4$, we see no evidence of clustering in the mean path lengths.

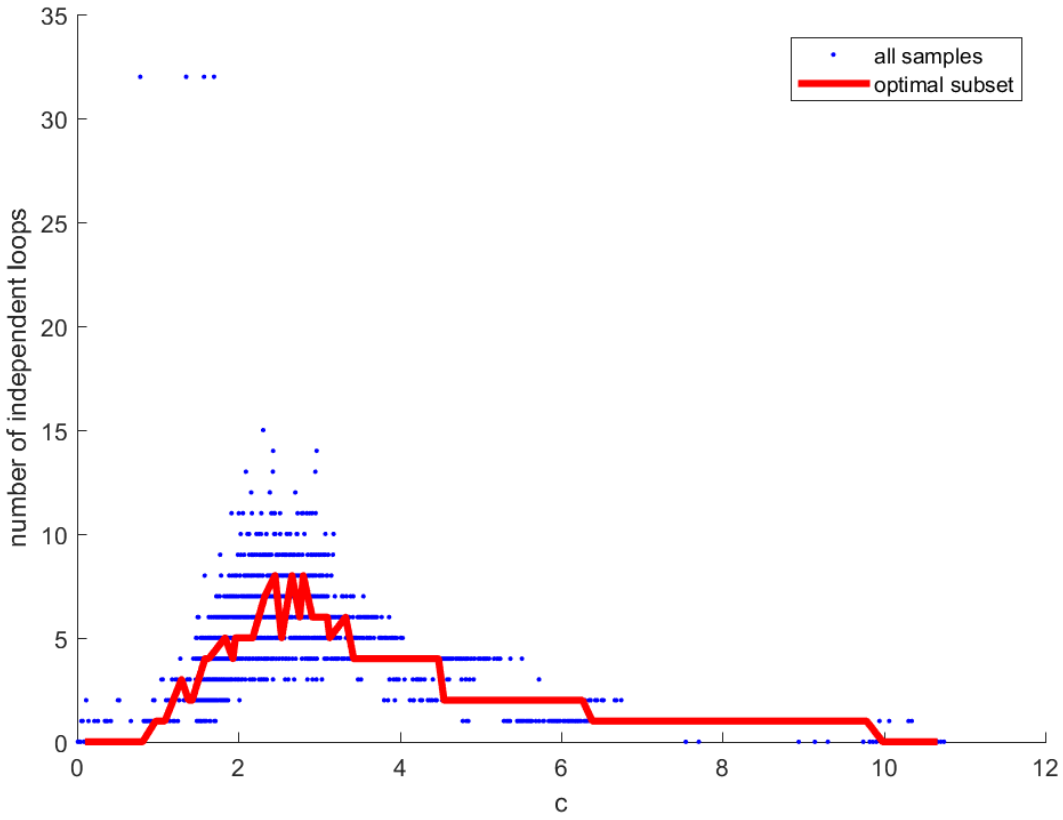


Figure 4.3.12: Number of independent loops in the conductance networks as a function of c for local (blue) and global (red) optima chosen from dividing the entire range equally into 100 intervals. Among the global optima, the minimum possible Betti number is 0 and the maximum is 9. The maximum possible Betti-number for a sama

In Fig. 4.3.13 we plot the average path length and the number of independent loops for globally optimal networks chosen one for each search interval. As the average path length of optimal networks decreases, less space is needed in the ambient grid for long tortuous paths. The space is filled by loopy subnetworks containing more direct paths to the bottleneck. This suggests that a low Betti number for small c followed by a steep increase is due to removing the spatial constraint on the size of loopy subnetwork as optimal networks transition from the conjectured NME optimizer at $c = 0$, the tour.

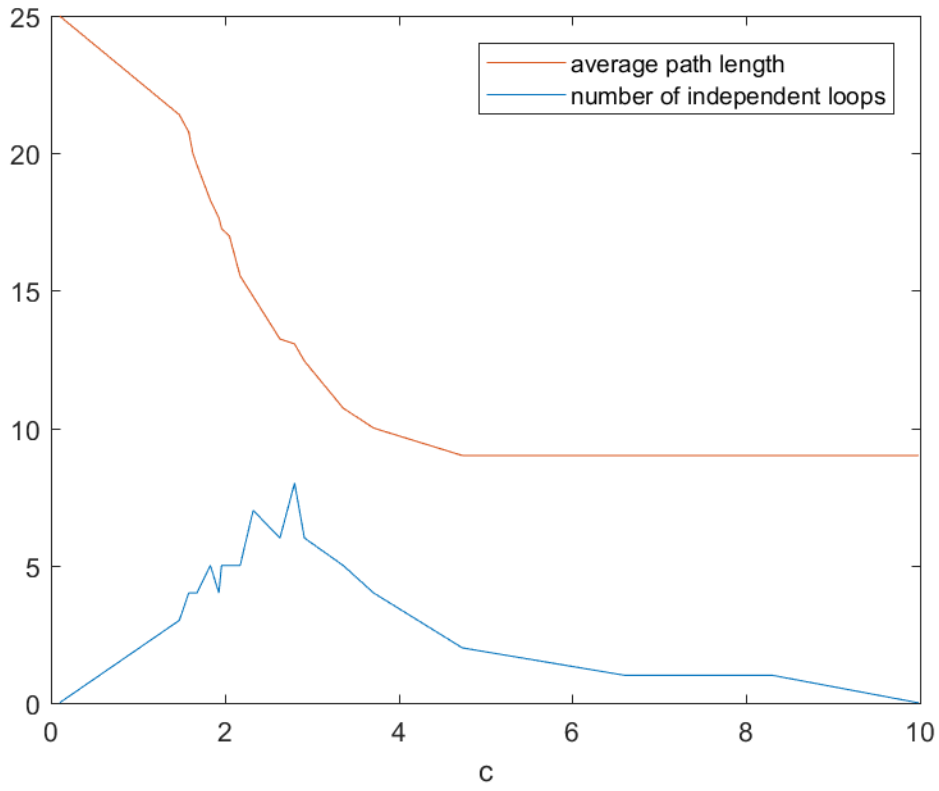
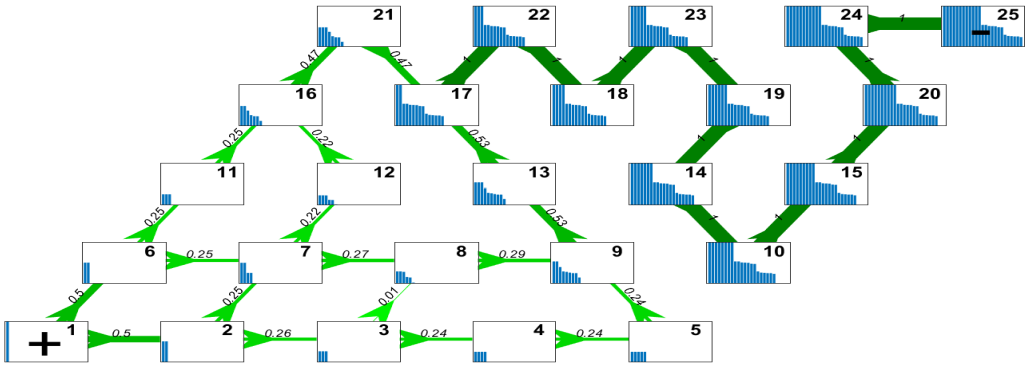
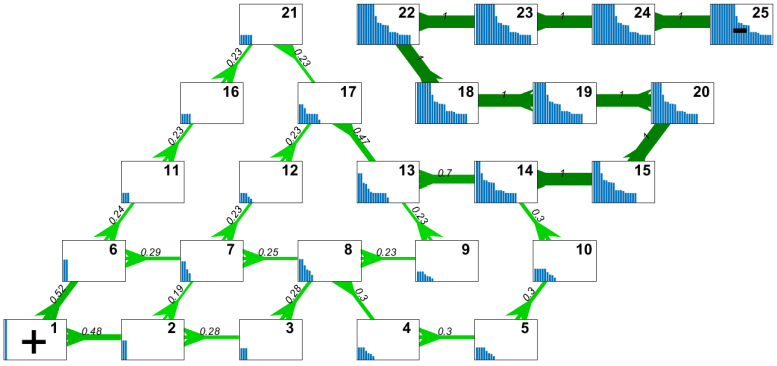


Figure 4.3.13: The average path length and the number of independent loops for the globally optimal networks (one network for each search interval) in figures 4.3.7 and 4.3.9.

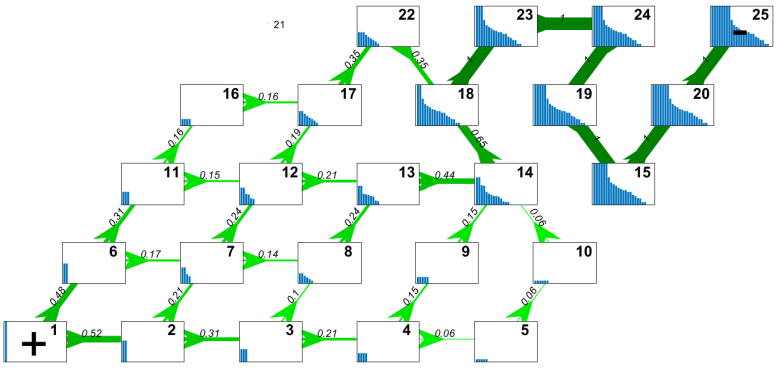
4.3.4 Signal distributions



(a) Globally optimal network 9, $c = 2.0531$, CMD $= -16.785$, NME $= -38.904$, $D = 10.774$



(b) Globally optimal network 10, $c = 2.1762$, CMD $= -15.442$, NME $= -34.8$, $D = 8.895$



(c) Globally optimal network 10, $c = 2.8029$, CMD $= -10.6$, NME $= -27.548$, $D = 6.0468$

Figure 4.3.14: Visualization of flow networks from the globally optimal networks in Fig. 4.3.9 The bar graphs at each node i depict the distribution of arriving signal strengths $\tilde{q}_{ji}$. The bar graphs at each node i depict the distribution of arriving signal strengths $\tilde{q}_{ji}$ with the signals sorted by decreasing magnitude.

The flow networks in Fig. 4.3.14 contain many different loops with different path-lengths. In figure (a) note that the distribution of nodes at the most down stream node of the loopy subnetwork has a hierarchical appearance, with several signals having strength 1, with a sharp drop to a strength of about $\frac{1}{2}$ followed by another drop to $\frac{1}{4}$. This characteristic shape appears to smooth out to a continuous decline to 0. For these networks, it appears that there is a higher density of loops closer to the source. This leads to more mixed distributions of signals being transferred down stream. Unlike the case of the ambient square grid, there is a variety of path lengths from the source to the final re-combination of flows in the loopy subnetwork. This may be due to the connectivity of the triangular grid allowing for every path length.

4.4 Two sources and two sinks on the square ambient grid with side length of 5 nodes for $\gamma = .5$

So far we have restricted our numerical experiments to networks with a single source and single sink. Our analytical results will be largely confined to such networks (see Chapter 5). However, the flexibility of our numerical algorithm allows us to explore other source and sink geometries. In this section we use numerical experiments to analyze optimal networks between a pair of sinks and pair of sources, that are distributed on opposite sides of a square ambient grid. The major change that we expect from such a network is that the optima for $c \to 0$ and for $c \to \infty$ are topologically different. For $c \to 0$ the global optimum network remains the tour (although the order in which the nodes are visited is important), why for $c \to \infty$ the optimal network now consists of two straight line paths, directly linking each source to its counterpart sink. Our numerical investigations probe the set of morphological transitions that occur between these two limits.

For two sources and two sinks it is outside of the scope of this study to state an all encompassing proposition about the structure of the mixing-dissipation cost function minima

curve. Instead, through a high number of samples we find several interesting features about how optimal networks differ as c is increased and about the associated jumps in the cost curves.

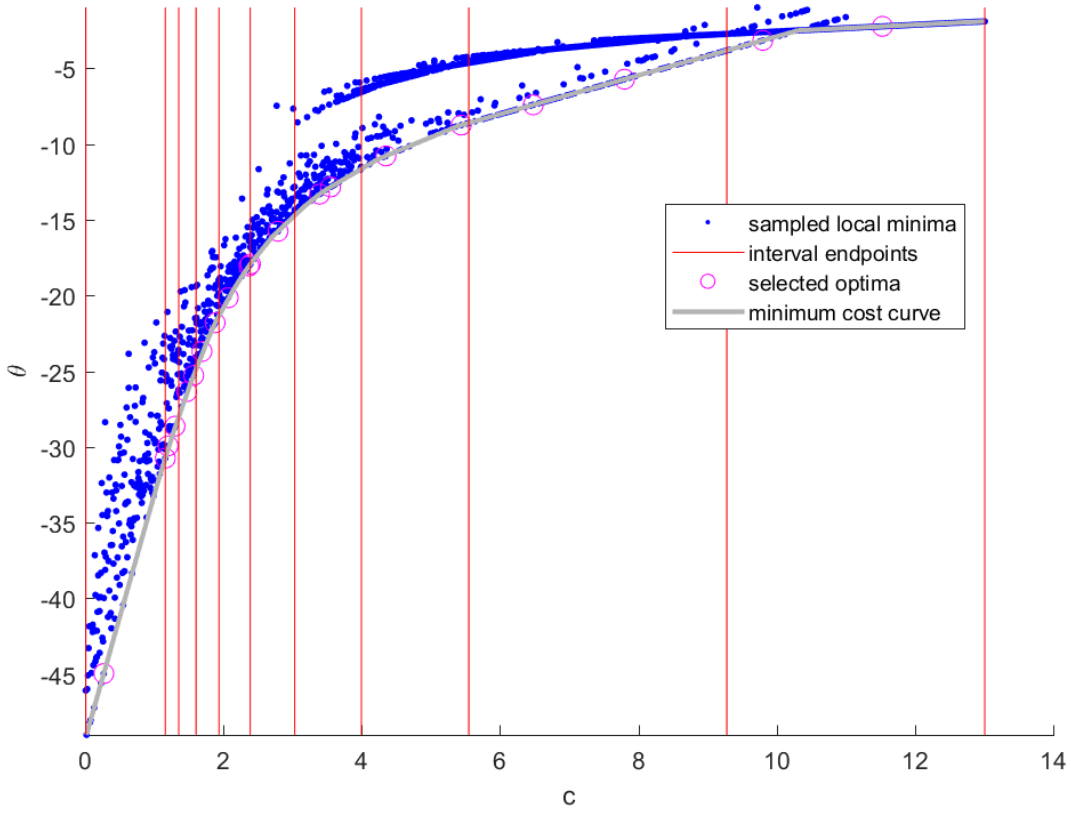


Figure 4.4.1: Scatter plot of the cost function NME + cD for locally optimal networks computed by numerical optimization (blue points). The grey is the minimal cost curve. The vertical red lines represent the end points of intervals over which we choose globally optimal networks, as networks that minimize the distance from the minimal cost curve (See Chapter 6 Section 8). We refer to these intervals as **search intervals** The c and NME + cD values for globally optimal networks are shown as magenta points.

The optimal mixing-dissipation cost curve (Fig. 4.4.1) seems to be comprised of 2 main curves made up of local optima. The locally optimal networks contain either one connected component or two connected components. The optimization algorithm is initialized with random conductances. As c is increased and more emphasis is placed on keeping dissipation

low, the probability that our optimization procedure can find a favorable sequence of steps towards the one-connected component optimum from it's initial condition, rather than immediately jump to any of increasingly favorable 2 connected component networks, decreases. We found that this caused a sudden discontinuous jump in the set of globally optimal networks. The sampling was repeated, converging to the set of mixing-dissipation minimizers pictured in Fig. **??**.

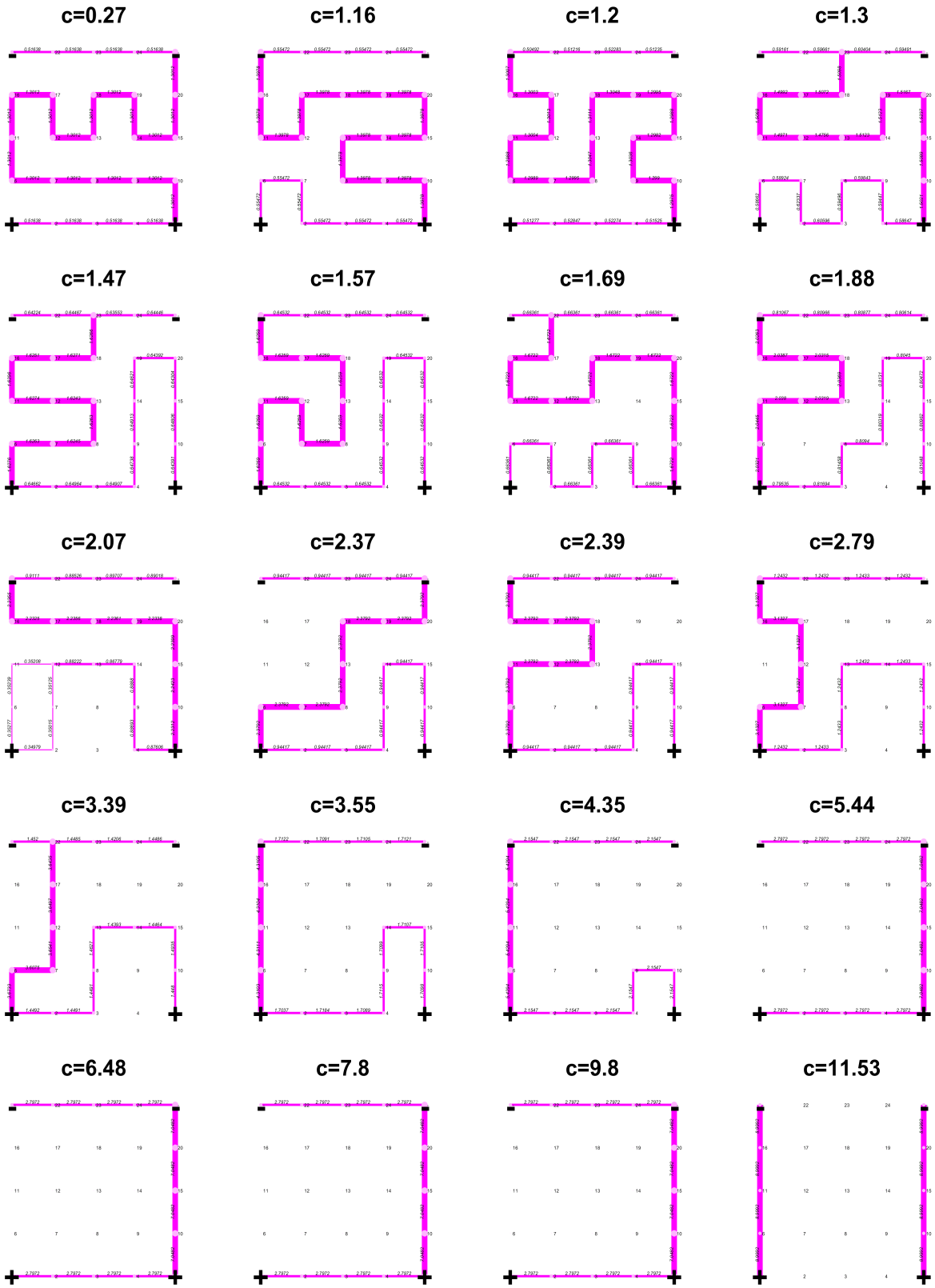


Figure 4.4.2: Selected optimal conductance networks by increasing c from right to left and top to bottom.

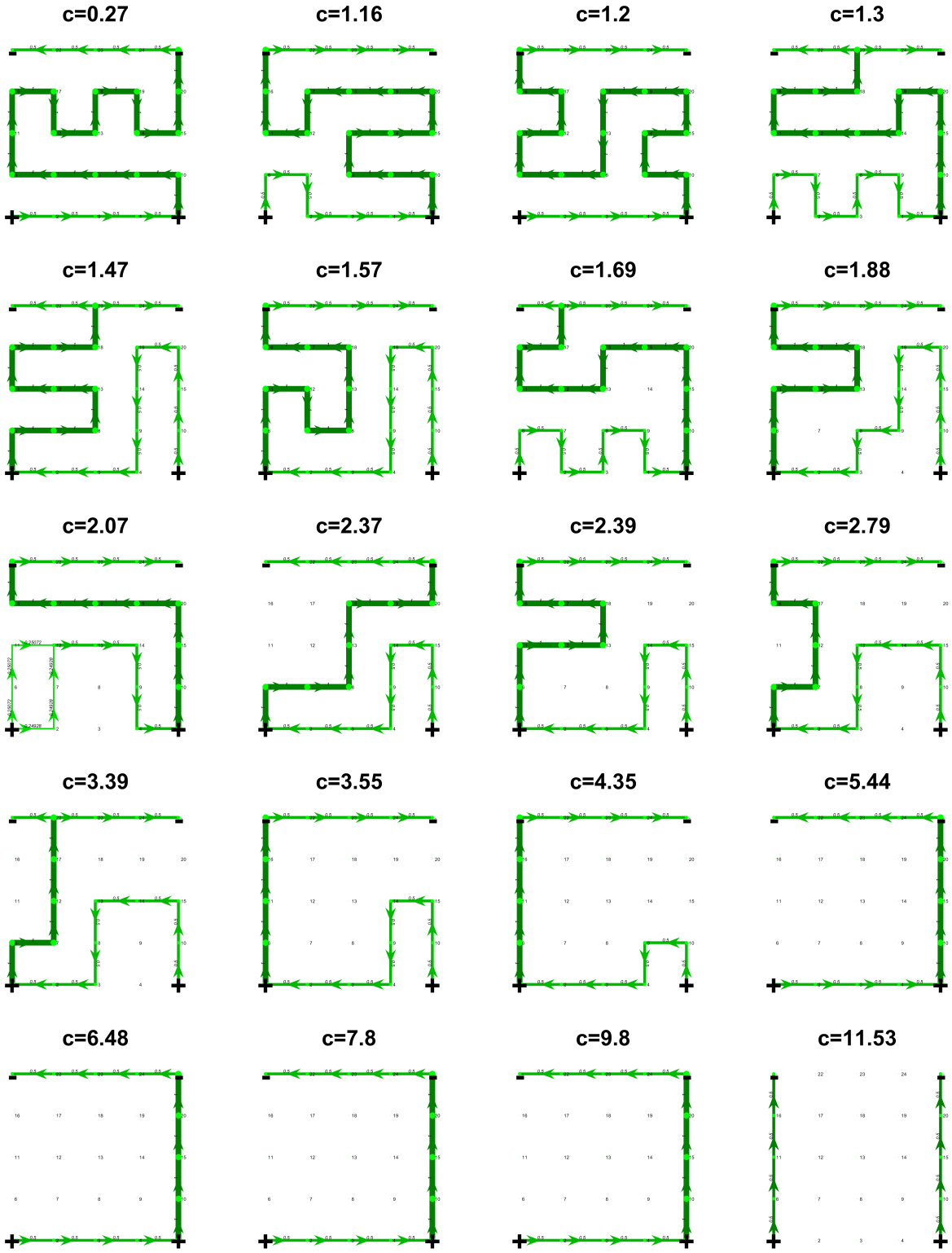


Figure 4.4.3: Selected optimal flow networks by increasing c from right to left and top to bottom.

A sequence of globally optimal flow networks are shown in Fig. 4.4.3. Similarly to the single source, single sink case, the majority of our networks are trees with loops being seen only for a very narrow range of c.

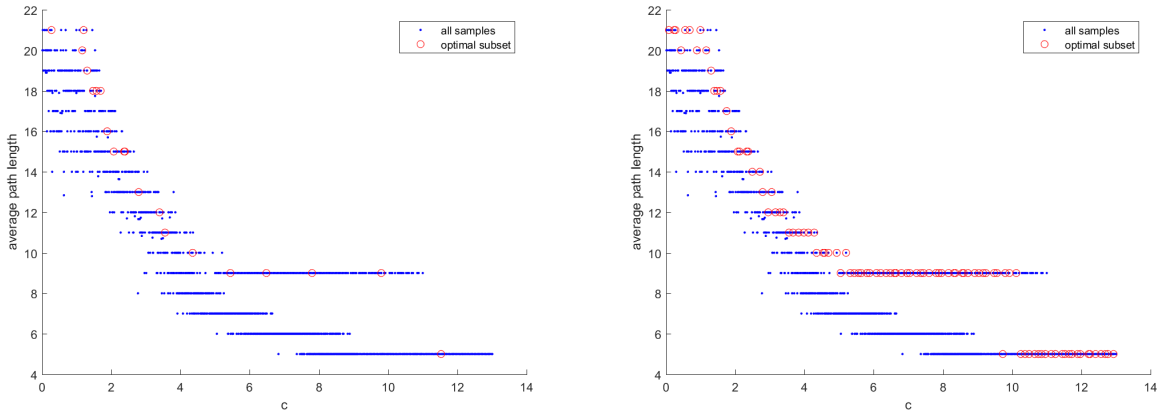


Figure 4.4.4: Average path lengths of every sampled network. Globally optimal networks in figures 4.4.1 and 4.4.3 are circled in red. On the right we divide up the entire domain of c into 100 equal parts, increasing the number of global optima that are assayed. The transitions between path lengths is stepwise in c. Over a large range of c values ($5 < c < 10$) the globally optimal network is the minimally dissipative tree, there is an abrupt transition at $c = 4.5$, in which this tree is replaced by the pair of geodesics as the global optimum.

CHAPTER 5

Theorems

5.1 Introduction

In this section, we take a mathematically rigorous approach to precisely stating and proving discoveries about conductance and flow networks which optimize the NME, NSE, dissipation and the cost of mixing and dissipation. To start, we compute and prove formulas for the NME in isolated scenarios to gain intuition for the process that the NME describes and the techniques we use. We fully describe the properties and values of functions on path graphs as well the next incrementally more complex type of network, paths with a single loop in the front, similar to those we generated in the numerical results (see Figure 4.2.2).

After warming up we state and prove general theorems characterizing the structure of the graph of the minimal mixing-dissipation cost θ, which get to the heart of why the numerical approximations often appear piecewise linear for single-source single-sink networks. We also break ground on how decreasing the material cost exponent γ affects the set of optimal networks. Specifically, for small material cost exponents, only networks whose mixing falls within a vanishingly small radius of one of a finite set of NME curves remain, these NME curves are generated by path networks of consecutive length. More so, for small γ the remaining set of minimizers cannot differ by more than a hair-thin vessel from a single path.

The proof involves a few key insights. Among them is that we can only safely erase nodes that not only have small f_i in an absolute sense, but by a relative sense, even when compared to small junctions so that their signal is small compared to rest and is barely detectable at

most nodes. Another insight is the re-discovery of a version of Murray's law that can be used to predict a sharp blow up of dissipation on networks with more flows above a very small threshold, and this helps to prove the removal of loops, as they have more edges than a path network on the same number of nodes.

We also prove two theorems about the structure flow networks optimizing the NME. On the path to the proof of the above collection of theorems, we show that the flow optimizing NME on a set of n of points is in fact the path visiting all n points. This is the reason that for small c, we obtain the tour as the result for our numerical examples (see Figure 4.2.4.) Finally, returning, in a sense, to the example of a path with a loop attached at the source, we prove that, if a network consists of two path parts, and one loopy subnetwork such that one path connects the source to the loopy sub-network, and one path connects the loopy sub-network to the sink, then the optimal path lengths with the same total amount of nodes are for all the nodes in paths to be placed from the loopy subnetwork to the sink. In this proof we construct a new probability distribution on the network and use tools from information theory to understand exactly how nodes gain or lose information based on shifting the network. For an example of such networks are the numerical results for $\gamma = .8$ on either the square or triangular grid (Figures 4.3.3 and 4.3.3).

5.2 Example Computations for NME

In this section we carry out two computations of the negative mixing entropy for the first for a path with unit total flow and flow network with a loop at the source representing a bifurcation that splits the flow into two symmetric parts, with one node on each side of the loop, and that then combines into a path that leads to the sink.

5.2.1 The NME and the minimal dissipation for a path flow of strength 1 through 9 points.

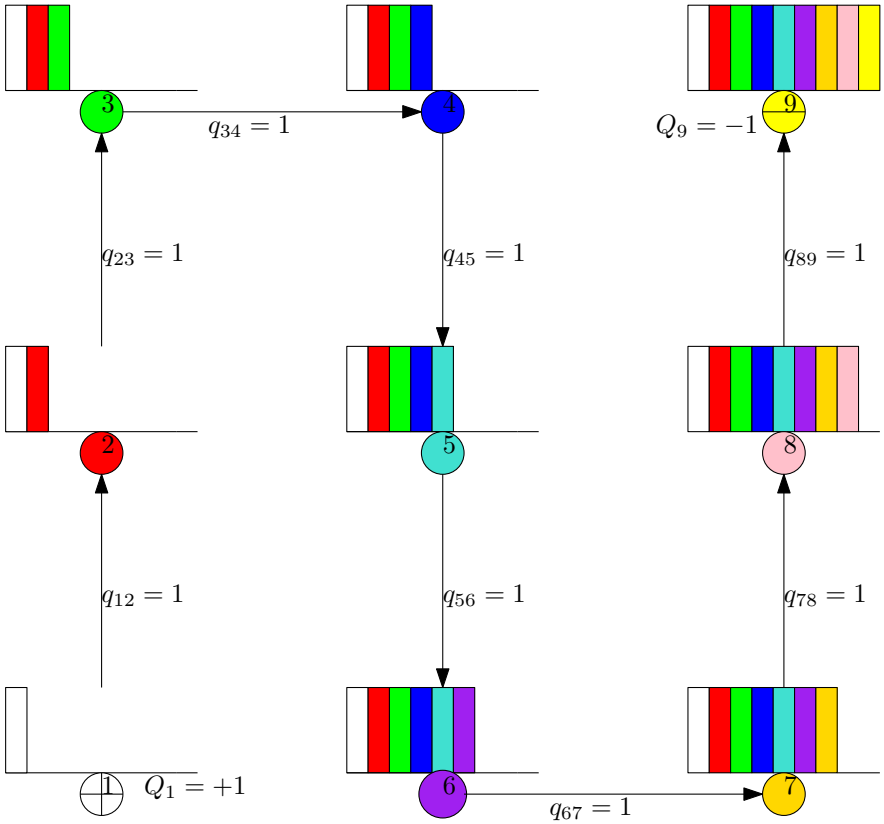


Figure 5.2.1: The path flow flow of strength 1. The bar charts are the strengths of the signals from each node represented by their corresponding colors.

We start our exposition by computing the NME and the minimal possible dissipation for the tour on the 3×3 grid with a single source at the bottom left corner, and a single sink at the top right corner, both of strength 1 given the material cost constraint $\sum_{(i,j)} \kappa_{ij}^{\gamma} = C$ for some $C > 0$. This is illustrated in figure 5.2.1.

For each node $i \in 1, 2, \ldots, 8$ there is only one non-zero out flow, $q_{i,i+1} = 1$. Therefore, the probability that a random walker transitions to $i+1$ given that it is on som $1 \leq i \leq 8$ is 1. Recall the definition of $\tilde{q}_{mn}$ (see Chapter 2. Section 2.1) as being the strength of node m, f_m multiplied by the probability of eventually visiting node n. That is, $\tilde{q}_{mn} = f_m P_{mn}$. Then

for any two nodes $1 \leq m < n \leq 9$, since the total out-flow of m is $f_m = \sum_{v:q_{mv}>0} q_{mv} = q_{m,m+1} = 1$ and $P_{mn} = 1$, $\tilde{q}_{mn} = 1$. Therefore, we calculate the probability distribution of signals arriving at node $1 \leq n \leq 9$ to be

$$\mathcal{P}_n(i) \quad = \quad \frac{\tilde{q}_{in}}{\sum_{j \leq n} \tilde{q}_{jn}} = \begin{cases} \frac{1}{n} & i \leq n \\ 0 & i > n \end{cases}.$$

This is the uniform distribution on n signals. The local negative mixing entropy at node n is

$$\mathrm{NME}_n \quad = \quad \sum_{i \leq n} \frac{1}{n} \log\left(\frac{1}{n}\right) = -\log(n).$$

The NME of the flow is the sum of the NME_n weighted by the f_n,

$$\sum_{n=1}^{9} f_n \mathrm{NME}_n \quad = \quad \sum_{n} -\log(n) = -\log(9!).$$

5.2.2 Example 2: Computing the NME of a path with a loop at the source

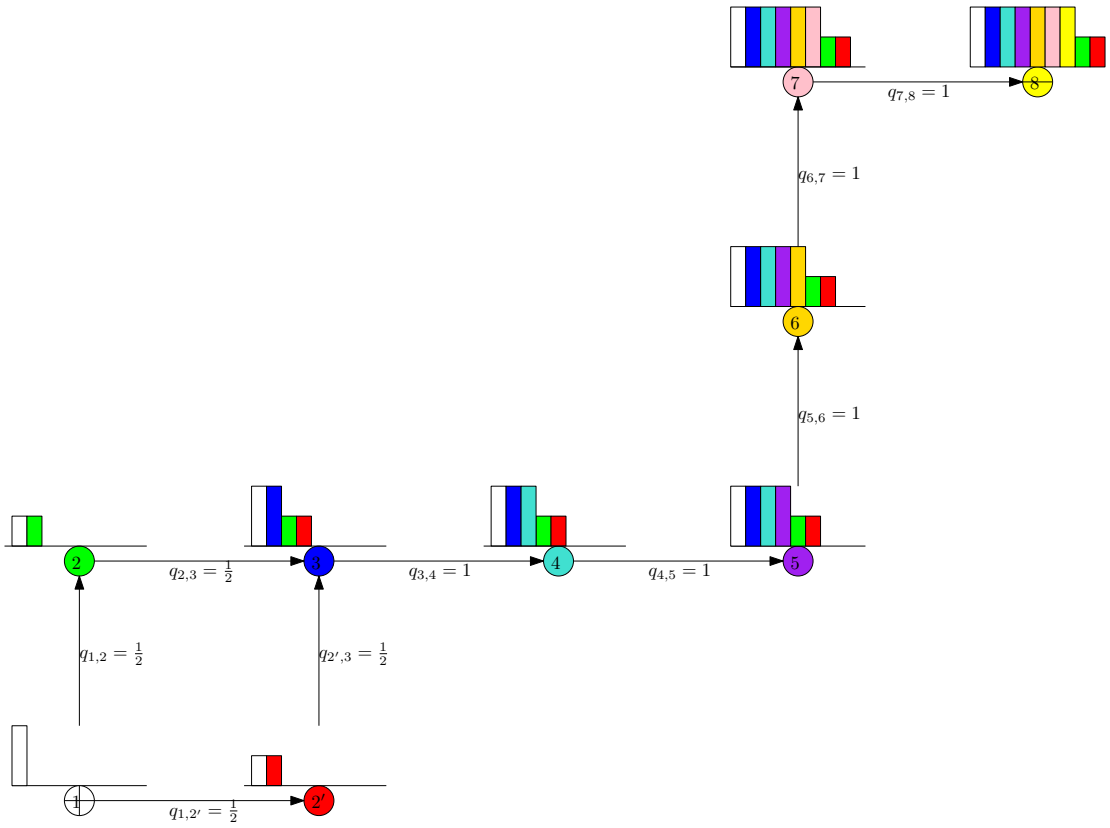


Figure 5.2.2: Diagram of the flows and bar graphs representing the un-normalized distributions of $\tilde{q}_{ij}$ at each node. The bars are ordered in order of their magnitude to be the clearest visual expression of how close to uniform signal distributions are at each node.

We now repeat the last exercise on a modified version of a path flow of strength 1. The flow network we consider is depicted in 5.2.2. For $3 \leq i \leq 7$ the flows $q_{i,i+1}$ form a path flowing from 3 to the sink at 8. At the source node 1 there is a bifurcation of the flow into two flows both of strength $\frac{1}{2}$. The separate $\frac{1}{2}$ flows pass through one node after the source until they meet at 3 closing a loop. That is, at the source, we have two outgoing flows $q_{1,2} = \frac{1}{2}$ and $q_{1,2'} = \frac{1}{2}$ that gather at 3 via the flows $q_{2,3} = \frac{1}{2}$ and $q_{2',3} = \frac{1}{2}$.

We first compute the distributions of signals arriving at each node. Node 1, the source, receives only the signal it generates, $\tilde{q}_{11} = 1$. Recall the convenient fact, in general networks for which 1 is assumed to be the only source, at node i $\tilde{q}_{1i} = \tilde{q}_{ii} = f_i$. Applying this on nodes

2 we have $\tilde{q}_{22} = \tilde{q}_{2'2'} = \tilde{q}_{12} = \tilde{q}_{12'} = f_{2'} = q_{12'} = \frac{1}{2}$.

Therefore $\mathcal{P}_2(1) = \frac{1}{2}$, $\mathcal{P}_2(2) = \frac{1}{2}$, $\mathcal{P}_{2'}(1) = \frac{1}{2}$ and $\mathcal{P}_{2'}(2') = \frac{1}{2}$. We represent the distribution of strengths of signals up to re-ordering arriving at either 2 or $2'$ as the sequence ordered by decreasing strength: $\left(\frac{1}{2}, \frac{1}{2}\right)$. The negative entropies of these distributions is the then equal to $-\log(2)$.

Now, at the meet of the flows from the bifurcation, node 3 we have nodes 2 and $2'$ feed into 3 equally $\tilde{q}_{23} = \tilde{q}_{2'3} = \frac{1}{2}$ while $\tilde{q}_{13} = \tilde{q}_{33} = f_3 = 1$. The sequence of $\tilde{q}_{i3} > 0$ ordered by decreasing strength is $\left(1, 1, \frac{1}{2}, \frac{1}{2}\right)$. Normalizing this to a probability distribution gives us the probabilties of choosing signals at node 3 $\mathcal{P}_3$ ordered in decreasing strength $\left(\frac{1}{3}, \frac{1}{3}, \frac{1}{6}, \frac{1}{6}\right)$. Note that this is equal to $\left(\frac{1}{3}, \frac{1}{3}, \frac{1}{2} \cdot \frac{1}{3}, \frac{1}{2} \cdot \frac{1}{3}\right)$, and so the negative local mixing entropy at 3 is

$$\begin{aligned} \mathrm{NME}_3 &= -\left(2 \cdot \frac{1}{3}\log(3) + 2 \cdot \frac{1}{2} \cdot \frac{1}{3}\log(2 \cdot 3)\right) \\ &= -\log(3) - \frac{1}{3}\log(2) \end{aligned}$$

For 4, $\tilde{q}_{14} = \tilde{q}_{34} = \tilde{q}_{44} = 1$ while $\tilde{q}_{24} = \tilde{q}_{2'4} = \frac{1}{2}$. Then $\tilde{q}_{i4} > 0$ up to re-ordered is represented by the decreasing sequence: $\left(1, 1, 1, \frac{1}{2}, \frac{1}{2}\right)$. Then the distribution $\mathcal{P}_4$ up to re-ordering can be represented by $\left(\frac{1}{4}, \frac{1}{4}, \frac{1}{4}, \frac{1}{8}, \frac{1}{8}\right)$. Computing NME_4 we get $\mathrm{NME}_4 = -\log(4) - \frac{1}{4}\log(2)$. We can repeat this calculation to get the negative local mixing entropies at nodes nodes $5, 6, \ldots 8$.

Alternatively we generalize the collection of signals for this sequence. Since $3, 4, \ldots, 8$ form a path flow with $T_{i,i+1} = 1$ for all $3 \leq i \leq 7$, if $4 \leq n \leq 8$ then $\tilde{q}_{mn} = f_m P_{mn} = f_m P_{m,n-1} T_{n-1,n} = f_m P_{m,n-1} = \tilde{q}_{m,n-1}$. Therefore, to compute the strengths of signals arriving at $n \geq 4$ we take a copy of the sequence of $\tilde{q}_{m,n-1} > 0$ and append $1 = \tilde{q}_{nn}$. Starting with the sequence represent signal strengths $\tilde{q}_{i3}$ $\left(2 \text{ copies of } 1, \frac{1}{2}, \frac{1}{2}\right)$ if $3 \leq n \leq 8$ then by repeating the last step $n-3$ times we have that the strengths of signals arriving at

n $\tilde{q}_{in}$are (up to re-ordering)

$$\left(n-1 \text{ copies of } 1, \frac{1}{2}, \frac{1}{2}\right).$$

The probability distribution obtained from normalizing this sequence is then

$$\left(n-1 \text{ copies of } \frac{1}{n}, \frac{1}{2}\cdot\frac{1}{n}, \frac{1}{2}\cdot\frac{1}{n}\right).$$

Computing the negative local mixing entropy at node n by taking the negative entropy of these probabilities we get

$$\begin{aligned}
\mathrm{NME}_n &= -H\left(\left(n-1 \text{ copies of } \frac{1}{n}, \frac{1}{2}\cdot\frac{1}{n}, \frac{1}{2}\cdot\frac{1}{n}\right)\right) \\
&= -\left((n-1)\frac{1}{n}\log(n) + 2\cdot\frac{1}{2}\cdot\frac{1}{n}\log(2\cdot n)\right) \\
&= -\left(\log(n) + \frac{1}{n}\log(2)\right).
\end{aligned}$$

We summarize these results in the following table:

	$\tilde{q}_{ji}$ at i in decreasing order by strength	$\mathcal{P}_i(j)$	NME_i
$i=1$	(1)	(1)	$-\log(1)=0$
2	$\left(\frac{1}{2},\frac{1}{2}\right)$	$\left(\frac{1}{2},\frac{1}{2}\right)$	$\frac{1}{2}\log\left(\frac{1}{2}\right)+\frac{1}{2}\log\left(\frac{1}{2}\right)=-\log(2)$
$2'$	$\left(\frac{1}{2},\frac{1}{2}\right)$	$\left(\frac{1}{2},\frac{1}{2}\right)$	$-\log(2)$
3	$\left(1,1,\frac{1}{2},\frac{1}{2}\right)$	$\left(\frac{1}{3},\frac{1}{3},\frac{1}{3}\cdot\frac{1}{2},\frac{1}{3}\cdot\frac{1}{2}\right)$	$-\log(3)+-\frac{1}{3}\log(2)$
4	$\left(1,1,1,\frac{1}{2},\frac{1}{2}\right)$	$\left(\frac{1}{4},\frac{1}{4},\frac{1}{4},\frac{1}{4}\cdot\frac{1}{2},\frac{1}{4}\cdot\frac{1}{2}\right)$	$-\log(4)+-\frac{1}{4}\log(2)$
5	$\left(1,1,1,1,\frac{1}{2},\frac{1}{2}\right)$	$\left(4 \text{ copies of } \frac{1}{5},\frac{1}{5}\cdot\frac{1}{2},\frac{1}{5}\cdot\frac{1}{2}\right)$	$-\log(5)+-\frac{1}{5}\log(2)$
6	$\left(5 \text{ copies of } 1,\frac{1}{2},\frac{1}{2}\right)$	$\left(5 \text{ copies of } \frac{1}{6},\frac{1}{6}\cdot\frac{1}{2},\frac{1}{6}\cdot\frac{1}{2}\right)$	$-\log(6)+-\frac{1}{6}\log(2)$
7	$\left(6 \text{ copies of } 1,\frac{1}{2},\frac{1}{2}\right)$	$\left(6 \text{ copies of } \frac{1}{7},\frac{1}{7}\cdot\frac{1}{2},\frac{1}{7}\cdot\frac{1}{2}\right)$	$-\log(7)+-\frac{1}{7}\log(2)$
8	$\left(5 \text{ copies of } 1,\frac{1}{2},\frac{1}{2}\right)$	$\left(7 \text{ copies of } \frac{1}{8},\frac{1}{8}\cdot\frac{1}{2},\frac{1}{8}\cdot\frac{1}{2}\right)$	$-\log(8)+-\frac{1}{8}\log(2)$

Table 5.2.1: Distributions of signals arriving at nodes in the path with a loop at the source, normalized probability distributions and the local negative mixing entropies computed at each node i.

The NME of the path with a loop at the source then comes out to be

$$\begin{aligned} \text{NME} &= \text{NME}_1 + \frac{1}{2}\text{NME}_2 + \frac{1}{2}\text{NME}_{2'} + \sum_{n=3}^{8} \text{NME}_n \\ &= 0 - \log(2) + \sum_{n=3}^{8} \left(-\log(n) - \frac{1}{n}\log(2) \right) \\ &= -\log(8!) - \log(2) \left(\sum_{n=3}^{8} \frac{1}{n} \right). \end{aligned}$$

From this computation we see that the path of length 8 with a loop at the source has lower NME than a path network of length 8, which, if we compute the NME via the method in Example 5.2.1, we get $-\log(8!)$.

5.3 Overview of section

5.3.1 Introduction and reason for one source one sink

In this part of the article we develop a mathematically rigorous story about the structure of networks with optimal NME, optimal combined mixing-dissipation cost, and the trends in optimal mixing-dissipation costs and optimal networks as $\gamma \to 0^+$. We do all of this for networks on ambient networks fed by exactly one source and feeding exactly one sink. The theorems we formulate and objects we choose to study are mostly inspired by the numerical results. This is a restricted case, but it provides us with a many different avenues for mathematical investigation. The results in chapter four inspire many questions that we will try to answer now: Why are minimal mixing dissipation costs piecewise linear on 5×5 ambient grids with $\gamma = .5$? Why are the optimal networks mostly paths of lengths changing by the minimal possible increments? What is the effect of increasing or decreasing the material cost exponent? What is being gained by having the most reticulated regions of the network located near the source?

5.3.2 Description of parts

5.3.2.1 Structure of the first order approximation to the minimal mixing-dissipation cost curve (Section 5.5):

In Section 5.5 we investigate a first order approximation to the minimal mixing-dissipation cost curve by minimizing over the set of paths connecting source and sink. We prove, with two steps, that when consideration is restricted to this set, the path length decreases at sharp non-differentiable transitions between paths with consecutive lengths. By consecutive lengths we mean for the square grid, since all paths connecting the same source and sink must have the same parity, the lengths decrease by two at the sharp transitions. On the triangular grid, which includes paths of lengths differing by 1 the sharp transitions switch between path lengths differing by 1.

5.3.2.2 Connection between dissipation and flow (Section 5.7)

In Section 5.7 we state and prove a lower bound on the dissipation given knowledge of the physical flows of a conductance network and that the conductances obey $\sum_{(i,j)} \kappa_{ij}^{\gamma} = C$. We do not assume knowledge of the exact conductances that give rise to the physical flows. Roughly we show that the dissipation is $> \left(\frac{\#\text{ non-zero flows}}{C}\right)^{1+\frac{1}{\gamma}}$ as $\gamma \to 0^+$. This will be used to compare the preference of flows of loopy networks with number of edges lying between the number of edges of two optimally mixing networks and show that loopy networks become infinitely more dissipative for low γ.

5.3.2.3 Total negative mixing entropy can be reduced to the subnetwork of the strongest nodes (Section 5.9.2)

Inspired by the structure of the NME function itself, we define a set of nodes in the flow network the effect of whose removal would be miniscule on the NME. These can be thought of as the weakest signals in the network. If we turn them off there is a very small change

to the entropy. Since NME $= \sum_i f_i \text{NME}_i$ and NME_i are bounded on finite networks, one might think to eliminate all networks with small f_i. This would be the same as drawing an absolute cut-off for total flow and everything above the absolute cut-off stays (strong nodes) and every node below the cutoff is removed with their flows (weak nodes). But then if a node barely makes the cut-off, it is possible that for the weak node that was retained, most of the flow into that node was supplied by a weak node that was eliminated, preventing us from accurately computing the local negative mixing entropy of that node. Therefore more than the absolute contribution needs to be taken into consideration, but the proportion of contribution as well. Therefore, the cut-off has a multiplicative gap. That is for a small mutliplicative constant δ we say for all i below the cut-off and j above the cut-off $\frac{f_i}{f_j} < \delta$. This guarentees that weak-nodes not only contribute small summands $f_i \text{NME}_i$ but their effect on the local negative mixing entropies of the strong nodes is very small. So they can be removed leaving the total negative mixing entropy unchanged. This is the content of definition 5.6 and theorem 5.8.

5.3.2.4 Removal of loops from optimal networks as $\gamma \to 0^+$

We investigate the shape of the cost curve and the structure of the exact nature as we vary γ, specifically for γ small. This gives us insight into why loops are penalized more heavily as γ is lowered. As we see from our second order approximations, larger loops being preferred might lead us to conjecture that more loops are preferred for high γ as well as large ambient networks.

In section 5.9 we prove a theorem (theorem 5.6) giving the description of the networks that are removed as $\gamma \to 0^+$. Note, this theorem applies to ambient networks where every path between the minimum and maximum are accessible. This means that consecutive path lengths differ by 1. This theorem applies to the triangular grid. It does not apply to the square grid.

We show that for all m between the minimal path and maximal path, networks with

NME outside of small neighborhoods outside of the best NME on m and $m+1$ nodes for all m are not optimal for any weight on dissipation. In mathematical terms $-\log((m+1)!)+\epsilon <$ NME $< -\log(m!)-\epsilon$That is, transitional networks with loops (at the source node) between the best networks on m and $m+1$ nodes are removed from consideration as $\gamma \to 0^+$. In this way, the transitions along the cost curves are sharpened, becoming closer to sharp jumps between optima from path networks.

5.3.2.5 Choosing building cost C so that the constraint satisfies $\sum_{(i,j)} \kappa_{ij}^{\gamma} = C$

In the next section we will describe in general terms how we prove the theorem about removing loopy networks between two consecutive paths. Before that we discuss the significance of the building constant C. In section 5.6 we show that if every other parameter is unchanged, then changing the building constant acts on the set of optimal costs by re-scaling the weight of dissipation c by $(C'/C)^{\frac{1}{\gamma}}$. That means, minimal mixing dissipation cost curves with different building materials carry the same information about the types of transitions and the specific networks from which optima arise. This invariance does not render C useless. Instead we use it as a way to focus on networks with between m and $m+1$ points as $\gamma \to 0^+$. As we have mentioned, the lower bound on dissipation yields $> \left(\frac{\#\text{ non-zero flows}}{C}\right)^{1+\frac{1}{\gamma}}$. To focus on networks with m flows, such as the path on $m+1$points we take $C = m$. Then, from computations in the next section, the dissipation for the path with m nodes goes to 0, the path with $m+1$ nodes goes to $m+1$ and a network with $m+1$ flows may be somewhere between. Networks with more flows have dissipation of ∞.

Therefore, the graphs of lines from τ_m, a network with m flows and τ_{m+1} is a flat line, a line with slope $\in (0, m+1)$ and a line with slope $m+1$. Now we are freed from the difficulty of computing ratios between slopes at every γ and we can instead look at the limiting value and the behavior as $\gamma \to 0^+$ can be derived from the intersections of several lines with known values at $c = 0$ and slopes.

5.3.2.6 Idea for proof of removal of loops from optimal networks as $\gamma \to 0^+$. A combination of mathematical connection of NME to the flows and dissipations to flows.

The idea for the proof of theorem 5.6 is to combine the connections we discovered between the optimal mixing networks and the flow network and the dissipation and the flow network.

We use the fact that optimal NMEnetworks on m nodes are paths with NME $= -\log(m!)$ as a scaffold to attach the loopy networks. The first thread connecting networks with $-\log((m+1)!) + \epsilon <$ NME $< -\log(m!) - \epsilon$ to the scaffold is theorem 5.8. We can take the dominance factor in definition 5.6 to partially transform the assumption of being transitioning between $m+1$ node and m node paths and the distance ϵ to be bounded away from the consecutive optimal situations into a statement about flows. We use this idea to discard the weak nodes while not changing the NME of the loopy network by more than (for example) $\epsilon/2$. Now the network has less nodes and none of the nodes are much less influential than the other nodes. That is, $f_i > \delta f_j$ for all i and j.

We know that to beat the entropy of $-\log(m!)$ the remaining strong nodes must have size n at least $m+1$. Otherwise it would be sub-optimal to the path on m nodes. Ordering the remaining nodes in order of f_i we see that $f_1 = 1$ because it is the source, $f_2 > \delta f_1$ and in general, $f_i > \delta f_{i+1}$. Then we can bound the decreasingly ordered total flows $f_1, f_2, \ldots, f_n$ by a geometric sequence similar to $1, \delta, \ldots, \delta^{n-1}$. Since the network is connected and we assume a maximal degree of k we can find a sub-tree with at least $n-1$ flows bounded below by dividing the total flows into k equivalents: $q_l \geq \frac{\delta^{l-1}}{k}$. In the cases where $n > m+1$ or $n = m+1$ and there is a loop in the network of these flows, there are at least $m+1$ flows bounded below by the geometric sequence $\frac{\delta^{l-1}}{k}$ $l = 1, 2, \ldots, m+1$.

Now we bring in the connection to dissipation to show how the network becomes progressively more unfavorable as $\gamma \to 0^+$.The network is such that NME $> -\log((m+1)!)$. We use the skill of taking $C = m$ to pin the dissipation of the path of $m+1$ nodes at m. By assumption

the lower bound theorem 5.7, the dissipation of our loopy network can be thought of as being $> \left(\frac{\#\text{ non-zero flows}}{m}\right)^{1+\frac{1}{\gamma}} = \left(\frac{m+1}{m}\right)^{1+\frac{1}{\gamma}} \to \infty$. This divergence is uniform because δ was specified by m and ϵ. Therefore the cost of the loopy network eventually exceeds that of the path on $m+1$ nodes so it is removed from our consideration on its transitional domain of c.

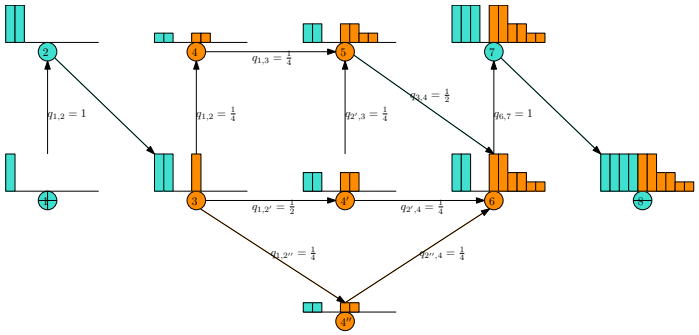
The one last case is if there are $m+1$ nodes left after eliminating the weak nodes, and there are at most m flows in the set bounded. This case is patched up by bringing in the assumption that NME $> -\log((m+1)!) + \epsilon$ to show that cannot only have m flows, but there must be some strong enough flows diverting enough flow away from the path. And these strong flows again feed into our lower bound on dissipation to eliminate this loopy network.

5.3.2.7 Loopy sub-networks occur at the source

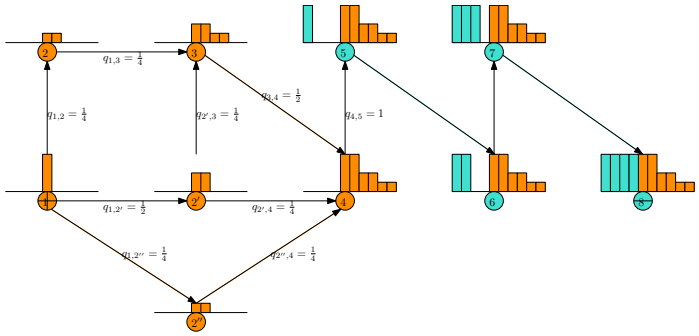
In the last section of this chapter, we prove that for a single source and single sink network, that the observation we made in Chapter 4 about the placement of loops within the network has a rigorous proof. To do this we appeal to the NSE instead of the NME because recall in Chapter 2 we showed that NSE is equal to the NME with the signs on the boundary flows reversed. We need the concept of conditional entropy to describe how nodes lose and gain entropy depending on their position in the network and their distributions of down-stream nodes. This theorem is proven by first proving that if it is possible to move any node from the path that links the loopy subnetwork to the sink to the path that links it with the source, then we will obtain a lower NME. This step is shifting the loopy subnetwork towards the sink. Repeating this until the loopy subnetwork is at the sink then proves the theorem.

5.4 Definitions and Basic Facts

We introduce some helpful facts and definitions for thinking about optimally mixing, dissipative and combined mixing-dissipation networks which we will use through out the paper. We find



(a)



(b)

Figure 5.3.1: Figure 5.3.1a. is an example of $P_1 = \{1, 2\}$ $\mathcal{L}$ the networks with loops and $P_2 = \{7, 8\}$. Figure 5.3.1b. is an example of a network with same number of nodes as that depicted in figure 5.3.1a except with the loopy network at the source $P_1 = \emptyset$ and $P_2 = \{5, 6, 7, 8\}$.

it simplifying to consider for every flow, the set of conductances which would result in that physical flow that have the lowest dissipation. We first explicitly note the simple fact, which is necessary to allow thinking of minimal mixing-dissipation networks in terms of the directed flow network q_{ij}.

This is true because NME is a function on the flows and does not need the conductances besides to find the physical flows.As an abuse of this notation, we also refer to a network of flows q_{ij} which are physical flows of some conductances and compatible with the boundary flows. Since we are concerned about minimizing our cost function, we also abuse the notation of D, in Chapter 2 was previously argued to be a function of κ_{ij} to write D (recall $D(\kappa_{ij}) = \sum_{(i,j)} \frac{q_{ij}^2}{\kappa_{ij}}$), as a function of q_{ij} to mean the minimal possible dissipation of a network of

conductances obeying the building cost conrtraint with its physical flows being q_{ij}. Therefore we introduce the definition:

Definition 5.1. For a fixed building cost $\sum_{(i,j)} \kappa_{ij}^{\gamma} = C$, we define the dissipation of a flow $D(q_{ij})$ to be minimum of the dissipation over the conductances obeying the building cost and giving q_{ij} as a physical flow. That is

$$D(q_{ij}) \quad = \quad \min_{\sum_{(i,j)} \kappa_{ij}^{\gamma} = C : q_{ij} \text{ physical flows of } \kappa_{ij}} D(\kappa_{ij}).$$

Then we define the mixing-dissipation cost on flow networks using this definition to be

$$\mathrm{NME}(q_{ij}) + cD(q_{ij}).$$

Written mathematically, the dissipation of a network of flows is

$$D(q_{ij}) \quad = \quad \min \left\{ \sum_{(i,j)} \frac{q_{ij}^2}{\kappa_{ij}} : \kappa_{ij} \text{ s.t. } \sum_{(i,j)} \kappa_{ij}^{\gamma} = C \,\text{and}\, q_{ij} \text{ are physical flows of } \kappa_{ij} \right\}.$$

This allows us to use specific flow-networks to build approximations for the minimal mixing-dissipation cost curves without going through the trouble of repeatedly choosing the best conductances. As we have observed, the curve $\theta(c)$ in many cases with one source and one sink appears to be approximated well by minimizing over a smaller set of networks, just trees connecting sources and sinks. In the single source and sink case these are path graphs. We have the following definition.

Definition 5.2. The path of length m is any flow network and its minimally dissipative conductance network on m nodes which is isomorphic to labeling the nodes $1, 2, \ldots, m$ setting 1 to be the source, m to be the sink and taking the edge set to be the directed edges $(1,2), (2,3), \ldots, (i, i+1), \ldots, (m-1, m)$. We denote this τ_m.

In the next fact we deduce that for minimal dissipation the conductances must be uniform.

We also note that if a network has a path sub-network, that is a sub-graph including all edges connecting it to ambient graph isomorphic to τ_m for some m, then it must also have uniform conductances.

Theorem 5.1. *Let τ_m be a path flow that is the physical flow of conductances with total material cost $\sum_{(i,j)} \kappa_{ij}^{\gamma} = C$. The dissipation of the path graph is*

$$D(\tau_m) \;=\; \sum_{(i,j)\in\tau_m} \frac{q_{ij}^2}{\kappa_{ij}} = \frac{(m-1)^{1+\frac{1}{\gamma}}}{C^{\frac{1}{\gamma}}}$$

and the total negative mixing entropy is

$$NME(\tau_m) \;=\; -\log(m!).$$

Proof. Since the flow along each edge has magnitude 1 the dissipation is

$$\sum_{(i,j)} \frac{q_{ij}^2}{\kappa_{ij}} \;=\; \sum_{i=1}^{m} \frac{q_{i,i+1}^2}{\kappa_{i,i+1}} = \sum_{i=1}^{8} \frac{1}{\kappa_{i,i+1}}$$

We use now use Lagrange multipliers to solve for the minimizing κ_{ij} on the constraint maniforld. We do this by adding the dissipation to a real number λ times the material cost and find the critical points. For $1 \le n < m$ we get

$$\frac{\partial}{\partial \kappa_{n,n+1}} \sum_{i=1}^{m-1} \frac{1}{\kappa_{i,i+1}} + \lambda \sum_{i=1}^{m-1} \kappa_{i,i+1}^{\gamma} \;=\; -\frac{1}{\kappa_{n,n+1}^2} + \gamma\lambda\kappa_{n,n+1}^{\gamma-1}.$$

Thus $\frac{1}{\gamma\lambda} = \kappa_{n,n+1}^{\gamma+1}$ for all $1 \le n < m$. This implies that the conductances along the edges are all equal. Therefore $\kappa_{i,i+1} = \kappa > 0$ for all $1 \le i < m$. And we get $\sum_{i=1}^{m-1} \kappa_{i,i+1}^{\gamma} = (m-1)\kappa^{\gamma} = C$. Therefore $\kappa_{i,i+1} \equiv \left(\frac{C}{m-1}\right)^{\frac{1}{\gamma}}$ and the dissipation is equal to $\frac{(m-1)^{1+\frac{1}{\gamma}}}{C^{\frac{1}{\gamma}}}$.

We now compute the NME of the path graph. Let $i \in \tau_m$ be a node. Then for every up-stream node of i, that is all j such that $\tilde{q}_{ji} > 0$, $\tilde{q}_{ji} = 1$. Note, this includes $\tilde{q}_{ii}$ which we

define to be f_i. Therefore, if the nodes are named $1, 2 \dots m$ as they are in the last argument then if $1 \leq l \leq m$, the distribution of signals at node l is $\tilde{q}_{jl} = 1$ for l distinct nodes. Therefore, the normalized distribution of the species is $\frac{\tilde{q}_{jl}}{\sum_{j:\tilde{q}_{jl}>0} \tilde{q}_{jl}} = \frac{1}{l}$ for l different nodes. Thus the local negative mixing entropy at i is

$$\mathrm{NME}_l \quad = \quad \sum_{j:\tilde{q}_{jl}>0} \frac{1}{l} \log\left(\frac{1}{l}\right) = -\log(l).$$

Repeating this for every node $1, 2, \dots, m$ and summing the NME_l we get

$$\mathrm{NME}(\tau_m) \quad = \quad \sum_{l=1}^{m} -\log(l) = -\log(m!).$$

□

To discuss the graph of θ in detail, for a network of conductances κ_{ij} or of flows q_{ij} we introduce the concept of the line from the network. That is, the line of mixing dissipation cost applied to the network for all c. This is important for discussing the graph of the minimal mixing-dissipation cost and the optimal networks because the order of the intersection, whether one line is first below another and then above another (and visa versa) fully describe the relative preferences for different networks. We have the definition:

Definition 5.3. Define the line from conductance network κ_{ij} as the map $c \mapsto \mathrm{NME}(\kappa_{ij}) + cD(\kappa_{ij})$ and define the line from the network of flows q_{ij} as the map $c \mapsto \mathrm{NME}(q_{ij}) + cD(q_{ij})$.

Because $\theta(c)$ be written as the image of c on the line from the network κ_{ij}which is the minimum out of all possible conductances networks obeying $\sum_{(i,j)} \kappa_{ij}^{\gamma} = C$ this definition comes in handy. Especially when we are studying the results of our simulations. We take the envelope of our results. This is defined as computing the line from each sampled optimum and minimizing over all of them. It gives the tightest upperbound on the graph of $\theta(c)$ due to the concavity of θ.

In [Dur07] the minimal dissipation network on an ambient network with fixed boundary

flows with a fixed material cost was proven to always be a tree.

Theorem 5.2. *The conductance network* κ_{ij} $\sum_{(i,j)} \kappa_{ij}^{\gamma} = C$ $0 < \gamma < 1$ *connecting the sources and sinks which sum to* 0 *which minimizes* $D(\kappa_{ij})$ *is a tree.*

This theorem states that minimally dissipative networks which obey the building cost constraint are trees. Therefore it allows us to frame optimally mixing networks that contain loops as departing in character from minimally dissipative networks obeying Murray's law. This has been a starting point for many studies of biological networks (such as leaf venation) because organisms improve their survival through being as conservative with energy as possible while real networks have loops. Our networks also have loops, so we are interested in characterizing them.

We have next a corollary of this describes why the networks minimizing the mixing-dissipation cost with high weights c are all geodesics.

Theorem 5.3. *On an ambient network with boundary flows being a single source* u *and single sink* v *of magnitude* 1 *(i.e.* $Q_u = 1$, $Q_v = -1$*) the minimally dissipative network compatible with the flows is the path with the least amount of vertices connecting* u *to* v*.*

Now that we have layed down many facts which we repeatedly make use of, we start with new mathematical results about minimal mixing-dissipation cost networks. The first object we look at is what we reference as a back-bone to the curve of mixing-dissipation optima.

5.5 Structure of mixing-dissipation cost minima over set of paths

5.5.1 Definitions needed for studying minimization over tree networks

Definition 5.4. Recall that the minimal dissipation cost is defined

$$\theta(c) \quad = \quad \min_{\sum_{(i,j)} \kappa_{ij}^{\gamma} = C} \mathrm{NME}(\kappa_{ij}) + cD(\kappa_{ij})$$

As a first order approximation, we restict the set of all possible conductance networks to trees:

$$\theta_{\text{tree}}(c) \quad = \quad \min_{\kappa_{ij} \text{ path } : \sum_{(i,j)} \kappa_{ij}^{\gamma} = C} \text{NME}(\tau_m) + cD(\tau_m).$$

In the case of one source and one sink, the set of trees is reduced to the set of paths connecting the source to the sink:

$$\theta_{\text{path}}(c) \quad = \quad \min_{\text{all possible } \tau_m : \sum_{(i,j)} \kappa_{ij}^{\gamma} = C} \text{NME}(\tau_m) + cD(\tau_m).$$

- Borrowing common graph theoretic terms for the important networks in the paper, we refer to the shortest path connecting the source to the sink for the single-source, single-sink boundary flow assumption as the *geodesic* or the *minimal path.*

- If N side length of a square gird is odd or the grid is triangular, then there is a path connecting the source to the sink that contains each of the N^2 nodes. We call this path, if it exists, the *tour.*

- If N is even on square grids, there is no such tour, but there is a path connecting the source to the sink containing $N^2 - 1$ nodes (i.e., all but 1 node). To capture both settings, we call the path with the most nodes connecting the source to the sink the *maximal path.*

Definition 5.5. Let $0 < m < n$ odd integers such that $2N - 1 \leq m < n \leq N^2$. Then the c where the lines from τ_m and τ_n intersect is denoted by c_{nm} or, equivalently, c_{mn}. That is c_{mn} is the positive real such that

$$\text{NME}(\tau_m) + c_{mn}D(\tau_m) \quad = \quad \text{NME}(\tau_n) + c_{mn}D(\tau_n).$$

Fact 5.1. *The intersection point* $c_{mn} = \frac{NME(\tau_m)-NME(\tau_n)}{D(\tau_n)-D(\tau_m)} = C^{\frac{1}{\gamma}}\left(\frac{-\log(m!)+\log(n!)}{(n-1)^{1+\frac{1}{\gamma}}-(m-1)^{1+\frac{1}{\gamma}}}\right)$. *The simplest form of this expression is if we take* $C = 1$ *to obtain* $\frac{-\log(m!)+\log(n!)}{(n-1)^{1+\frac{1}{\gamma}}-(m-1)^{1+\frac{1}{\gamma}}}$.

The preliminary numerical result leads us to formulate the first mathematically rigorous claim of this chapterthat the graph of the minimal mixing-dissipation costs at weights c where the networks being minimized over are path graphs is piece-wise affine, and that as c is increased, from one interval in the domain where the graph is affine to the subsequent interval, the graph switches from the a line segment corresponding to line from a path graph with a fixed number of nodes, to a line with a path graph with fewer nodes. In fact, the minimizer changes from a path graph to the next smallest path graph. The consecutive sizes of path graphs between the same two points on a triangular grid differ by 1. The consecutive sizes of path graphs between the same two points on a square grid differ by 2 nodes because path graphs connecting the same two nodes all have the same parity of number of nodes (and edges).

Fact 5.2. *If* $c = 0$ *then the mixing-dissipation minimizer over paths is the longest possible path connecting source and sink. At the other extreme, the graph of of minimizers over paths is the line from the geodesic connecting source and sink as* $c \to \infty$.

Let M be the maximum number of nodes in a path connecting the source to sink (N^2 if N is odd and $N^2 - 1$ if $N$is even) and $\mu = 2N - 1$ be the minimal number of nodes in a path connected the source to the sink. Since changing C has the same effect as uniformly scaling the weight for the mixing-dissipation cost, c and the next lemma and proposition only depend on the order of the intersection points c_{mn}, we can simplify the proofs of the following statements taking effect for all C by taking the building cost to be $C = 1$ (discussed in the immediately following section). Then we have the proposition describing the graph of θ_{tree} :

Theorem 5.4. *Consider networks of conductances on an ambient network* G *on* N *nodes and boundary flows of one source* $Q_1 = 1$ *and one sink* $Q_N = -1$. *Suppose that the maximal*

path length between source and sink is M. Then the minimizer of the mixing-dissipation cost over paths is the maximal path τ_M for $c \leq c_{M,M-\iota}$, the path of length m for all odd $m > 2N-1$ on $c_{m+\iota,m} \leq c \leq c_{m,m-\iota}$ and the path of shortest length $\mu \geq 2$ on $c_{\mu+2,\mu} \leq c < \infty$. Correspondingly, the graph of minima for the mixing-dissipation costs varying c can be written

$$\theta_{path}(c) = \begin{cases} NME(\tau_M) + cD(\tau_M) & 0 \leq c \leq c_{M,M-\iota} \\ NME(\tau_m) + cD(\tau_m) & c_{m+\iota,m} \leq c \leq c_{m,m-\iota} \text{ for all } \mu+\iota \leq m \leq M-\iota \cdot \\ NME(\tau_\mu) + cD(\tau_\mu) & c_{\mu+\iota,\mu} \leq c \end{cases}$$

Proof. Let $m \in \mathbb{N}$ such that $\mu+\iota \leq m \leq M-\iota$. Let us restrict the cases to when $M \neq \mu+\iota$ because in that case there are only two lines, $\mathrm{NME}(\tau_M) + cD(\tau_M)$ and $\mathrm{NME}(\tau_\mu) + cD(\tau_\mu)$ and since $\mu < M$, $\mathrm{NME}(\tau_M) < \mathrm{NME}(\tau_\mu)$ while $D(\tau_M) > D(\tau_\mu)$. Thus the line from τ_M starts below the line from τ_μ and is the mixing-dissipation cost minimizer from $c = 0$ up to $c = c_{\mu M}$, intersects $\tau_\mu$at a positive $c_{\mu M} > 0$ and for $c > c_{\mu M}$, τ_μ is the minimizer.

By lemma 5.1 $c_{m+\iota,m} < c_{m,m-\iota}$. Assume there exists $l \in \mathbb{N}$ such that $\mu \leq l < m-\iota < m$. Then showing the inequality $\mathrm{NME}(\tau_m) + cD(\tau_m) < \mathrm{NME}(\tau_l) + cD(\tau_l)$ for all $c \in [c_{m+\iota,m}, c_{m,m-\iota}]$ is equivalent to showing that $c < \frac{\mathrm{NME}(\tau_l) - \mathrm{NME}(\tau_m)}{D(\tau_m) - D(\tau_l)} = c_{lm}$ because $D(\tau_m) - D(\tau_l) > 0$. This is the same as $c_{m,m-\iota} < c_{lm} = c_{ml}$. On the other hand, assume that there exists $k \in \mathbb{N}$such that $m < m+\iota < k \leq M$. Then proving $\mathrm{NME}(\tau_m) + cD(\tau_m) < \mathrm{NME}(\tau_k) + cD(\tau_k)$ for all $c \in [c_{m+\iota,m}, c_{m,m-\iota}]$ is equivalent to showing $c > \frac{\mathrm{NME}(\tau_m) - \mathrm{NME}(\tau_k)}{D(\tau_k) - D(\tau_m)} = c_{km}$, which is the same as showing $c_{m,m+\iota} > c_{mk}$. Therefore it suffices to show that the following two statements hold: if there exists $l \in \mathbb{N}$ such that $\mu \leq l < m-\iota < m$ then $c_{m,m-\iota} < c_{ml}$, and if there exists $k \in \mathbb{N}$such that $m < m+\iota < k \leq M$ then $c_{m+\iota,m} > c_{km}$.

This is equivalent to showing for all $m > 2+\iota$ and $2 \leq \lambda \leq \frac{m-2}{\iota}$ that $c_{m,m-\iota} < c_{m,m-\lambda\iota} < c_{m-(\lambda-1)\iota,m-\lambda\iota}$. To see why this is true, let $2 < l < m$ and $\lambda = \frac{m-l}{\iota}$ then by the left inequality $c_{m,m-\iota} < c_{m,m-\lambda\iota} = c_{ml}$. On the other hand we can re-assign $k = m$ and $\lambda = \frac{k-m}{\iota}$ and

the right inequality will imply $c_{km} = c_{k,k-\frac{k-m}{\iota}\iota} < c_{m+\iota,m}$. We pick the upper bound on λ $\frac{m-2}{\iota}$ because this is the value where $c_{m-(\lambda-1)\iota,m-\lambda\iota} = c_{2+\iota,2}$ and we utilize that lemma 5.1$c_{m-\iota,m} > c_{m+\iota,m}$ for all $m \geq 2+\iota$.

We will now prove this. For all i let $a_i = \text{NME}(\tau_{i-\iota}) - \text{NME}(\tau_i)$ when i is a possible path length and 0 otherwise. Similarly, let $b_i = D(\tau_i) - D(\tau_{i-\iota})$ when i is a possible path-length and 0 otherwise. Then $c_{m,m-\iota} = \frac{a_i}{b_i}$. We also have $\text{NME}(\tau_{m-\lambda}) - \text{NME}(\tau_m) = \sum_{i=0}^{\lambda-1} a_{m-t\iota}$, $D(\tau_m) - D(\tau_l) = \sum_{i=0}^{\lambda-1} b_{m-t\iota}$ and $c_{m,m-\lambda\iota} = \frac{\sum_{i=0}^{\lambda-1} a_{m-t\iota}}{\sum_{i=0}^{\lambda-1} b_{m-t\iota}}$.

Let $m > 2+\iota$. We prove that $c_{m,m-\iota} < c_{m,m-\lambda\iota} < c_{m-(\lambda-1)\iota,m-\lambda\iota}$ on the differences $\lambda = 2, 3, \ldots, \frac{m-2}{\iota}$ by induction. For base case let $\lambda = 2$. Then

$$c_{m,m-2\iota} = \frac{a_m + a_{m-\iota}}{b_m + b_{m-\iota}}.$$

We know that $\frac{a_m}{b_m} = c_{m,m-\iota} < c_{m-\iota,m-2\iota} = \frac{a_{m-\iota}}{b_{m-\iota}}$. Thus, by $a_m b_{m-\iota} < a_{m-\iota} b_m$ and so $a_m(b_m + b_{m-\iota}) = a_m b_m + a_m b_{m-\iota} < a_m b_m + a_{m-\iota} b_m = b_m(a_m + a_{m-\iota})$, and $b_{m-\iota}(a_m + a_{m-\iota}) < a_{m-\iota}(b_m + b_{m-\iota})$. Therefore $c_{m,m-\iota} < c_{m,m-2\iota} < c_{m-\iota,m-2\iota}$. Therefore we have the base case.

For the inductive step assume that for some λ $2 \leq \lambda \leq \frac{m-2}{\iota} - 1$, we have $c_{m,m-\iota} < c_{m,m-\lambda\iota} < c_{m-(\lambda-1)\iota,m-\lambda\iota}$. We wish to show that $c_{m,m-\iota} < c_{m,m-(\lambda+1)\iota} < c_{m-\lambda\iota,m-(\lambda+1)\iota}$. By the first inequality

$$\frac{a_m}{b_m} < \frac{\sum_{t=0}^{\lambda-1} a_{m-t\iota}}{\sum_{t=0}^{\lambda-1} b_{m-t\iota}}$$

and by lemma 5.1, since $m - (\lambda+1)\iota \geq 2$ and $m - \lambda\iota \geq 2 + \iota$, $c_{m,m-\iota} < c_{m-\lambda\iota,m-(\lambda+1)\iota}$. Therefore

$$a_m \left(\sum_{t=0}^{\lambda-1} b_{m-t\iota} \right) = a_m \left(\sum_{t=0}^{\lambda-1} b_{m-t\iota} \right) + a_m(b_{m-\lambda\iota}) < b_m \left(\sum_{t=0}^{\lambda-1} a_{m-t\iota} \right) + b_m(a_{m-\lambda\iota}) = b_m \left(\sum_{t=0}^{\lambda} a_{m-t\iota} \right)$$

and so $c_{m,m-\iota} = \frac{a_m}{b_m} < \frac{\sum_{t=0}^{\lambda} a_{m-t\iota}}{\sum_{t=0}^{\lambda} b_{m-t\iota}} = c_{m,m-(\lambda+1)\iota}$.

On the other hand since, $c_{m-\lambda\iota,m-(\lambda+1)\iota} > c_{m-(\lambda-1)\iota,m-\lambda\iota}$the second inequality, $c_{m,m-\lambda\iota} < c_{m-(\lambda-1)\iota,m-\lambda\iota} < c_{m-\lambda t,m-(\lambda+1)t}$

$$\left(\sum_{t=0}^{\lambda} a_{m-t\iota}\right) b_{m-\lambda\iota} = \left(\sum_{t=0}^{\lambda-1} a_{m-t\iota}\right) b_{m-\lambda\iota} + a_{m-\lambda\iota} b_{m-\lambda\iota}$$
$$< \left(\sum_{t=0}^{\lambda-1} b_{m-t\iota}\right) a_{m-\lambda i} + b_{m-\lambda\iota} a_{m-\lambda\iota} = a_{,m-\lambda}\left(\sum_{t=0}^{\lambda} b_{m-t\iota}\right).$$

Which imples that $c_{m,m-(\lambda+1)\iota} < c_{m-\lambda\iota,m-(\lambda+1)\iota}$. This completes the proof. □

Lemma 5.1. *Let $\iota = 1$ or 2. For all $2 + \iota \leq m$ $c_{m,m+\iota} < c_{m,m-\iota}$.*

Proof. We first prove the lemma for $\iota = 1$ and then we use the solution to prove the theorem for $\iota = 2$.

We show that $c_{m+1,m} < c_{m,m-1}$. We have that $c_{m+1,n} = \frac{\log(m+1)}{m^{1+\frac{1}{\gamma}} - (m-1)^{1+\frac{1}{\gamma}}}$ and $c_{m,m-1} = \frac{\log(m)}{(m-1)^{1+\frac{1}{\gamma}} - (m-2)^{1+\frac{1}{\gamma}}}$. To show that

$$c_{m,m-1} = \frac{\log(m)}{(m-1)^{1+\frac{1}{\gamma}} - (m-2)^{1+\frac{1}{\gamma}}} \quad > \quad c_{m+1,m} = \frac{\log(m+1)}{m^{1+\frac{1}{\gamma}} - (m-1)^{1+\frac{1}{\gamma}}} = c_{m+1,m}.$$

We introduce some new functions to simplify the argument. Since $\gamma \leq 1$ we define $t = 1 + \frac{1}{\gamma} \geq 2$. Now we define the function

$$A_t(m) \quad = \quad (m-1)^t - (m-2)^t.$$

We now aim to show that $\frac{\log(m)}{A_t(m)} > \frac{\log(m+1)}{A_t(m+1)}$ for all $m \geq 3$ and all $t \geq 2$. First we show this to be true for all $m \geq 3$ for $t = 2$. This is equivalent to showing

$$A_2(m+1)\log(m) - A_2(m)\log(m+1) \quad > \quad 0$$

for all $m \geq 3$. It suffices to show that this inequality holds at $m = 3$ and that

$$\frac{d}{dm} A_2(m+1)\log(m) - A_2(m)\log(m+1) \quad > \quad 0$$

for all $m \geq 3$. At $t = 2$, $A_2(m) = 2m - 3$, and so

$$A_2(m+1)\log(m) - A_2(m)\log(m+1) \quad = \quad (2m-1)\log(m) - (2m-3)\log(m+1).$$

At $m = 3$this expression is equal to $5\log(3) - 3\log(4) = \log\left(\frac{243}{64}\right) > 0$.

The derivative of this expression is then

$$\frac{d}{dm} A_2(m+1)\log(m) - A_2(m)\log(m+1) \quad = \quad \frac{2m-1}{m} - \frac{2m-3}{m+1} + 2\log\left(\frac{m}{m+1}\right)$$

and this has as a horizontal asymptote $\frac{2m-1}{m} - \frac{2m-3}{m+1} + 2\log\left(\frac{m}{m+1}\right) \to 0$ as $m \to \infty$. To show that this quantity is positive for $m \geq 3$ it suffices to show that it is positive at $m = 3$ and that it only has one critical point on the positive real line (i.e. where the second derivative of $A_2(m+1)\log(m) - A_2(m)\log(m+1)$ is 0) and that it less than 3. This would forbid it from dipping below the x axis after 3 because then it would need to have a critical point to achieve the horizontal asymptote. Evaluating the first derivative at $m = 3$ gives us

$$2\log(m) + \frac{2m-1}{m} - 2\log(m+1) - \frac{2m-3}{m+1} \quad = \quad 2\left(\log(3) - \log(4)\right) + \frac{5}{3} - \frac{3}{4} = 0.3413 > 0.$$

Taking the second derivative

$$\frac{d^2}{dm^2} A_2(m+1)\log(m) - A_2(m)\log(m+1) \quad = \quad \frac{-2m^2 + 4m + 1}{m^2(m+1)^2}.$$

The zeros of the numerator are then $1 \pm \sqrt{\frac{3}{2}}$. The positive root is $1 + \sqrt{\frac{3}{2}} < 3$.This shows

that $\frac{d}{dm}A_2(m+1)\log(m) - A_2(m)\log(m+1) > 0$ for all $m \geq 3$. Therefore

$$c_{n,n-1} - c_{n+1,n} = A_2(m+1)\log(m) - A_2(m)\log(m+1) \quad > \quad 0$$

for all $m \geq 3$ when $t = 2$.

It is now left to show that this is true for $t > 2$. We first show that for all $n \geq 3$

$$\frac{A_t(m+1)}{A_t(m)} = \frac{m^t - (m-1)^t}{(m-1)^t - (m-2)^t} \quad > \quad \frac{m^2 - (m-1)^2}{(m-1)^2 - (m-2)^2} = \frac{A_2(m+1)}{A_2(m)}.$$

This is equivalent to showing

$$\frac{m^t - (m-1)^t}{m^2 - (m-1)^2} \quad > \quad \frac{(m-1)^t - (m-2)^t}{(m-1)^2 - (m-2)^2}.$$

Note that $m^t - (m-1)^t = \int_{m-1}^{m} tx^{t-1}dx$ and $m^2 - (m-1)^2 = \int_{m}^{m+1} 2xdx$. We have that $\min_{m-1\leq x\leq m} \frac{tx^{t-1}}{2x} = \frac{t}{2}(m-1)^{t-2}$ and $tx^{t-1} > 2x \cdot \frac{t}{2}(m-1)^{t-2}$ for every $x \in [m-1, m]$ except for at the left end-point $x = m-1$. Therefore $m^t - (m-1)^t = \int_{m-1}^{m} tx^{t-1}dx > \frac{t}{2}(m-1)^{t-2}\int_{m-1}^{m} 2xdx = \frac{t}{2}(m-1)^{t-2}\left(m^2 - (m-1)^2\right)$. On the interval $m-2 \leq x \leq m-1$, $\max_{n-1\leq x\leq n} \frac{tx^{t-1}}{2x} = \frac{t}{2}(m-1)^{t-2}$ and $tx^{t-1} < 2x \cdot \frac{t}{2}(m-1)^{t-2}$ on for ever x in $[m-2, m-1]$ except for the right end-point. We then have $(m-1)^t - (m-2)^t = \int_{m-2}^{m-1} tx^{t-1}dx < \frac{t}{2}(m-1)^{t-2}\int_{m-2}^{m-1} 2xdx = \frac{t}{2}(m-1)^{t-2}\left((m-1)^2 - (m-2)^2\right)$. These two strict inequalities combine to give us

$$\frac{m^t - (m-1)^t}{m^2 - (m-1)^2} \quad > \frac{t}{2}(m-1)^{t-2} > \quad \frac{(m-1)^t - (m-2)^t}{(m-1)^2 - (m-2)^2}.$$

Now we can finish the proof of the theorem. We have for all $m \geq 3$ and $t > 2$,

$$\begin{aligned} A_t(m)\log(m+1) &= \frac{A_t(m)}{A_2(m)} A_2(m)\log(m+1) \\ &< \frac{A_t(m)}{A_2(m)} A_2(m+1)\log(m) \\ &= A_t(m)\frac{A_2(m+1)}{A_2(m)}\log(m) \\ &< A_t(m)\frac{A_t(m+1)}{A_t(m)}\log(m) = A_t(m+1)\log(m). \end{aligned}$$

This proves the case when $\iota = 1$. This completes the proof for when path lengths are allowed to differ by 1 as in the triangular grid. We now address cases where path lengths are only allowed to differ by 2 as in the square grid. That is, assume $\iota = 2$ and $m \geq 2 + \iota$. First we prove a rudimentary statement about fractions. Assume that a, b, c, d are positive real numbers such that $\frac{a}{b} < \frac{c}{d}$. Then since $ad > bc$, $d(a+c) = da + dc < dc + bc = c(b+d)$. Likewise $a(b+d) = ab + ad < ab + bc = b(a+c)$. Therefore $\frac{a}{b} < \frac{a+c}{b+d} < \frac{c}{d}$.

Now we apply this to prove the bounds $c_{m-1,m-2} < c_{m-2,m} < c_{m,m-1}$. For the intersection of lines from paths with lengths differing by 2 we have

$$\begin{aligned} c_{m-2,m} &= \frac{\log(m) + \log(m-1)}{(m-1)^{1+\frac{1}{\gamma}} - (m-3)^{1+\frac{1}{\gamma}}} \\ &= \frac{\mathrm{NME}(\tau_{m-2}) - \mathrm{NME}(\tau_{m-1}) + \mathrm{NME}(\tau_{m-1}) - \mathrm{NME}(\tau_m)}{D(\tau_{m-1}) - D(\tau_{m-2}) + D(\tau_m) - D(\tau_{m-1})}. \end{aligned}$$

By definition $c_{m-2,m-1} = \frac{\mathrm{NME}(\tau_{m-2}) - \mathrm{NME}(\tau_{m-1})}{D(\tau_{m-1}) - D(\tau_{m-2})}$ and $c_{m-1,m} = \frac{\mathrm{NME}(\tau_{m-1}) - \mathrm{NME}(\tau_m)}{D(\tau_m) - D(\tau_{m-1})}$ and the claim follows from the $c_{m-1,m} < c_{m-2,m-1}$. Now using this ordering twice, we find

$$c_{m+2,m+1} < c_{m,m+2} = c_{m,m+\iota} < c_{m+1,m} < c_{m,m-1} < c_{m-2,m} = c_{m-\iota,m} < c_{m-2,m-1},$$

so $c_{m,m+2} = c_{m,m+\iota} < c_{m-2,m} = c_{m-\iota,m}$ for all $m > 2 + \iota = 4$. We have shown that $c_{m,m+\iota} < c_{m-\iota,m}$ for all $m > 2 + \iota$ and $\iota \in \{1, 2\}$. □

We make use of the following property in the next few theorems. We have found it useful especially when studying networks as the exponent goes to 0. It also shows that changing the amount of material will not change our results in a fundamental way.

5.6 Invariance mixing-dissipation cost curves on choice of C for $\sum_{(i,j)} \kappa_{ij}^{\gamma} = C$

Let $C, C' > 0$. Suppose κ_{ij}obeys the building cost with exponent γ with total material C, $\sum_{(i,j)} \kappa_{ij}^{\gamma} = C$. Then a new network of conductances can be obtained by scaling every conductance by the same constant across the whole network so that it obeys the building cost constraint with the same exponent γ but with a new total material C'. We have the following fact.

Fact 5.3. *Altering the building cost amount of a network is equivalent to rescaling the argument by the ratio of the building costs raised to the $\frac{1}{\gamma}$ power.*

$$\begin{aligned} \theta(c) \textit{ with total material} C' &= \min_{\kappa_{ij} \sum_{(i,j)} \kappa_{ij}^{\gamma} = C'} NME(\kappa_{ij}) + cD(\kappa_{ij}) \\ &\textit{is the same as} \\ \theta\left(\left(\frac{C}{C'}\right)^{\frac{1}{\gamma}} c\right) \textit{ with total material} C &= \min_{\kappa_{ij} \sum_{(i,j)} \kappa_{ij}^{\gamma} = C} NME(\kappa_{ij}) + \left(\frac{C}{C'}\right)^{\frac{1}{\gamma}} cD(\kappa_{ij}). \end{aligned}$$

Essentially, the weight c is multiplied by the ratio of the second building cost to the first building cost.

We get this via $\tilde{\kappa}_{ij} = \left(\frac{C'}{C}\right)^{\frac{1}{\gamma}} \kappa_{ij}$ for all edges (i,j). Then $\sum_{(i,j)} \tilde{\kappa}_{ij}^{\gamma} = \sum_{(i,j)} \frac{C'}{C} \tilde{\kappa}_{ij}^{\gamma} = C'$. Now we turn to study the effect this has on the dissipation of κ_{ij} and $\tilde{\kappa}_{ij}$. Note that κ_{ij} and $\tilde{\kappa}_{ij}$ have the same physical flows since by Thomson's principle if q_{ij} is the physical flow of κ_{ij} then it is compatible with the boundary flows and it minimizes $\sum_{(i,j)} \frac{q_{ij}^2}{\kappa_{ij}}$. Considering $\hat{q}_{ij}$ a set of flows compatible with the boundary flows, the dissipation by the network of conductances κ_{ij} is $\sum_{(i,j)} \frac{\hat{q}_{ij}^2}{\kappa_{ij}}$ and the dissipation on the network of conductances $\tilde{\kappa}_{ij}$ is

$\sum_{(i,j)} \frac{\hat{q}_{ij}^2}{\tilde{\kappa}_{ij}} = \left(\frac{C}{C'}\right)^{\frac{1}{\gamma}} \sum_{(i,j)} \frac{\hat{q}_{ij}^2}{\kappa_{ij}}$. That is the dissipation by $\tilde{\kappa}_{ij}$ as a function on the set of compatible flows is a positive scalar multiple of the dissipation by κ_{ij} as a function on the set of compatible flows. Therefore, the network of compatible flows minimizing both functions is the same– i.e., the physical flows over both networks is the same. Further more $D(\tilde{\kappa}_{ij}) = \left(\frac{C}{C'}\right)^{\frac{1}{\gamma}} D(\kappa_{ij})$.

We now understand the effect this has on the mixing-dissipation cost. We have that the mixing-dissipation cost for weight c at κ_{ij} is $\mathrm{NME}(\kappa_{ij}) + cD(\kappa_{ij})$ and for the same c at $\tilde{\kappa}_{ij}$ we have that

$$\begin{aligned} \mathrm{NME}(\tilde{\kappa}_{ij}) + cD(\tilde{\kappa}_{ij}) &= \mathrm{NME}(\tilde{\kappa}_{ij}) + c\left(\frac{C}{C'}\right)^{\frac{1}{\gamma}} D(\kappa_{ij}) \\ &= \mathrm{NME}(\kappa_{ij}) + c\left(\frac{C}{C'}\right)^{\frac{1}{\gamma}} D(\kappa_{ij}). \end{aligned}$$

The second equality is due to the fact that NME is a function on the physical flows for the network of conductances in the argument. This shows us that the mapping $c \mapsto \mathrm{NME}(\tilde{\kappa}_{ij}) + cD(\tilde{\kappa}_{ij})$ is the same as the mapping $c \mapsto \mathrm{NME}(\kappa_{ij}) + cD(\kappa_{ij})$ except the argument is scaled by a constant factor $\left(\frac{C}{C'}\right)^{\frac{1}{\gamma}}$.

Fact 5.4. *A simple important consequence is that that for an increasing sequence real numbers $a_0 = 0 < a_1 < a_2 < \cdots < a_n$ if there is any mathematical characterization of the minimizers or statements about the continuity, the monotonicity, the existence or sign of the derivative of θ at $a_i$or on each interval endpoints a_i, a_{i+1} (closed, left/right half-closed or open interval included) then the same statements hold for $\left(\frac{C}{C'}\right) a_0 = 0 < \left(\frac{C}{C'}\right) a_1 < \left(\frac{C}{C'}\right) a_2 < \cdots < \left(\frac{C}{C'}\right) a_n$and the intervals with endpoints of the same indices in this sequence.*

The work in this paper is about qualities of the mixing-dissipation cost at different building cost exponents γ which are unchanged by altering the building cost. The set minimizers of the mixing-dissipation cost depending on the relative magnitudes of c, and the shape of the curve in terms of whether it is piece-wise linear/ whether there are corners/

or rounded corners are all unchanged with switching the building cost. A specific example interection c of lines from τ_m and τ_n c_{mn} 5.5, are also scaled by $\left(\frac{C'}{C}\right)^{\frac{1}{\gamma}}$ when changing building cost from C to C'. Therefore the study of the mixing-dissipation cost and optimal networks at different building cost exponents γ can be carried out by first choosing a fixed material cost. The advantage is that the material cost can be chosen to be the one which facilitates logic, computation and intuition. Perhaps the most important way we use it is studying the dissipations of networks in which the number of positive flows differs by at most 1.

This is important in proving that networks with loops are removed as $\gamma \to 0^+$. To understand why consider three flow networks G_1, G_2, G_3 with $\mathrm{NME}(G_1) < \mathrm{NME}(G_2) < \mathrm{NME}(G_3)$ and the number of edges of G_1, G_2 and G_3 are $10, 12, 11$ respectively with $G_1 = \tau_{11}$ and $G_3 = \tau_{12}$. That is, we are studying an intermediate loopy network between two path networks differing by one node. We need to pick the same C for each network. Picking an arbitrary C, say 1 gives us that $D(G_1) = (10)^{1+\frac{1}{\gamma}}$,$D(G_2) \geq \left(\sum_{(i,j)} \left(q_{ij}^2\right)^{\frac{\gamma}{\gamma+1}}\right)^{1+\frac{1}{\gamma}}$ and $D(G_3) = (11)^{1+\frac{1}{\gamma}}$ each of these approaches ∞ as $\gamma \to 0$, so we cannot draw any conclusions. If we set $C = 11$, the number of edges in the larger of the paths, then we find $D(G_1) = 10(\frac{10}{11})^{\frac{1}{\gamma}}$,$D(G_2) \geq \left(\sum_{(i,j)} \left(q_{ij}^2\right)^{\frac{\gamma}{\gamma+1}}\right)\left(\frac{\sum_{(i,j)}\left(q_{ij}^2\right)^{\frac{\gamma}{\gamma+1}}}{11}\right)^{\frac{1}{\gamma}}$ and $D(G_3) = (11)^1$ using the lower bound we are about to prove in theorem 5.5. It is not difficult to see that $D(G_1) \to 0$, a flat line, $D(G_2) \to \infty$ and $D(G_3) = 11$. Therefore we expect the line from G_2 to intersect the line from G_1 before that of G_3 does for all γ small enough. We repeatedly use this idea in analyzing the behavior of optimal networks as $\gamma \to 0^+$.

5.7 Lower bound on dissipation D in terms of flows q_{ij}

Theorem 5.5. *Let q_{ij}be a network of flows compatible with the the boundary flows Q_i then we have the following lower bound on the dissipation for conductances κ_{ij} with q_{ij} for physical*

flows that obey the building cost constraint $\sum_{ij} \kappa_{ij}^{\gamma} = C$

$$\min \left\{ \sum_{(i,j)} \frac{q_{ij}^2}{\kappa_{ij}} : \kappa_{ij} \geq 0 \text{ with physical flow } q_{ij} \text{ with } \sum_{(i,j)} \kappa_{ij}^{\gamma} = C \right\} \geq \frac{\left(\sum_{(i,j)} \left(q_{ij}^2 \right)^{\frac{\gamma}{\gamma+1}} \right)^{1+\frac{1}{\gamma}}}{C^{\frac{1}{\gamma}}}.$$

NB: This is a different way of writing Murray's law.

Proof. Fixing the flows q_{ij} we minimize the dissipation $\sum_{(i,j)} \frac{q_{ij}^2}{\kappa_{ij}}$ over conductances κ_{ij} which obey the building constraint $\sum_{(i,j)} \kappa_{ij}^{\gamma} = C$. Note: we are not assuming that q_{ij} is the physical flow associated to κ_{ij}. For a set of conductances $\tilde{\kappa}_{ij}$ to minimize the dissipation while holding the building constraint constant, it must be a critical point of the function over the level surface given by the building constraint. This is due to two facts. The first is at the boundary of the constraint manifold at least one of the conductances is 0, so the dissipation $\to \infty$ as the conductances approach a point in the boundary of the region of non-negativity, and the approach is uniform in the sense that , we can find a distance such that points within that distance of the boundary are at least as large as a given large number. The second follows from this in that there is a compact neighborhood in the feasible region for which outside of which (and on the boundary), the dissipation is much larger than some chosen attainable dissipation, which is greater than or equal to the minimal dissipation. And on the interior of this compact region, the function attains a minimum. We use the method of Lagrange multipliers to find a critical point via finding $\tilde{\kappa}_{ij}$ such that for some Lagrange multiplier $\lambda \neq 0$

$$\frac{\partial}{\partial \kappa_{ab}} \left(\sum_{(i,j)} \frac{q_{ij}^2}{\kappa_{ij}} + \lambda \sum_{(i,j)} \kappa_{ij}^{\gamma} \right) = 0$$

for all edges (a, b). From this we get that $\frac{q_{ij}^2}{\tilde{\kappa}_{ij}^2} = \lambda(\gamma - 1)\tilde{\kappa}_{ij}^{\gamma-1}$.This implies the equations $q_{ij}^2 = T\tilde{\kappa}_{ij}^{\gamma+1}$ and $\frac{q_{ij}^2}{\tilde{\kappa}_{ij}} = T\tilde{\kappa}_{ij}^{\gamma}$ where we take $T = \lambda(\gamma - 1)$. We get that $\sum_{(i,j)} \frac{q_{ij}^2}{\tilde{\kappa}_{ij}} = T \sum_{ij} \tilde{\kappa}_{ij}^{\gamma} = TC$ so $T = \frac{\sum_{(i,j)} \frac{q_{ij}^2}{\tilde{\kappa}_{ij}}}{C}$, the ratio of the minimal dissipation and the building cost. From $q_{ij}^2 = T\tilde{\kappa}_{ij}^{\gamma+1}$ we

get $\tilde{\kappa}_{ij} = \left(\frac{q_{ij}^2}{T}\right)^{\frac{1}{\gamma+1}}$. Thus $\sum_{(i,j)} \frac{q_{ij}^2}{\tilde{\kappa}_{ij}} = \sum_{(i,j)} \frac{q_{ij}^2}{\left(\frac{q_{ij}^2}{T}\right)^{\frac{1}{\gamma+1}}} = \frac{\left(\sum_{(i,j)} \frac{q_{ij}^2}{\tilde{\kappa}_{ij}}\right)^{\frac{1}{\gamma+1}}}{C^{\frac{1}{\gamma+1}}} \sum_{(i,j)} \left(q_{ij}^2\right)^{\frac{\gamma}{\gamma+1}}$ and so $\left(\sum_{(i,j)} \frac{q_{ij}^2}{\tilde{\kappa}_{ij}}\right)^{\frac{\gamma}{\gamma+1}} = \frac{\sum_{(i,j)} \left(q_{ij}^2\right)^{\frac{\gamma}{\gamma+1}}}{C^{\frac{1}{\gamma+1}}}$. From this we get the only possible critical value $\sum_{(i,j)} \frac{q_{ij}^2}{\tilde{\kappa}_{ij}} = \frac{\left(\sum_{(i,j)} \left(q_{ij}^2\right)^{\frac{\gamma}{\gamma+1}}\right)^{1+\frac{1}{\gamma}}}{C^{\frac{1}{\gamma}}}$.

Since $\frac{\left(\sum_{(i,j)} \left(q_{ij}^2\right)^{\frac{\gamma}{\gamma+1}}\right)^{1+\frac{1}{\gamma}}}{C^{\frac{1}{\gamma}}}$ is the minimum of the dissipation for a given set of flows over all positive κ_{ij} obeying the building cost constraint it must hold that $\frac{\left(\sum_{(i,j)} \left(q_{ij}^2\right)^{\frac{\gamma}{\gamma+1}}\right)^{1+\frac{1}{\gamma}}}{C^{\frac{1}{\gamma}}} \leq \sum_{(i,j)} \frac{q_{ij}^2}{\kappa_{ij}}$ for κ_{ij}obeying the building cost constraint with q_{ij} as its physical flows. □

5.8 Networks with optimal NME are paths

Proposition 5.1. *Consider networks of flows which are physical flows corresponding to a network of conductances and compatible with boundary flows Q_i such that $Q_1 = +1$ and $Q_{N^2} = -1$ and all other boundary flows are zero. Suppose that there are m nodes in a network. Then the lowest possible attainable total negative mixing entropy is $-\log(m!)$. And the unique network on m nodes that attains this minimum is the path through the n nodes starting at the source and ending at the sink*

Proof. Consider a network of flows compatible with the boundary conditions q_{ij}. For a node i we say that j is upstream from i if $\tilde{q}_{ji} > 0$. Consider $V_k \subset \mathcal{N}$ a subset of nodes in the network such that there are k upstream nodes in $\mathcal{N}$ of i for all $i \in V_k$. That is $V_k = \{i \in \mathcal{N} : |\{j \in \mathcal{N} : \tilde{q}_{ji} > 0\}| = k\}$. Let $u \notin V_k$ $v \in \mathcal{N}$. Assume that $\tilde{P}_{uv} > 0$. Then $\tilde{P}_{vu} = 0$ because there are no loops in a flow physical flow. Therefore $\{w \in \mathcal{N} : \tilde{q}_{wu} > 0\} \subsetneq \{w \in \mathcal{N} : \tilde{q}_{wv} > 0\}$ because $v \in \{w \in \mathcal{N} : \tilde{q}_{wv} > 0\}$ while $v \notin \{w \in \mathcal{N} : \tilde{q}_{wu} > 0\}$ contradicting the assumption that $\{w \in \mathcal{N} : \tilde{q}_{wu} > 0\}$ and $\{w \in \mathcal{N} : \tilde{q}_{wv} > 0\}$ contain the same number of

elements. Consider the flow through V_k defined as

$$P(\text{hit } V_k|\text{start at node 1}) \;=\; P\left(\bigvee_{i\in V_k} \text{hit } i|\text{start at node 1}\right)$$

which by principle of inclusion and exclusion is equal to

$$\begin{aligned} P(\text{hit } V_k|\text{start at node 1}) = &\sum_{i\in V_k} P(\text{hit } i|\text{start at node 1}) \\ &+ (-1)^2 \sum_{i_1 i_2 \in V_k} P(\text{hit } i_1, i_2|\text{start at node 1}) + \cdots \\ &+ (-1)^{|V_k|} P(\text{hit } i \forall i \in V_k|\text{start at node 1}). \end{aligned}$$

Note that$P(\text{hit } i_1, i_2, \ldots, i_n|\text{start at node 1}) < \sum_s \sum_{t\neq s} P(\text{hit } i_t \text{ after hitting } i_s|\text{start at node 1}) = \sum_{s=1}^{n} P(\text{hit } i_s|\text{start at node 1}) \sum_{1\leq t\leq n, t\neq s} P_{i_s i_t} = 0$ since we have shown that no node in V_k is an up-stream node of another node in V_k. Thus

$$P(\text{hit } V_k|\text{start at node 1}) = \sum_{i\in V_k} P(\text{hit } i|\text{start at node 1}) = \sum_{i\in V_k} f_i < 1.$$

This fact can be understood as saying the flow through V_k is partitioned over the nodes in V_k. Let $i \in V_k$. Then the largest negative value that the local NME at i can take on is at the distribution of up-stream nodes contributing to node i is uniform. That is

$$\text{NME}_i = \sum_{j,\tilde{q}_{ji}>0} \frac{f_j q_{ji}}{\sum_{l,\tilde{q}_{li}>0} f_l q_{li}} \log\left(\frac{f_j q_{ji}}{\sum_{l,\tilde{q}_{li}>0} f_l q_{li}}\right) > \sum_{j,\tilde{q}_{ji}>0} \frac{1}{k} \log\left(\frac{1}{k}\right) = -\log(k).$$

Also, since the nodes in V_k are such that $\sum_{i\in V_k} f_i < 1$, we can bound the NME below by the case that a flow of 1 passes through each non-emptyV_k and the maximum number of k with $|V_k| > 0$ which is m. Precisely, combining the lower bound on NME_i for each i and picking the arrangement of up-stream nodes for each node to give the best possible mixing entropy

we have

$$\begin{aligned}\mathrm{NME} = \sum_i f_i \mathrm{NME}_i &= \sum_{k, V_k \neq \emptyset} \sum_{i \in V_k} f_i \mathrm{NME}_i \\ &\geq \sum_{k, V_k \neq \emptyset} -\log(k) \\ &\geq \sum_{k=1}^{m} -\log(k) = -\log(m!).\end{aligned}$$

Having proven Theorem 5.1 we know that the path on m nodes attains this value of NME. We now prove uniqueness of the solution. Suppose that q_{ij} is a network which minimizes the NME as a single source and sink both of magnitude 1.

We first show that containing a path of length m is equivalent to every predecessor set being non-empy. Suppose that for q_{ij} and some $1 \leq k \leq m$, $V_k = \emptyset$ and q_{ij} contains a path of length m . Let x_k be the k-th node in the path. Then it has at least k predecessors. The other $m - k$ vertices have yet to be visited along the path of length m so none of them are predecessors to x_k. Therefore, x_k has precisely k predecessors. This proves one half of the claim. Now suppose that each V_k has at least one element. Let $x_1, x_2, \ldots, x_n$ be a selection of nodes such that $x_i \in V_i$ for each $1 \leq i \leq n$. Suppose there is a V_k such that there exists $y \in V_k$ $y \neq x_k$. Then we know that neither elements in $V_k$are predecessors of each other and this contradicts that there are only n nodes in the network. Let j be the smallest $j \in 1, 2, \ldots, n-1$ such that there is no edge flowing from x_j to x_{j+1}. In order of $x_{j+1}$to have j predecessors, not including itself, it must have one of x_k for $k > j+1$ as a predecessor because there only $j-1$ x_i $i < j$. But then x_{j+1} would have at least as many as $k+1$ predecessors.

If some predecessor set V_k in q_{ij} is empty then the NME is bounded below $-\log(m!) + \log(k)$ by the above equation and is therefore not a minimizer. Therefore we conclude that $q_{ij}$is a path and each V_k contains a single node x_k. Since we know that the unique minimum of the negative entropy on probability spaces with k atoms is $-\log(k!)$, and $0 \leq f_{x_k} \leq 1$ we

know that $f_{x_k}\text{NME}_{x_k} \geq \text{NME}_{x-k} \geq -\log(k)$ for all $1 \leq k \leq n$. For the the first non-strict inequality to be an equality, it is necessary that $f_i = 1$ for all k. That implies that there are no flow bifurcations, and so q_{ij} is necessarily a path. □

5.9 Optimal mixing-dissipation cost and networks as $\gamma \to 0^+$

We aim to describe the set of flow topologies that will be minimizers of the mixing-dissipation cost at various weights c as γ decreases to 0. It is important to note that we prove this specifically for ambient networks where every length of path between source and sink between the minimum and maximum lengths are possible. The example that we study in our experiments for which this is true is the triangular grid. An example where this is not the case, as we know, is the square grid. All paths between two points must have lengths of the same parity for the square grid. In the last section we partially address this case, but cannot conclude that we have the same theorem for square grids.

The first theorem we prove, theorem 5.6 shows that networks which differ by more than pre-determined amount from an optimal level of mixing are uniformly removed from consideration of being optimal as $\gamma \to 0^+$. If we study path lengths $m-1$, m and $m+1$ then the optimal NME of the networks are $-\log((m-1)!)$,$-\log(m!)$ and $-\log((m+1)!)$ respectively. We show that for every positive "error" ϵ there exists a γ_ϵ such that the following is true. If the flow network q_{ij} is such that $\text{NME}(q_{ij})$ is bounded ϵ away from these three optimal values, then that network is suboptimal on the interval $(c_{m+1,m}, c_{m,m-1})$ for all $\gamma < \gamma_\epsilon$.

By Proposition 5.5. For a flow network q_{ij} we can bound the dissipation below by $D(q_{ij}) > \frac{\left(\sum_{(i,j)}\left(q_{ij}^2\right)^{\frac{\gamma}{\gamma+1}}\right)^{1+\frac{1}{\gamma}}}{C^{\frac{1}{\gamma}}}$. The sum converges to the number of non-zero flows on the network. Therefore we know that if the network has more non-zero flows than the optimally mixing network on $m+1$ nodes, which we know to be the path with m flows then its dissipation and therefore cost grows to be infinitely larger than the network on $m+1$ flows.

We use this to show networks with a lower magnitude of NME and more edges are eliminated.

Along with proving that, for γ small, networks with NME bounded away from optimal NME on a given number of nodes are eliminated, we show that the remaining networks are very close to path networks. By this we mean, since we do not show that on the interval $[c_{m+1,m}, c_{m,m-1}]$ that networks with $\text{NME}(q_{ij}) \in [-\log(m!)-\epsilon, -\log(m!)+\epsilon]$ are eliminated, we cannot precisely conclude that we what are left with are paths, but if ϵ is very small then the remaining flow networks differ from path networks by a very small amount. When we say a path network is preferred at a c in the interior of $[c_{m+1,m}, c_{m,m-1}]$, we are actually referring to the preference to an approximate path, as we cannot know for sure if it is preferred.

Theorem 5.6. *Let G be an ambient network with N nodes, which we refer to as $1, 2, \ldots, N$. Let Q_i be boundary flows with $Q_1 = 1$ and $Q_N = -1$ and $Q_i = 0$ for all $i \neq 1, N$. Let $1 < m < N$ be such that there exists undirected paths of lengths m, $m-1$ and $m+1$ using distinct nodes and edges in G. Let $\frac{\log(m)}{2} > \epsilon > 0$. Let $C > 0$. Consider conductance networks on G κ_{ij} such that $\sum_{(i,j)} \kappa_{ij}^{\gamma} = C$. Then there exists γ_ϵ such that if $\gamma < \gamma_\epsilon$ then if q_{ij} is a flow network on G such that NME$(q_{ij}) \notin$*

$$(-\log((m+1)!)-\epsilon, -\log((m+1)!)+\epsilon) \cup (-\log(m!)-\epsilon, -\log(m!)+\epsilon) \cup (-\log((m+1)!)-\epsilon, -\log((m+1)!)+\epsilon)$$

then

$$\text{NME}(q_{ij}) + cD(q_{ij}) \quad > \quad \text{NME}(\tau_m) + cD(\tau_m)$$

for all $c > c_{m+1,m}$ (in particular $c_{m+1,m} < c < c_{m,m-1}$).

Proof. We split the proof into 4 cases in Section 5.9.1 after which the rest of this section is dedicated to describing concepts needed for this proof until finally, in Section 5.9.6 we finish off the proof for all 4 cases. □

This last theorem does not dictate whether networks with $\text{NME} \in (-\log(m!)-\epsilon, -\log(m!)+$

ϵ) remain or are removed. We do not prove a theorem about which networks in this range become sub-optimal, with the exception of applying theorem 5.6 to $\frac{\epsilon}{2}$ (or any positive number smaller than ϵ) as the new distance from the path optima. These networks can potentially persist through out the interval $[c_{m+1,m}, c_{m,m-1}]$. We can state rigorously that flow networks which differ very little from the paths remain. This is the content of theorem 5.7.

Theorem 5.7. *Let G be an ambient network with N nodes and assume that the boundary flows Q_i are such that $Q_1 = 1$ and $Q_N = -1$ and $Q_i = 0$ for all $i \neq 1, N$. Let $1 \leq m \leq N$ be such that there exists an undirected path of length m in G . Let $\epsilon_1 > 0$. Then there exists $\epsilon > 0$ such that if q_{ij} is a flow on G with $-\log(m!) - \epsilon < NME(q_{ij}) < -\log(m!) + \epsilon$ then there exists a path of length m with flows 1 such, $\tau_m = \sigma_{ij}$ such that $\min_{(i,j)\in G} |q_{ij} - \sigma_{ij}| < \epsilon_1$ where the minimum is over all of the edges in the ambient network G.*

Proof. We prove this in Section 5.9.3. □

5.9.1 Proof of the removal of paths bounded away from NME optima on $m-1$, m and $m+1$ nodes. (Theorem 5.6)

5.9.1.1 Cases and proof strategy

We break this theorem into four parts corresponding to the four connected components in the complement of $(-\log((m+1)!) - \epsilon, -\log((m+1)!) + \epsilon) \cup (-\log(m!) - \epsilon, -\log(m!) + \epsilon) \cup (-\log((m+1)!) - \epsilon, -\log((m+1)!) + \epsilon)$. Specifically these are:

Case	Connected interval
A	$(-\log(m!) + \epsilon, -\log((m-1)!) - \epsilon)$
B	$(-\log((m+1)!) + \epsilon, -\log(m!) - \epsilon)$
C	$(-\infty, -\log((m+1)!) - \epsilon)$
D	$(-\log((m-1)!) + \epsilon, +\infty)$

Case A:

Assume $-\log(m!) + \epsilon < \mathbf{NME}(q_{ij}) < -\log((m-1)!) - \epsilon$. The proofs for cases A and B are similar. We will use the fact that the optimal NMEon m nodes is $-\log(m!)$. Since we are exceeding that by an amount atleas as large as ϵ, we can conclude that there are at least $m+1$ nodes with flows above a level that guarentees leaving the remaining nodes out of the calculation of the NME will barely affect it. If there are more than $m+1$ nodes with large enough f_i, then there must be at least $m+1$ edges with large flows. We can show that for having larger magnitude NME than the path τ_m while being sub-optimal to τ_{m+1} that if there are $m+1$ nodes with large enough f_i then there must be a cycle.

Case B:

Assume $-\log((m+1)!) + \epsilon < \mathrm{NME}(q_{ij}) < -\log(m!) - \epsilon$ we use the same argument as for case A except we shift by adding one extra node to the paths on the boundary.

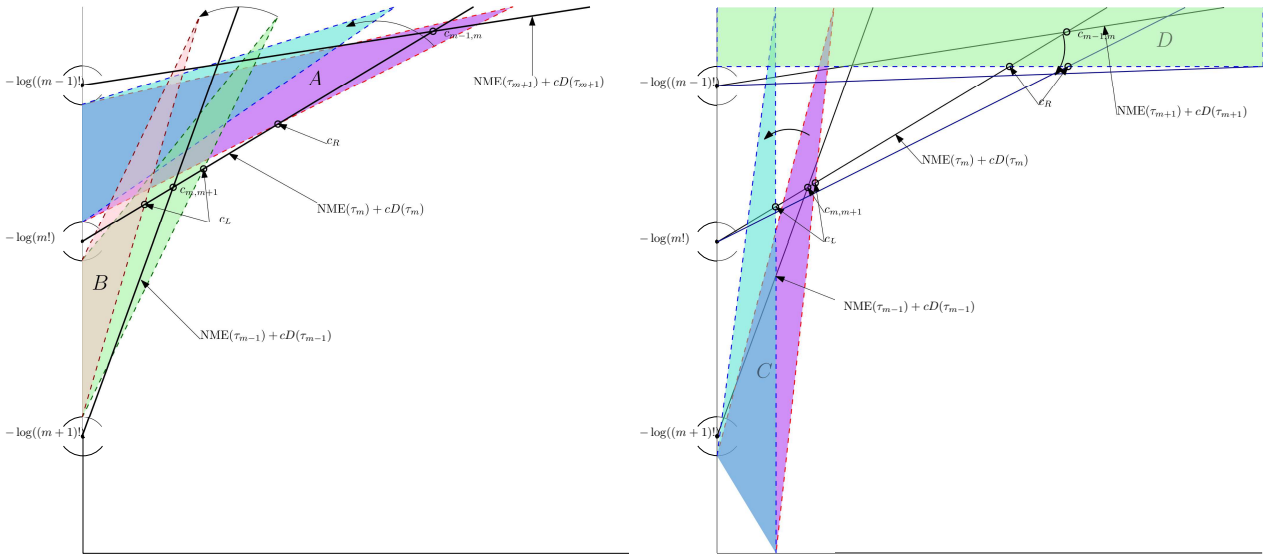
We prove both cases A and B by proving B first in Section

Case C:

Assume that $\mathrm{NME}(q_{ij}) < -\log((m+1)!) - \epsilon$. We use a similar technique to A and B to show that since the NME is a larger negative number than $\mathrm{NME}(\tau_{m+1})$ by ϵ then there must be at least $m+1$ flows, which causes the dissipation to grow at a rate infinitely faster than networks which are optimal on $[c_{m+1,m}, c_{m,m-1}]$.

Case D:

Assume that $\mathrm{NME}(q_{ij}) > -\log((m-1)!) + \epsilon$. For this case we recognize that the mixing dissipation cost of $c_{m,m-1}$ on lines from τ_m and τ_{m-1} drops below any number with fixed negative value above $-\log((m-1)!)$.



(a) (b)

Figure 5.9.1: Figure 5.9.1a is a diagram illustrating cases A and B of our proof. The black lines are lines from path networks with $m-1$, m and $m+1$ nodes. Finding flows uniformly bounded below allows us to apply the lower bound on the dissipation. The lower bound on the dissipation geometrically transforms the plot of θ as $\gamma \to 0^+$ by magnifying the relative magnitude of slopes in the regions labeled A (resp. B) to the slope of the line from τ_m (resp. τ_{m+1}). This is represented by the arsc with arrows showing a steepening of these regions. Also labeled are points illustrating the concept of c_L and c_R the left and right transition points. They appear as the supremum of the triangular region representing B and its intersection with the line from τ_m and infimum of the triangular region representing case A and its intersection with τ_m. There are still cases C and D, so these are not exact representations of the transition points. In figure 5.9.1b we have a similar legend. Now the boundary cases C and D are labeled. Case D is proven by showing that as $\gamma \to 0^+$ the mixing dissipation costs on τ_m and τ_{m-1} shift so that $c_{m,m-1}$ falls below $\mathrm{NME}(\tau_{m-1})$ for all small γ. Region C follows a similar proof as the removal of networks in cases B and D by showing that the dissipation grows infinitely faster than that of paths.

Before we dig into the proof of this theorem we discuss the fall back of the triangular regions B and D past $c_{m+1,m}$ to the left and the removal of the regions A and C upwards past $c_{m,m-1}$ in figure 5.9.1. Visual inspection can be helpful to understand when each case A, B, C or D are eliminated from the set of optimal networks. Their shrinking overlap with the frontier of minima shows how cut corners are gradually made to sharp transitions. Note how the intersections of the right sides of the regions B and D with the line from τ_m

indicate whether the networks in those components are optimal. The graphic indicates that the intersections of B and D with τ_m recede to the left, and once they fall behind $c_{m+1,m}$ networks from B and D are no longer optimal. On the other hand, note that the intersections of the bottom and right sides of A and C with the line from $\tau_m$indicate their optimality. As $\gamma \to 0$ the intersections move upwards and to the right until they are tucked behind $c_{m,m-1}$. We prove the following lemma that we invoke in section 5.9.6 to show that the cut corners are removed at a rate uniform over all networks dependent only on the values m and ϵ.

Lemma 5.2. *Let q_{ij} and ν_{ij} be flow networks on the ambient network G such that NME$(q_{ij}) <$ NME(τ_m) and NME$(\nu_{ij}) >$ NME(τ_m). Let c_L be the intersection of the lines from q_{ij} and τ_m and assume $c_L > 0$. Let c_R be the intersection of the lines from ν_{ij} and τ_m and assume $c_R > 0$. If $c_L < c_{m+1,m}$ then q_{ij} is not a mixing-dissipation cost minimizer in $[c_{m+1,m}, c_{m,m-1}]$ and if $c_R > c_{m,m-1}$then ν_{ij} is not an mixing-dissipation cost minimizer on $[c_{m+1,m}, c_{m,m-1}]$.*

Proof. Note that the assumption $c_L > 0$ is equivalent to $D(q_{ij}) > D(\tau_m)$. This, we will see in our applications is satisfied for small γ. For ν_{ij} we do not have to worry about $D(\nu_{ij}) \geq D(\tau_m)$ because then the line from ν_{ij} will always lie above the line from τ_m therefore we assume $D(\nu_{ij}) < D(\tau_m)$. Since the line from q_{ij} starts below the line from τ_m and has a steeper slope, for all $c > c_L$, $\mathrm{NME}(q_{ij}) + cD(q_{ij}) > \mathrm{NME}(\tau_m) + cD(\tau_m)$. In particular if $c_L < c_{m+1,m}$ then this inequality holds on all of $c > c_{m+1,m}$. On the other hand, since $\mathrm{NME}(\nu_{ij}) > \mathrm{NME}(\tau_m)$ we know that the line from ν_{ij} lies above the line from $\tau_m$up until they intersect. Therefore, if their intersection is greater than $c_{m,m-1}$, the line from ν_{ij} will be above the line from τ_m on the interval $[c_{m+1,m}, c_{m,m-1}]$. □

5.9.2 The main influences on NME: Strong Nodes

As we have stated in the synopsis of this section, a core ingredient in this proof is to find more flows which are uniformly bounded below than the number of flows in the next largest path. The summands contributing to the NME, the local negative mixing entropies, are

dependent on the relative contributions of flows from up-stream nodes. There is less emphasis on the absolute size of the total flows because the entropy of the distribution of flows *after* normalized to sum to one are computed. Therefore, instead of bounding absolute size, we bound the maximum effect that a set of nodes can have on a summand $f_i \text{NME}_i$. Therefore we define the concept of a dominance factor, a multiplicative constant for which a set of nodes with low flow fall below the set of nodes with greater flow.

Definition 5.6. Let q_{ij} be a network of flows with with N nodes. Order the nodes non-zero f_i in decreasing order of of total flow. That is nodes $1, 2, \ldots, n$ such $f_1 \geq f_2 \geq \cdots \geq f_n$. Let $0 \leq \delta \leq 1$. For our definition we call this number the **dominance factor**. Then the **set of strong nodes above the dominance factor** δ is the set of nodes $1, 2, \ldots, i$ where i is the first index such that $\frac{f_{i+1}}{f_i} < \delta$. When the choice of δ is clear of context we simply refer to $\mathcal{F}_\delta$ as the set of strong nodes.

Fact 5.5. *Alternatively, the set of strong nodes above dominance factor δ can be taken to be the intersection of all sets of nodes $\mathcal{F}$ such that for all $i \in \mathcal{F}$and $j \notin \mathcal{F}$, $f_i > \delta f_j$.*

Fact 5.6. *The set of strong nodes is not barred from being every node i with non-zero f_i. On the other hand the set of strong nodes always contains the nodes with the largest total flow. Therefore it is never empty.*

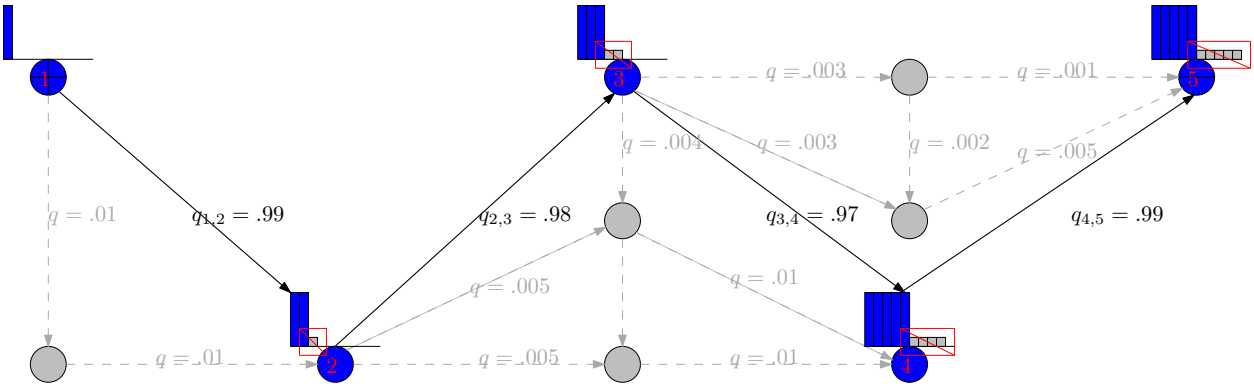


Figure 5.9.2: A flow network on 10 nodes. Node 1 is the source and node 5 is the sink. This is an illustration of the strong-node and dominance factor concept. The blue nodes represent the strong nodes in the network. The grey nodes represent the weak nodes below a dominance factor of .02. Their total flows f_i range from .003 to .01. The total flows of the strong nodes, numbered $1, 2, 3, 4, 5$ are approximately 1. The grey bars in the histograms represent the contribution of weak nodes to these distributions.

To make use of the notion of the set of strong nodes to form a lower bound on the magnitudes of flows, we introduce a generalizations of the total negative mixing entropy and the local negative mixing entropy. We re-iterate that for a flow network q_{ij}NME$_i$ and NME can be taken as functions of $\tilde{q}_{ij}$ for all i, j such that $j \in n(i)$ and $\tilde{q}_{ij} > 0$ as free variables (where f_i appearing in the sum NME $= \sum_i f_i$NME$_i$ is included as $f_i = \tilde{q}_{ii}$). We can do this instead of as functions of q_{ij}, which would then put constraints on $\tilde{q}_{ij}$.

Definition 5.7. Consider a flow network q_{ij} and a sub-set of the nodes $\mathcal{F}$. Let $\tilde{q}_{ij}$ be the flows between nodes which are not necessarily adjacent computed from q_{ij} as have defined previously in the paper. We define the **local negative mixing entropy restricted to $\mathcal{F}$ at node** i , denoted**NME**$_{\mathcal{F}i}$, to be

$$\text{NME}_{\mathcal{F}i} \quad = \quad \sum_{j:\tilde{q}_{ji}>0, j\in\mathcal{F}} \frac{\tilde{q}_{ji}}{\sum_{l:\tilde{q}_{li}>0, l\in\mathcal{F}} \tilde{q}_{jl}} \log\left(\frac{\tilde{q}_{ji}}{\sum_{l:\tilde{q}_{li}>0, l\in\mathcal{F}} \tilde{q}_{jl}}\right).$$

Also define the **total negative mixing entropy restricted to $\mathcal{F}$**, denoted $\mathrm{NME}_{\mathcal{F}}$, to be

$$\mathrm{NME}_{\mathcal{F}} \;=\; \sum_{i\in\mathcal{F}} f_i \mathrm{NME}_{\mathcal{F}i}.$$

This definition can be described as a total negative mixing entropy applied to the flow network q_{ij} while ignoring species coming from nodes $i \notin \mathcal{F}$ as well as the summands coming from nodes not in $\mathcal{F}$, $f_i\mathrm{NME}_i$. It is not the NME of the flow obtained after equilibrium is again reached after deleting nodes and their adjacent conductances. Still, it does have some physical meaning in that if the nodes outside of $\mathcal{F}$were rendered non-conducive or inoperable at a point in time then the NME for an infinitesimal division of time would be the NME restricted to the remaining set.

In the next theorem, we show how a small enough dominance factor can be chosen so that the resulting set of strong nodes has an NME that is close to the NME of the entire network.

Theorem 5.8. *Let G be an ambient network with N nodes and maximum total degree of a node D. Let Q_i be boundary flows on G such that $\sum_{i\in G} Q_i = 0$ and $\sum_{i\in sources} Q_i = 1$ (the total flow is 1). Consider a flow network q_{ij} on G compatible with Q_i. Let $\epsilon > 0$. Then there exists $1 > \delta > 0$ such that if $\mathcal{F} = \mathcal{F}_\delta$ is the set of strong nodes above dominance factor δ then $|NME_{\mathcal{F}_\delta}(q_{ij}) - NME(q_{ij})| < \epsilon$.*

Proof. We show this in three parts. First we show that there exists a $a > 0$ such that if $\alpha \le a$ and $\mathcal{F} = \mathcal{F}_\alpha$ then for all $i \in \mathcal{F}$, $|f_i\mathrm{NME}_i - f_i\mathrm{NME}_{\mathcal{F}i}| < \frac{\epsilon}{2N}$. Since the total flow of the network is 1, $f_i \le 1$ so it suffices to show $|\mathrm{NME}_i - \mathrm{NME}_{\mathcal{F}i}| < \frac{\epsilon}{2N}$.

To start this part of the proof, first take $\frac{1}{e} > a > 0$ to be arbitrary (e is Euler's constant), Let $\alpha \le a$ and define $\mathcal{F} = \mathcal{F}_\alpha$. Then

$$\mathrm{NME}_i \;=\; \sum_{j\in\mathcal{F}} \frac{\tilde{q}_{ji}}{\sum_l \tilde{q}_{li}} \log\left(\frac{\tilde{q}_{ji}}{\sum_l \tilde{q}_{li}}\right) + \sum_{j\notin\mathcal{F}} \frac{\tilde{q}_{ji}}{\sum_l \tilde{q}_{li}} \log\left(\frac{\tilde{q}_{ji}}{\sum_l \tilde{q}_{li}}\right).$$

Let $j \notin \mathcal{F}$ such that $\tilde{q}_{ji} > 0$. We have $\frac{\tilde{q}_{ji}}{\sum_l \tilde{q}_{li}} = \frac{f_j P_{ji}}{\sum_l \tilde{q}_{li}} \leq \frac{f_j P_{ji}}{\tilde{q}_{ii}} = \frac{f_j P_{ji}}{f_i} \leq \frac{f_j}{f_i} < \alpha < a$. Because $x\log(x)$ is decreasing on $[0, \frac{1}{e}]$, we know that

$$\left| \sum_{j \notin \mathcal{F}} \frac{\tilde{q}_{ji}}{\sum_l \tilde{q}_{li}} \log\left(\frac{\tilde{q}_{ji}}{\sum_l \tilde{q}_{li}} \right) \right| \quad < \quad |\mathcal{F}^C| \, |\alpha \log(\alpha)| \leq N \, |a \log(a)|$$

Let $j \in \mathcal{F}$. We write $\frac{\tilde{q}_{ji}}{\sum_l \tilde{q}_{li}} = \frac{\tilde{q}_{ji}}{\sum_{l \in \mathcal{F}} \tilde{q}_{li} + \sum_{l \notin \mathcal{F}} \tilde{q}_{li}}$. Then

$$\begin{aligned}
\frac{\tilde{q}_{ji}}{\sum_{l \in \mathcal{F}} \tilde{q}_{li}} - \frac{\tilde{q}_{ji}}{\sum_l \tilde{q}_{li}} &= \frac{\tilde{q}_{ji} \sum_{l \in \mathcal{F}} \tilde{q}_{li} + \tilde{q}_{ji} \sum_{l \notin \mathcal{F}} \tilde{q}_{li} - \tilde{q}_{ji} \sum_{l \in \mathcal{F}} \tilde{q}_{li}}{\left(\sum_{l \in \mathcal{F}} \tilde{q}_{li}\right) \left(\sum_l \tilde{q}_{li}\right)} \\
&= \frac{\tilde{q}_{ji} \sum_{l \notin \mathcal{F}} \tilde{q}_{li}}{\left(\sum_{l \in \mathcal{F}} \tilde{q}_{li}\right) \left(\sum_l \tilde{q}_{li}\right)} \\
&\leq \frac{\tilde{q}_{ji} \sum_{l \notin \mathcal{F}} \tilde{q}_{li}}{f_i^2} < \alpha^2 |\mathcal{F}^C| \leq a^2 N.
\end{aligned}$$

Since $x\log(x)$ is continuous on $[0,1]$, a compact set, it is uniformly continuous. Therefore we can find $a_1 > 0$ such that for all $x, y \in [0,1]$ if $|x - y| < 2a_1$ then $|x\log(x) - y\log(y)| < \frac{\epsilon}{4N^2}$. As a specific case, taking $y = 0$ implies for all $0 < x < 2a_1$ so $x \log x < \frac{\epsilon}{4N^2}$. Since a was arbitrary, we may choose $a = \min\left(a_1, \sqrt{\frac{a_1}{N}}\right)$ and have the above inequalities hold. Then for all $i \in \mathcal{F}$ and $j \notin \mathcal{F}$ such that $\tilde{q}_{ji} > 0$, $\frac{\tilde{q}_{ji}}{\sum_{l \in \mathcal{F}} \tilde{q}_{li}} - \frac{\tilde{q}_{ji}}{\sum_l \tilde{q}_{li}} < a^2 N < a_1 < 2a_1$. We have

$$\begin{aligned}
|\mathrm{NME}_{\mathcal{F}i} - \mathrm{NME}_i| &= \left| \sum_{,j \in \mathcal{F}} \frac{\tilde{q}_{ji}}{\sum_{l \in \mathcal{F}} \tilde{q}_{li}} \log\left(\frac{\tilde{q}_{ji}}{\sum_{l \in \mathcal{F}} \tilde{q}_{li}} \right) - \sum_j \frac{\tilde{q}_{ji}}{\sum_l \tilde{q}_{li}} \log\left(\frac{\tilde{q}_{ji}}{\sum_l \tilde{q}_{li}} \right) \right| \\
&\leq \left| \sum_{,j \in \mathcal{F}} \frac{\tilde{q}_{ji}}{\sum_{l \in \mathcal{F}} \tilde{q}_{li}} \log\left(\frac{\tilde{q}_{ji}}{\sum_{l \in \mathcal{F}} \tilde{q}_{li}} \right) - \sum_{j \in \mathcal{F}} \frac{\tilde{q}_{ji}}{\sum_l \tilde{q}_{li}} \log\left(\frac{\tilde{q}_{ji}}{\sum_l \tilde{q}_{li}} \right) \right| \\
&+ \left| \sum_{j \in \mathcal{F}} \frac{\tilde{q}_{ji}}{\sum_l \tilde{q}_{li}} \log\left(\frac{\tilde{q}_{ji}}{\sum_l \tilde{q}_{li}} \right) \right| \\
&< \sum_{,j \in \mathcal{F}} \left| \frac{\tilde{q}_{ji}}{\sum_{l \in \mathcal{F}} \tilde{q}_{li}} \log\left(\frac{\tilde{q}_{ji}}{\sum_{l \in \mathcal{F}} \tilde{q}_{li}} \right) - \frac{\tilde{q}_{ji}}{\sum_l \tilde{q}_{li}} \log\left(\frac{\tilde{q}_{ji}}{\sum_l \tilde{q}_{li}} \right) \right| + Na \log a \\
&< |\mathcal{F}| \frac{\epsilon}{4N^2} + N \frac{\epsilon}{4N^2} < \frac{\epsilon}{2N}.
\end{aligned}$$

for all $0 < \alpha < a$ and $\mathcal{F} = \mathcal{F}_\alpha$.

For the second part we show that there exists a $0 < b < 1$ such that for all $\beta \leq b$ if we take $\mathcal{F} = \mathcal{F}_\beta$ then

$$\left| \sum_{i \notin \mathcal{F}} f_i \mathrm{NME}_i \right| \quad < \quad \frac{\epsilon}{2}.$$

Note that since there are N nodes in the network, for a node i, the absolute value $|\mathrm{NME}_i| = \left| \sum_j \frac{\tilde{q}_{ji}}{\sum_l \tilde{q}_{li}} \log \left(\frac{\tilde{q}_{ji}}{\sum_l \tilde{q}_{li}} \right) \right|$ attains its maximum magnitude of $\log(N)$ when the flow from all of the nodes in the network to i are positive and have the same magnitude resulting in $\frac{\tilde{q}_{ji}}{\sum_l \tilde{q}_{li}} \equiv \frac{1}{N}$. Therefore

$$\left| \sum_{i \notin \mathcal{F}} f_i \mathrm{NME}_i \right| \quad < \quad \sum_{i \notin \mathcal{F}} |f_i \mathrm{NME}_i| < \sum_{i \notin \mathcal{F}} f_i \log N < \beta N \log N < bN \log N.$$

The third inequality in the relationship above is due to $f_i < \beta f_j$ for all $i \notin \mathcal{F}$, $j \in \mathcal{F}$ and that $f_j < \sum_{i \, \mathrm{source}} Q_i = 1$. Take $b = \frac{\epsilon}{2N \log N}$.

Now define $\delta = \min(a, b)$. Then $\delta \leq a$ and $\delta \leq b$. Therefore, the inequality from part a and part b hold. Let $\mathcal{F} = \mathcal{F}_\delta$. Then

$$\begin{aligned}
|\mathrm{NME}(q_{ij}) - \mathrm{NME}_{\mathcal{F}}(q_{ij})| &= \left| \sum_i f_i \mathrm{NME}_i - \sum_{i \in \mathcal{F}} f_i \mathrm{NME}_{\mathcal{F}i} \right| \\
&\leq \left| \sum_{i \in \mathcal{F}} f_i \mathrm{NME}_i - \sum_{i \in \mathcal{F}} f_i \mathrm{NME}_{\mathcal{F}i} \right| + \left| \sum_{i \notin \mathcal{F}} f_i \mathrm{NME}_i \right| \\
&< \sum_{i \in \mathcal{F}} \left| f_i \mathrm{NME}_i - \sum_{i \in \mathcal{F}} f_i \mathrm{NME}_{\mathcal{F}i} \right| + bN \log N \\
&\leq \sum_{i \in \mathcal{F}} |\mathrm{NME}_i - \mathrm{NME}_{\mathcal{F}i}| + bN \log N \\
&< |\mathcal{F}| \frac{\epsilon}{2N} + \frac{\epsilon}{2} \leq \epsilon,
\end{aligned}$$

which is what we wanted. □

5.9.3 Counting the number of strong nodes and high magnitude flows

Lemma 5.3. *Let $\mathcal{F} = \mathcal{F}_\delta$ be a subset of nodes of a physical flow network for some conductance network (to be certain we can define $\tilde{q}_{ij}$). Suppose that $\mathcal{F}$ contains n nodes. Then the largest possible value NME can take on is $-\log(n!)$.*

Proof. Let $n = |\mathcal{F}|$. Recall $\mathrm{NME}_{\mathcal{F}}(q_{ij}) = \sum_{i\in\mathcal{F}} f_i \mathrm{NME}_{\mathcal{F}i}$. We customize the definition of V_k from the proof of theorem 5.1 to assist study the optimal values of $\mathrm{NME}_{\mathcal{F}}(q_{ij})$. Let $1 \leq k \leq n$. Define V_k to be the set of $i \in \mathcal{F}$ such that $\#\{j \in \mathcal{F} : \tilde{q}_{ji} > 0\} = k$. Suppose there are $u, v \in V_k$ such that $\tilde{q}_{uv} > 0$. Then $P_{uv} > 0$. Let $i \in \mathcal{F}$ such that $\tilde{q}_{iu} > 0$. Then $P_{iv} \geq P_{iu}P_{uv} > 0$. Therefore $\tilde{q}_{iv} = f_i P_{iv} > 0$ and so $\{j \in \mathcal{F} : \tilde{q}_{ju} > 0\} \subset \{j \in \mathcal{F} : \tilde{q}_{jv} > 0\}$. By definition $\tilde{q}_{uv} > 0 \Rightarrow \tilde{q}_{vu} = 0$ $v \notin \{j \in \mathcal{F} : \tilde{q}_{ju} > 0\}$. at the same time $v \in \{j \in \mathcal{F} : \tilde{q}_{jv} > 0\}$, so this containership is strict. Therefore $|\{j \in \mathcal{F} : \tilde{q}_{jv} > 0\}| > |\{j \in \mathcal{F} : \tilde{q}_{ju} > 0\}|$, which contadicts the assumption that they both have equal number of elements k. Therefore $\tilde{q}_{uv} = 0$ and $P_{uv} = 0$ for all $u, v \in V_k$. The event of a random walker starting at the source and passing (taking on a value) in V_k before it reaches the sink is the union of the events that is passes any $u \in V_k$. These are disjoint events because the probability of a walker reaching $u \in V_k$ at time t_1 and then some $v \in V_k$ at a later time t_2 is zero because the conditional probability the walker will arrive at v at some time $s > t_2$ given that it was at u at time $t_1$is 0. Therefore the probability of a random walker starting at the source and passing through V_k is $\sum_{i\in V_k} P_{1i} = \sum_{i\in V_k} f_i$. From this we get

$$\mathrm{NME}_{\mathcal{F}}(q_{ij}) \quad = \quad \sum_{k=1}^{n} \sum_{i\in V_k} f_i \mathrm{NME}_{\mathcal{F}i}.$$

Since for $i \in V_k$, $\mathrm{NME}_{\mathcal{F}i}$ is the negative entropy of a probability distribution on a set with

k elements, its smallest possible value is $-\log(k)$. Since $\sum_{i \in V_k} f_k = 1$ we have that

$$\begin{aligned} \mathrm{NME}_{\mathcal{F}}(q_{ij}) &= \sum_{k=1}^{n} \sum_{i \in V_k} f_i \mathrm{NME}_{\mathcal{F}i} \\ &\geq \sum_{k=1}^{n} \sum_{i \in V_k} -f_i \log(k) \\ &\geq \sum_{k=1}^{n} 1_{V_k \neq \emptyset} \log(k) \geq -\log(n!). \end{aligned}$$

□

Lemma 5.4. *Let q_{ij} be a flow network with the hypotheses of the theorem such that $NME(q_{ij}) < -\log(m!) - \epsilon$. Let $\delta > 0$ be a dominance factor such that for $\mathcal{F} = \mathcal{F}_\delta$, $|NME_{\mathcal{F}}(q_{ij}) - NME(q_{ij})| < \frac{\epsilon}{2}$. Then there are at least $m+1$ strong nodes. That is, $|\mathcal{F}| \geq m+1$.*

Proof. The hypothesis$|\mathrm{NME}_{\mathcal{F}}(q_{ij}) - \mathrm{NME}(q_{ij})| < \frac{\epsilon}{2}$ is equivalent to $-\frac{\epsilon}{2} < \mathrm{NME}_{\mathcal{F}}(q_{ij}) - \mathrm{NME}(q_{ij}) < \frac{\epsilon}{2}$ and $\mathrm{NME}(q_{ij}) < -\log(m!) - \epsilon$ is equivalent to $\epsilon < -\log(m!) - \mathrm{NME}(q_{ij})$. We therefore have the lower bound

$$\begin{aligned} -\log(m!) - \mathrm{NME}_{\mathcal{F}}(q_{ij}) &= -\log(m!) - \mathrm{NME}(q_{ij}) - (\mathrm{NME}_{\mathcal{F}}(q_{ij}) - \mathrm{NME}(q_{ij})) \\ &> \epsilon - \frac{\epsilon}{2} = \frac{\epsilon}{2}. \end{aligned}$$

Thus $\mathrm{NME}_{\mathcal{F}}(q_{ij}) < -\log(m!) - \frac{\epsilon}{2}$.

Then by using Lemma 5.3

$$-\log(m!) - \frac{\epsilon}{2} \quad > \mathrm{NME}_{\mathcal{F}}(q_{ij}) \geq \quad -\log(n!).$$

Thus $\log(n!) > \log(m!) + \frac{\epsilon}{2}$ and since $\log(n!)$ is strictly increasing in n for $n \geq 1$, $n > m$. That is $|\mathcal{F}_\delta| = |\mathcal{F}| = n \geq m+1$, which is what we wanted. □

5.9.4 Proof that path networks persist as optimal for small material cost exponents γ (Theorem 5.7)

Proof. Assume that that q_{ij} is such that $-\log(m!) - \epsilon < \mathrm{NME}(q_{ij}) < -\log(m!) + \epsilon$. By Theoem 5.8 there is a δ such that for $\mathcal{F} = \mathcal{F}_\delta$, the strong nodes above the dominance factor δ, $|\mathrm{NME}_{\mathcal{F}}(q_{ij}) - \mathrm{NME}(q_{ij})| < \epsilon$. Then $\mathrm{NME}_{\mathcal{F}}(q_{ij}) \in (-\log(m!) - 2\epsilon, -\log(m!) + 2\epsilon)$. From this we know that $|\mathcal{F}| \geq m$. If it is greater than or equal to $m+1$ then we can find m flows uniformly bounded below depending only on m and ϵ and this will cause the dissiaption of this network to grow infinitely quickly comparted to $D(\tau_m)$ as $\gamma \to \infty$. Therefore we conclude that $\mathcal{F}$ has m nodes. Let V_k bet the set of nodes in $\mathcal{F}$with k predecessors in $\mathcal{F}$. As in the proof of uniqueness of the NME minimizer over networks with n vertices, if there is a $V_k = \emptyset$ then

$$\mathrm{NME}_{\mathcal{F}}(q_{ij}) \quad = \quad \sum_k \mathbf{1}_{V_k \neq \emptyset} \sum_{i \in V_k} f_i \mathrm{NME}_{\mathcal{F}i} \geq -\log(m!) + \log(k) > -\log(m!) + \epsilon.$$

(See proof of 5.1 for properties of V_k). Therefore $\mathcal{F}$ contains a path of length m. A network connecting source to sink on m nodes with a path of length m as a subnetwork has it so hat, each $V_k$contains only one node 5.1. For simplicity, denote by i be the single node in V_i for each $1 \leq i \leq m$. Then

$$\mathrm{NME}_{\mathcal{F}}(q_{ij}) \quad = \quad \sum_{i=1}^{m} f_i \mathrm{NME}_{\mathcal{F}_i}$$

and

$$\begin{aligned} 2\epsilon > \mathrm{NME}_{\mathcal{F}}(q_{ij}) - \left(-\sum_{i=1}^{m}\log(i)\right) &= \sum_{i=1}^{m}\left(\log(i) - f_i \mathrm{NME}_{\mathcal{F}_i}\right) \\ &\geq \sum_{i=1}^{m}\left(\log(i) - f_i\log(i)\right) \\ &= \sum_{i=1}^{m}(1-f_i)\log(i), \end{aligned}$$

which implies, that $\frac{2\epsilon}{\log(2)} > (1 - f_i)$ for each i. We can see by induction, that only at very small portion of the flow, $\frac{2\epsilon}{\log(2)}$ can be diverted from an edge flow along the longest path. $\square$

5.9.5 Lower Bound on Flows

In order to prove cases A and B, we develop a lower bound on the flows, showing that for a number of flows larger than the number in the next smallest path, there is a lower boun.

We carry though the definition of δ and $\mathcal{F}_\delta$ from the conclusion of lemma 5.4 with a small adjustment. We also assume that $\delta < \frac{1}{D}$ by perhaps redefining $\delta := \min(\delta, \frac{1}{100D})$. Indeed, any number $0 < \hat{\delta} < \delta$ is also dominance factor such that

$$\left|\mathrm{NME}_{\mathcal{F}_{\delta'}}(q_{ij}) - \mathrm{NME}(q_{ij})\right| \leq \left|\mathrm{NME}_{\mathcal{F}_\delta}(q_{ij}) - \mathrm{NME}(q_{ij})\right| < \frac{\epsilon}{2}.$$

Note that the definition of δ only depends on the ambient grid, m and ϵ (actually $\frac{\epsilon}{2}$). We now demonstrate a uniform lower bound on at least $m+1$ flows from q_{ij} dependent only on m and ϵ. Let i be a numbering of the nodes of $\mathcal{F}_\delta$ such that $f_i > f_{i+1}$for all $1 \leq i < |\mathcal{F}_\delta|$. Since we are studying the 1 source 1 sink case with both having magnitude 1 the source node has a total flow of 1 . Therefore $f_1 = 1$. By the definition of the set of strong nodes above the dominance factor (see definition 5.6 in section 5.9.2), ordering the strong nodes in order of decreasing total flow it must be true that $f_{i+1} \geq \delta f_i$ for all $1 \leq i < |\mathcal{F}_\delta|$. Since $f_1 = 1$, an application of induction implies $f_i \geq \delta^{i-1}$ for all $1 \leq i \leq |\mathcal{F}_\delta|$. In our proof of the theorem, it

will be more wieldy to use the weaker lower bound of $f_i \geq \delta^{|\mathcal{F}_\delta|-1}$. We summarize the result of the deduction in this paragraph in the following lemma.

Lemma 5.5. *If i is numbering of the set of strong nodes above the dominance factor δ, $\mathcal{F}_\delta$, in decreasing order of f_i, and there is one source and one sink each with a flow of 1 then $f_1 = 1, f_2 \geq \delta, \ldots, f_i \geq \delta^{i-1}, \ldots, f_{|\mathcal{F}_\delta|} \geq \delta^{|\mathcal{F}_\delta|-1}$. The weaker lower bound $f_i \geq \delta^{|\mathcal{F}_\delta|-1}$ is then guarenteed*

Since we are addressing case B of theorem 5.6 (see section 5.9.1), we have the assumption on q_{ij}: $-\log((m+1)!) + \epsilon < \mathrm{NME}(q_{ij}) < -\log(m!) - \epsilon$.

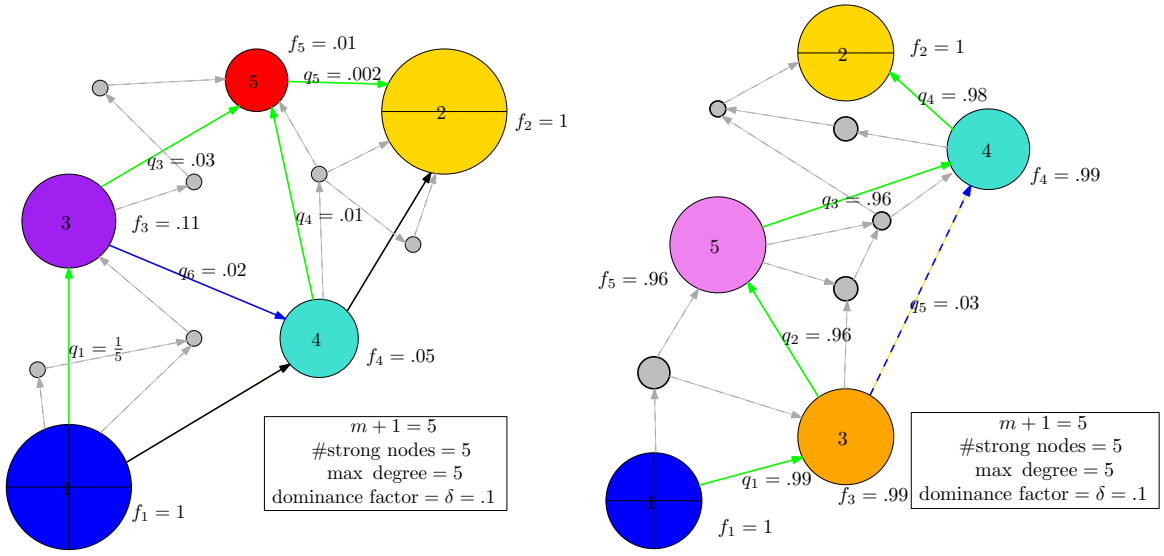
We now look at 3 cases to find a set of $m+1$ flows bounded below.

Lemma 5.6. *If $|\mathcal{F}_\delta| > m+1$ then there exists $m+1$ flows in the flow network q_{ij} such that $q_{ij} \geq \frac{\delta^{m+1}}{D}$ where D is the max total degree of the undirected ambient network G.*

Proof. First number the nodes in $|\mathcal{F}_\delta|$ in decreasing order of f_i. Then by lemma 5.5 $f_i \geq \delta^{i-1}$. Since $|\mathcal{F}_\delta| > m+1$ there are at least $m+2$ nodes. They obey the sequential lower bound $f_1 \geq 1, f_2 \geq \delta, \ldots, f_i \geq \delta^{i-1}, \ldots, f_{m+2} \geq \delta^{m+1}$ and as in the proof for lemma 5.5, since $0 < \delta < 1$ we have that $f_i \geq \delta^{m+1}$ for all $1 \leq i \leq m+2$. This puts a lower bound on the total-flows of $m+2$ nodes. Now define q_i to be the largest outgoing flow of node i for $1 \leq i \leq m+2$ i not the sink. Then since the total out-degree is $< D$, an upper bound on the total degree, and $q_i \geq \frac{f_i}{\text{out-degree of } i}$, being the largest outgoing flow, we get $q_i \geq \frac{f_i}{D} \geq \frac{\delta^{m+1}}{D}$. Since the sink has strength 1 it could very well be one of the first $m+2$ nodes in this sequence. But it is the only (strong node–for some weak nodes might have $f_i = 0$) node with no out-flowing edges in the one source one sink set up therefore there are at least $m+1$ flows q_i in the list of the first $m+2$ flows bounded below by $\frac{\delta^{m+1}}{D}$. □

Now that we have established the lower bound for $m+2$ strong nodes, the remaining case is to show there is a uniform lower bound on $m+1$ flows when there are $m+1$ strong flows. This case then can be divided into two other cases. First we use lemma 5.5 to obtain the

lower bound fact that $f_i \geq \delta^m$. For each of the nodes other than the sink, take the largest out-going flow and call this set $\mathcal{S}$. We know if i is not the sink and $i \in \mathcal{F}_\delta$ then, if q_{ij} is the strongest out going flow of i, $q_{ij} \geq \frac{f_i}{D} \geq \frac{\delta^m}{D}$. By the way we chose δ we have $\delta < \frac{1}{D}$. Therefore, if q_{ij} is the strongest outgoing flow of i then $f_j \geq q_{ij} \geq \frac{f_i}{D} > \delta f_i$, implying that j is also a strong node. There are two possible scenarios. Either the destination nodes of all of the m flows in $\mathcal{S}$ are distinct, or equivalently, every node except for the source is pointed to by an edge is $\mathcal{S}$. In this case, since there are $m+1$ nodes we can follow from each node down its strongest flow to the next node to get a path of length $m+1$. The other case is that the nodes together with the flows in $\mathcal{S}$ do not form a path. Then there must be one node which is the endpoint of two strong flows and one node which is not being pointed to by any flows in $\mathcal{S}$. We establish this second case in the next lemma. The first case, although initially less complex, actually has a more complicated proof.



(a) (b)

Figure 5.9.3: Figure 5.9.3a: As in the previous example, $m + 1 = 5$. This illustrates the **second** case in our proof. There are 5 strong nodes, the same number as there are in the next largest neighboring path (in terms of NME). The green edges represent the q_i strong flows which are chosen as outgoing flows. Note that there are 4 of these. Since the path with 5 nodes also have 4 flows, this is not sufficient to show that the dissipation grows infinitely large compared to this next largest path. Figure 5.9.3b: Continuing with the same $m+1 = 5$, the third case in our proof occurs if the the set of chosen strong flows (green edges) contains a path connecting every node. In this case, it seems we are cornered becuase the lower bound on the magnitude of the NME, $\text{NME} < -\log(m!) - \epsilon$ cannot be used to find more than m chosen flows. Indeed, the path τ_{m+1} fits the lower bound. To scale this obstacle, we use the upper bound (on $|\text{NME}| < \log((m + 1)!) - \epsilon$) from the tansition to the next larger set of nodes as a belay. To be bounded away from the NME of the path on $m+1$ nodes there must be a flow (represented by the dashed blue arrow) diverting flow away from the path with a strength above some amount depending on ϵ and $m + 1$.

Lemma 5.7. *Assume that* $|\mathcal{F}_\delta| = m + 1$. *Let* $\mathcal{S} = \{\max_{j \in n(i)} q_{ij} : i \in \mathcal{F}_\delta\}$. *Suppose that the collection of endpoints of the flows in* $\mathcal{S}$ *does not contain every* $i \in \mathcal{F}_\delta$ *excluding the source node. Then there are* $m + 1$ *flows with magnitudes greater than or equal to* $\frac{\delta^m}{D}$.

Proof. We know that $\mathcal{S}$has a flow for every node other than the sink. Therefore, by lemma 5.5 and since each $q_{ij} \in \mathcal{S}$ is such that $q_{ij} \geq \frac{f_i}{D}$, we have $q_{ij} \geq \frac{\delta^m}{D}$. This results in m flows with the uniform lower bound. To get the $m + 1$ flow, let $u \in \mathcal{F}_\delta$, u not the source, be

such that there does not exist a flow in $\mathcal{S}$ with u as its end point. We can then choose the maximal in-flow to u, say q_{vu} for node v assured that it was not already counted in $\mathcal{S}$. Then following the same reasoning that q_{vu} is larger than the average of the in-flows, or f_i divided by the number of in-flows, we get $q_{vu} \geq \frac{f_i}{D} \geq \frac{\delta^m}{D}$. This gives us $m+1$ flows with magnitudes greater than or equal to $\frac{\delta^m}{D}$. □

Now we tackle the case where $\mathcal{S}$ together with all $m+1$ nodes in $\mathcal{F}_\delta$ forms a path with $m+1$ points. Up until now, we have relied on $\text{NME}(q_{ij}) < -\log(m!) - \epsilon$. For this proof we need the other boundary, $-\log((m+1)!) + \epsilon < \text{NME}(q_{ij})$.

Lemma 5.8. *Assume that $|\mathcal{F}_\delta| = m+1$. Let $\mathcal{S} = \{\max_{j \in n(i)} q_{ij} : i \in \mathcal{F}_\delta\}$. Assume that the nodes in $\mathcal{F}_\delta$ and the flows in $\mathcal{S}$ form a path of length $m+1$, including all of the nodes in $\mathcal{F}_\delta$. Then there exists a small postive number $0 < \eta < 1$ depending only on m and ϵ such that there are $m+1$ flows $q_{ij} \geq \frac{\eta\delta^m}{D}$.*

Proof. As we have pointed out in figure 5.9.3. If the set of strong nodes has $m+1$ nodes and the set of chosen strong flows $\mathcal{S}$ forms a path, then we cannot use the assumption $\text{NME}(q_{ij}) < -\log(m!) - \epsilon$ to find another flow with magnitude $\geq \frac{f_i}{D}$ for some i. Indeed, the path with flows of magnitude 1 τ_{m+1} is such that $\text{NME}(\tau_{m+1}) = -\log((m+1)!) < -\log(m!) - \epsilon$ and it only has m non-zero flows. Therefore, for the first time in the proof we use the other assumption on q_{ij}: the lower bound on NME, $\text{NME}(q_{ij}) > -\log((m+1)!) + \epsilon$.

Let $1 \leq i \leq m+1$ be an enumeration of the set $\mathcal{S}$. Then $q_{i,i+1} > \frac{f_i}{D}$ for all $1 \leq i \leq m$. Let $0 < s \leq 1$ such that $\min_{1 \leq i \leq m} \frac{q_{i,i+1}}{f_i} = 1 - s$. Recall $\tilde{q}_{uv} = P_{uv} f_u$ where P_{uv} is the probability of hitting v given the first state of a random walk is u on the Markov chain as computed in Chapter 2. on the network with all of the nodes, for all $u, v \in G$. Therefore for strong nodes i, j $1 \leq i < j \leq m+1$

$$\tilde{q}_{ij} \quad = \quad f_i P_{ij} \geq f_i P\left(x_l = i + l \text{ for } 1 \leq l \leq j - i | x_0 = i\right) = f_i \prod_{l=i}^{j-1} \frac{q_{l,l+1}}{f_l}$$

where u_l, $l \geq 0$ represents random sequences distributed according to the Markov chain given by the flows. Basically, we are calling on the fact that the probability of taking every possible path from i to j given we start at i is bounded below by the probability of a single path, the one defined by following the flows in set $\mathcal{S}$ and hitting the nodes in $\mathcal{F}_\delta$ along the path of length $m+1$. Therefore

$$f_i \geq \tilde{q}_{ij} \quad \geq \quad f_i \prod_{l=i}^{j-1} \frac{q_{l,l+1}}{f_l} \geq f_i(1-s)^{j-i}.$$

We also have by the facts stated in the model section, $\tilde{q}_{ii} = f_i = \tilde{q}_{1i}$ because 1 is the only source and the flow $Q_1 = 1$. Therefore, applying this and the last inequality, $1 \geq f_i \geq f_1(1-s)^{i-1} = (1-s)^{i-1}$.

Since δ is defined so that $|\mathrm{NME}_{\mathcal{F}_\delta}(q_{ij}) - \mathrm{NME}(q_{ij})| < \frac{\epsilon}{2}$, and by the assumption that $\mathrm{NME}(q_{ij}) > -\log((m+1)!) + \epsilon$ we have

$$\begin{aligned} \mathrm{NME}_{\mathcal{F}_\delta}(q_{ij}) - (-\log((m+1)!)) &= \mathrm{NME}_{\mathcal{F}_\delta}(q_{ij}) - \mathrm{NME}(q_{ij}) + \mathrm{NME}(q_{ij}) - (-\log((m+1)!)) \\ &> -\frac{\epsilon}{2} + \epsilon = \frac{\epsilon}{2}. \end{aligned}$$

We now place a lower bound on s. Recall, that by the definition of $\mathrm{NME}_{\mathcal{F}_\delta i}$, for all $i \in \mathcal{F}_\delta$

$$\mathrm{NME}_{\mathcal{F}_\delta i} \quad = \quad \sum_{j \leq i} \frac{\tilde{q}_{ji}}{\sum_{k \leq i} \tilde{q}_{ki}} \log \left(\frac{\tilde{q}_{ji}}{\sum_{k \leq i} \tilde{q}_{ki}} \right)$$

and $\sum_{1 \leq i \leq m+1} f_i \mathrm{NME}_{\mathcal{F}_\delta i}$ is a continuous function of $\tilde{q}_{ij}$ as positive free variables. Therefore there exists t depending only on $\frac{\epsilon}{2}$ and $m+1$ such that if each $(1-t) < f_i \leq 1$ and each $(1-t) < \tilde{q}_{ij} \leq 1$ for all $1 \leq i \leq j \leq m+1$ ($i, j \in \mathcal{F}_\delta$ such that $\tilde{q}_{ij} \geq 0$) then

$$-\log((m+1)!) + \frac{\epsilon}{2} \quad \geq \mathrm{NME}_{\mathcal{F}_\delta} \geq \quad -\log((m+1)!).$$

Therefore since $(1-s)^{i-1} \leq f_i \leq 1$, and $(1-s)^{j-1} \leq f_i(1-s)^{j-i} \leq \tilde{q}_{ij} \leq 1$ for all $1 \leq i \leq j \leq$

$m+1$, in order for $\text{NME}_{\mathcal{F}_\delta} > -\log((m+1)!) + \frac{\epsilon}{2}$ it must hold that $\min_{1\leq i\leq m+1}(1-s)^{i-1} = (1-s)^m < (1-t)$. Note $s \mapsto (1-s)^m$ is strictly decreasing on $[0,1]$. Therefore there exists $0 < \eta < 1$ which depends on t and m such that in order for $(1-s)^m < (1-t)$ it is necessary for $s > \eta$. Since η depended only on m and t, and choice of t depends only on m and ϵ, choice of η depends only on m and ϵ. Since $1 - s = \min_{1\leq i\leq m} \frac{q_{i,i+1}}{f_i}$ there exists a node i such that $1 - \frac{q_{i,i+1}}{f_i} = s$. Furthermore, assuming that $s > \eta$, at this node it holds that $1 - \frac{q_{i,i+1}}{f_i} > \eta$ and so $f_i - q_{i,i+1} > \eta f_i$. Therefore there are other out-going flows from i. Since the total remaining out-going nodes of i is less than D we can lower bound the strongest remaining outgoing flow by $\frac{\eta f_i}{D}$. And so by lemma 5.5 we that this greater than or equal to $\frac{\eta\delta^m}{D}$. For convenience and since $0 < \eta < 1$ we take $\frac{\eta\delta^m}{D}$ as the lower bound on $m+1$ flows $q_{ij} \geq \frac{\eta\delta^m}{D}$. □

For 3 different situations, depending on the flow network topology, we have demonstrated the existence of three lower bounds on the flows along $m+1$ edges. They are: $q_{ij} > \frac{\delta^{m+1}}{D}, q_{ij} > \frac{\delta^m}{D}$ and $q_{ij} > \frac{\eta\delta^m}{D}$. All three of these choices only depend on m and ϵ. We unify these lower bounds into one lower bound. To conveniently mark our progress for further use , we state this lower bound on the flows in the following theorem and assign it to formal terminology.

Theorem 5.9. *If q_{ij} is such that $-\log((m+1)) + \epsilon < NME(q_{ij}) \leq -\log(m!) - \epsilon$ then there exists a positive number $\xi > 0$ depending only on m and ϵ such that there exist $m+1$ flows in the network q_{ij} with $q_{ij} > \xi$. We refer to ξ as the* ***uniform flow lower-bound.***

Proof. Let q_{ij} satisfy the hypothesis of the theorem that $-\log((m+1)) + \epsilon < \text{NME}(q_{ij}) \leq -\log(m!) - \epsilon$. Then there exists $\delta > 0$ and $\eta > 0$ depending only on ϵ and m such that, by the exhaustive trichotomy of the set of strong nodes proven in this section and lemmas 5.65.7 and 5.8 for each case, one of the three following statements must hold:

There exists a cardinality $m+1$ set of flows $\mathcal{S}$ such that

$$q_{ij} \quad \geq \quad \frac{\delta^{m+1}}{D}$$

for all $q_{ij} \in \mathcal{S}$.

There exists a cardinality $m+1$ set of flows $\mathcal{S}$ such that

$$q_{ij} \geq \frac{\delta^m}{D}$$

for all $q_{ij} \in \mathcal{S}$.

There exists a cardinality $m+1$ set of flows $\mathcal{S}$ such that

$$q_{ij} \geq \frac{\eta\delta^m}{D}$$

for all $q_{ij} \in \mathcal{S}$.

That one of these three statements holds, implies that there exists a cardinality $m+1$ set of flows $\mathcal{S}$ such that

$$q_{ij} \geq \min\left(\frac{\delta^{m+1}}{D}, \frac{\delta^m}{D}, \frac{\eta\delta^m}{D}\right).$$

Which is what we wanted because now we can take $\xi = \min\left(\frac{\delta^{m+1}}{D}, \frac{\delta^m}{D}, \frac{\eta\delta^m}{D}\right)$ to be our **uniform lower bound on flows**. □

5.9.6 Finishing the proof.

In this section we will complete the proofs of case A and case B. First we prove case B and then apply the same concept to the proof of case A.

Before we begin we decide on parameters that will shift our focus on paths of length $m+1$ to be held at constant dissipation of $D(\tau_{m+1}) = m+1$ and paths of length m will have dissipation that shrinks to zero. As we will see networks with a uniform flow lower-bound on a set of $m+1$ flows will have come to have arbitrarily large dissipation for small γ.

For the rest of the proof of case B assume that all conductance networks obey the building

cost constraint with varying γ and fixed total material $C = m$. That is $\sum_{(i,j)} \kappa_{ij}^{\gamma} = C = m$. This is to be assumed even when the conductance network is not explicitly mentioned, such as in talking about the dissipation of a set of flows. As a reminder, if we are talking about the dissipation of a flow network we mean

$$\begin{aligned} D(q_{ij}) &= \min_{\kappa_{ij}:\sum_{(i,j)} \kappa_{ij}^{\gamma}, q_{ij} \text{ physical flows of } \kappa_{ij}} \sum_{(i,j)} \frac{q_{ij}^2}{\kappa_{ij}} \\ \text{or equivalently } &= \min_{\kappa_{ij}:\sum_{(i,j)} \kappa_{ij}^{\gamma}, q_{ij} \text{ physical flows of } \kappa_{ij}} D(\kappa_{ij}) \end{aligned}$$

Lemma 5.9. ***(Case B)*** *Consider flow networks on the ambient network G compatible with with one source, one sink and a total flow of 1. There exists $0 < \gamma_\epsilon < 1$ such that if $\gamma < \gamma_\epsilon$ then for all q_{ij} if $-\log((m+1)!) + \epsilon < NME(q_i) < -\log(m!) - \epsilon$ then*

$$NME(q_{ij}) + cD(q_{ij}) \quad > \quad NME(\tau_m) + cD(\tau_m)$$

for all $c_{m+1,m} \leq c \leq c_{m,m-1}$.

Proof. First, recall that $D(\tau_m) = (m-1)\left(\frac{m-1}{m}\right)^{\frac{1}{\gamma}}$ and $D(\tau_{m+1}) = m$. We now establish a lower bound on the dissipation of q_{ij}.

By theorem 5.9 there exists uniform flow lower bound ξ depending only on m and ϵ such that there is a set of $m+1$ flows $\mathcal{S}$ with $q_{ij} > \xi$ for all $q_{ij} \in \mathcal{S}$. By theorem 5.5 we have

$$\begin{aligned} D(q_{ij}) &\geq \left(\sum_{(i,j)} \left(q_{ij}^2\right)^{\frac{\gamma}{\gamma+1}}\right) \frac{\left(\sum_{(i,j)} \left(q_{ij}^2\right)^{\frac{\gamma}{\gamma+1}}\right)^{\frac{1}{\gamma}}}{m^{\frac{1}{\gamma}}} \\ &\geq (m+1)\xi^{\frac{\gamma}{\gamma+1}} \frac{(m+1)^{\frac{1}{\gamma}} \xi^{\frac{1}{\gamma}}}{m^{\frac{1}{\gamma}}} \\ &= (m+1)\xi \left(\frac{m+1}{m}\right)^{\frac{1}{\gamma}}. \end{aligned}$$

It follows that $D(q_{ij}) \to \infty$ as $\gamma \to 0^+$. Recall that the intersection of lines from τ_{m+1} and

τ_m

$$\begin{aligned} c_{m+1,m} &= \frac{\mathrm{NME}(\tau_m) - \mathrm{NME}(\tau_{m+1})}{D(\tau_{m+!}) - D(\tau_m)} \\ &= \frac{\log(m+1)}{m - (m-1)\left(\frac{m-1}{m}\right)^{\frac{1}{\gamma}}} \\ &\overset{\gamma\to 0^+}{\longrightarrow} \frac{\log(m+1)}{m} \end{aligned}$$

Therefore $c_{m+1,m} \to \frac{\log(m+1)}{m}$ as $\gamma \to 0^+$. Let c_L be the intersection of $c \mapsto \mathrm{NME}(q_{ij}) + cD(q_{ij})$ and $c \mapsto \mathrm{NME}(\tau_m) + cD(\tau_m)$. Then

$$c_L = \frac{\mathrm{NME}(\tau_m) - \mathrm{NME}(q_{ij})}{D(q_{ij}) - D(\tau_m)}$$

The intersection of $\mathrm{NME}(q_{ij}) + cD(q_{ij})$ and $\mathrm{NME}(\tau_m) + cD(\tau_m)$ is bounded above by

$$\begin{aligned} \frac{\mathrm{NME}(\tau_m) - \mathrm{NME}(q_{ij})}{D(q_{ij}) - D(\tau_m)} &\leq \frac{-\log(m!) + \log((m+1)!)}{(m+1)\xi\left(\frac{m+1}{m}\right)^{\frac{1}{\gamma}} - (m-1)\left(\frac{m-1}{m}\right)^{\frac{1}{\gamma}}} \\ &\overset{\gamma\to 0^+}{\longrightarrow} 0. \end{aligned}$$

Therefore $c_L \to 0$ and the limit from before, $c_{m+1,m} \to \frac{\log(m+1)}{m} > 0$. Therefore there exists a γ_ϵ such that for $\gamma < \gamma_\epsilon$ $c_L < c_{m+1,m}$. By lemma 5.2 q_{ij} is not an optimum of the mixing-dissipation cost for all $c_{m+1,m} < c < c_{m,m-1}$ when $\gamma < \gamma_\epsilon$. □

To prove case A, we focus on the path of length m . It suffices to prove case A for conductance networks maintaining a building cost of $C = m - 1$.

Lemma 5.10. ***(Case A)*** *Consider flow networks on the ambient network G compatible with with one source, one sink and a total flow of 1. There exists $0 < \gamma_\epsilon < 1$ such that if $\gamma < \gamma_\epsilon$ then for all q_{ij} if $-\log(m!) + \epsilon < NME(q_i) < -\log((m-1)!) - \epsilon$ then*

$$NME(q_{ij}) + cD(q_{ij}) > NME(\tau_m) + cD(\tau_m)$$

for all $c_{m+1,m} \leq c \leq c_{m,m-1}$.

Proof. In theorem 5.9, the assumption $-\log((m+1)!) + \epsilon < \text{NME}(q_i) < -\log(m!) - \epsilon$ is for arbitrary m so long as there are paths of length m and $m+1$ in the ambient network. It can therefore be applied to $m-1$. Assume that q_{ij} is a flow network with $-\log(m!) + \epsilon < \text{NME}(q_i) < -\log((m-1)!) - \epsilon$. By theorem 5.9 there exists a uniform lower bound for the flows $\xi > 0$ and a set of $m-1+1 = m$ flows $\mathcal{S}$ such that for all $q_{ij} \in \mathcal{S}$ $q_{ij} \geq \xi$. Then

$$\begin{aligned} D(q_{ij}) &\geq \left(\sum_{(i,j)} \left(q_{ij}^2\right)^{\frac{\gamma}{\gamma+1}}\right) \frac{\left(\sum_{(i,j)} \left(q_{ij}^2\right)^{\frac{\gamma}{\gamma+1}}\right)^{\frac{1}{\gamma}}}{(m-1)^{\frac{1}{\gamma}}} \\ &\geq m\left(\xi^2\right)\left(\frac{m}{m-1}\right)^{\frac{1}{\gamma}}. \\ &\overset{\gamma\to 0^+}{\longrightarrow} \infty. \end{aligned}$$

As before, we also have $D(\tau_m) = m-1$ and $D(\tau_{m-1}) = (m-2)\left(\frac{m-2}{m-1}\right)^{\frac{1}{\gamma}}$. Then define c_R to be such that

$$\text{NME}(q_{ij}) + c_R D(q_{ij}) = \text{NME}(\tau_m) + c_R D(\tau_m).$$

Then

$$\begin{aligned} c_R &= \frac{\text{NME}(q_{ij}) - \text{NME}(\tau_m)}{D(\tau_m) - D(q_{ij})} \\ &\leq \frac{\log(m)}{D(\tau_m) - D(q_{ij})} \end{aligned}$$

and

$$\begin{aligned} c_{m,m-1} &= \frac{\mathrm{NME}(\tau_{m-1}) - \mathrm{NME}(\tau_m)}{D(\tau_m) - D(\tau_{m-1})} \\ &= \frac{\log(m)}{(m-1) - (m-2)\left(\frac{m-2}{m-1}\right)^{\frac{1}{\gamma}}} \\ &\overset{\gamma \to 0^+}{\longrightarrow} \frac{\log(m)}{m-1}. \end{aligned}$$

Since $\mathrm{NME}(q_{ij}) > \mathrm{NME}(\tau_m)$ we know that if $D(q_{ij}) \geq D(\tau_m)$ then $\mathrm{NME}(q_{ij}) + cD(q_{ij}) > \mathrm{NME}(\tau_m) + cD(\tau_m)$ for all c. Then since $D(q_{ij}) > m\left(\xi^2\right)\left(\frac{m}{m-1}\right)^{\frac{1}{\gamma}}$ and the lower bound is increasing to ∞ as$\gamma \to 0$, we take γ_ϵ such that $m\left(\xi^2\right)\left(\frac{m}{m-1}\right)^{\frac{1}{\gamma}} > D(\tau_m) = m-1$. This completes the proof. □

Case C. Now we consider case C. Let q_{ij} be such that $\mathrm{NME}(q_{ij}) < -\log(m+1) - \epsilon$, by lemma 5.4 there exists a δ that depends only on $m+1$ and ϵ such that there are at least $m+2$ with $f_i \geq \delta^{m+1}$. Then there are at least $m+1$ flows bounded below by $\frac{\delta^{m+1}}{D}$. Taking the building cost to be $C = m$, the line from τ_{m+1} has constant slope, while as we found in lemma 5.9 $c_{m+1,m} \geq \frac{\log(m+1)}{m}$ for all γ while the intersection of the line from q_{ij} and the line from τ_m approaches 0 because the line from τ_m has slope approaching 0 while the line from q_{ij} gets infinitely steep at a rate depending only on m and ϵ. Therefore we can find a γ_ϵ such that if $\gamma < \gamma_\epsilon$then the intersection of the line from q_{ij} with the line from τ_m is less than $c_{m,m-1.}$

Case D. Now consider q_{ij} with $\mathrm{NME}(q_{ij}) > -\log(m!) + \epsilon$. Set $C = m-1$. We have that $\mathrm{NME}(q_{ij}) + cD(q_{ij}) > \mathrm{NME}(q_{ij})$ for all $c \geq 0$. Since the building cost is $m-1$, the dissipation of τ_m is $m-1$ so the line from $\tau_m$is $c \mapsto -\log(m!) + c(m-1)$. This intersects the horizontal line $c \mapsto \mathrm{NME}(q_{ij})$ at a postitive c, denote it c_R. Since $D(\tau_{m-1}) \to 0$ as $\gamma \to 0^+$ there is a $\gamma_\epsilon$such that $\mathrm{NME}(\tau_{m-1}) + c_R D(\tau_{m-1}) < \mathrm{NME}(q_{ij})$.

5.9.6.1 Concluding proof of theorem 5.6

In the last section, we have defined a potentially different $\gamma_\epsilon$for each of case A, B, C, D. To conclude the proof, simply re-define γ_ϵ to be the minimum of all of the $\gamma_\epsilon$defined in each case above.

5.9.7 Background

In this section, we consider flows on conductance networks with boundary flows consisting of one source with boundary flow 1 and one sink with boundary flow -1.

From our experimental results on square and triangular grids with one source and one sink we notice that networks optimizing NME $+ cD$ are either path networks or they have a morphology that consists of a loopy part at the source node, that then connects via a path to the sink node. In Chapter 2 we proved (Theorem 2.1) an equivalence principle for the negative mixing entropy and the negative sending entropy; namely the negative mixing entropy for a given network is identical to the negative sending entropy on the same network if its source and sink were exchanged with each other. Hence, we are able to prove results about the negative sending entropy, and directly translate them to the negative mixing entropy.

We have two equivalent statements about the optimal location of the loops in networks that minimize either NME or the NSE. Consider a flow network formed by concatenating a loop section with two paths. Specifically, we consider a network that can be partitioned into three subgraphs: a path that links directly to the source, a path that links directly to the sink, and a middle section that bridges the two paths, and that may contain loops (we call this third subnetwork, the loopy subnetwork). We will then theoretically compare between versions of this network in which the total length of the two paths is kept fixed, but the partitioning of nodes between them is varied. Among this class of networks, we minimize NME$(q_{ij}) + cD(q_{ij})$, if the path connecting to the source has zero nodes. On the other hand,

for the network to be an optimum of $\mathrm{NME}(q_{ij})+cD(q_{ij})$ the path connecting the sub-network with loops to the sink must have zero nodes. Note that changing the partitioning of nodes between the two path sub-networks doesn't change the set of flows in the network, so doesn't affect the dissipation, hence, we only need to show that NME attains its minimum out of these possible flow networks if the subnetwork with loops appears at the source and the NSE attains its minimum out of these possible values on flow networks if it is at the sink. Formally we express these statements in the two following equivalent theorems, Theorem 5.10 and Theorem 5.11.

5.9.8 Theorem statement

To state the theorem we need the following definition.

Definition 5.8. Suppose we have T networks, $J_1, J_2, \ldots, J_T$, on disjoint sets of nodes, where each network consists of either a singleton node, or some set of nodes and edges with only a single sink and a single source with outflow/inflow f . Then we define the connected union of $J_1, J_2, \ldots, J_T$ in that order, denoted $J_1 \to J_2 \to \cdots \to J_T$, to be the network on nodes equal to the union of the sets of nodes of J_i $1 \leq i \leq T$ with edges containing the union of the sets of edges of J_i $1 \leq i \leq T$ and extra edges that carry a flow of size f connecting the sink of J_i (or the only node in J_i, if J_i is a single node) to the source of J_{i+1} or to the only node in J_{i+1}, if J_{i+1} is a single node network for each $1 \leq i \leq T-1$. If $J_i = \emptyset$ for some $1 < i < T$ then we connect the sink of the last $J_j \neq \emptyset$ with $j < i$ to the source of the next k with $J_k \neq \emptyset$, by an edge carrying a flow f. If there is no such J_j then the source of $J_1 \to \cdots \to J_T$ is the source of J_k. Likewise if there is no such k then the sink of $J_1 \to \cdots \to J_T$ is the sink of J_j. The connected union is compatible with boundary flows containing a single source at the source of J_1 and a single sink at the sink of J_T and zero boundary flow at all other nodes.

Theorem 5.10. *Let $\mathcal{R}$ be a flow network with single source and single sink of total flow 1 and let $n > 0$ and let $t \geq 0$ such that $0 \leq t \leq n$. Define G_t to be the connected union*

$\tau_t \to \mathcal{R} \to \tau_{n-t}$. *Then* $NME(G_0) \leq NME(G_t)$ *for all* $1 \leq t \leq n$ *with equality if and only if* $\mathcal{R}$ *is a path. That is, if* $\mathcal{R}$ *contains loops then the unique network minimizer is* $G_0 = \tau_0 \to \mathcal{R} \to \tau_n = \mathcal{R} \to \tau_n$.

Theorem 5.11. *Let* $\mathcal{R}$ *be a flow network with single source and single sink of total flow* 1 *and let* $n > 0$ *and let* $t \geq 0$ *such that* $0 \leq t \leq n$. *Define* G_t *to be the connected union* $\tau_t \to \mathcal{R} \to \tau_{n-t}$. *Then* $NSE(G_n) \leq NSE(G_t)$ *for all* $0 \leq t \leq n-1$ *with equality if and only if* $\mathcal{R}$ *is a path. That is, if* $\mathcal{R}$ *contains loops then the unique network minimizer is* $G_n = \tau_n \to \mathcal{R} \to \tau_0 = \tau_n \to \mathcal{R}$.

Since these theorems are equivalent, we only need to prove one to have the other. We will prove 5.11, that loops must occur near to and at the sink to minimize NSE. For this proof we construct a probability distribution on ordered pairs of nodes in a flow network. The first node, referred to as the chosen node and denoted by X is chosen with probability proportional to its flow–i.e. $\mathbb{P}(X = i) = \frac{f_i}{\sum_i f_j}$. The second node is a random variable called the receiver and denoted Y is then chosen with a probability distribution conditional on X that is equivalent to the probability distribution of receivers used in the definition of NSE. That is, $\mathbb{P}(Y = j | X = i) = \mathcal{P}_i(j) = \frac{\tilde{q}_{ij}}{\sum_k \tilde{q}_{ik}}$.

Defining this probability distribution enables us to use the information theoretic concept of conditional entropy. In Section 5.9.9 we will define the entropy of the receiver Y conditioned on the chosen node X, denoted $H(Y|X)$, and its relation to the NME and NSE. We state and prove a theorem in terms of $H(Y|X)$ that is equivalent to Theorem 5.11.

Consider a flow network q_{ij}. We refer to the entropy of Y conditioned on X where Y and X are constructed for q_{ij} using the definition above as $H(Y|X)$ on q_{ij}. In Section we derive the equation

$$H(Y|X) \text{ on } q_{ij} \quad = \quad -\frac{\text{NSE}(q_{ij})}{\sum_i f_i}.$$

In order to state the next theorem in terms of $H(Y|X)$ we take this as its definition.

Instead of attacking theorem 5.11 head on, requiring us to consider different combinations of path lengths, we can prove this theorem by showing a less complicated statement. We will prove the theorem:

Theorem 5.12. *Let $\mathcal{L}$ be a flow network with a single source and a single sink. Let $U = \{u\}$ be a network consisting of a single node. Then*

$$NSE(\mathcal{L} \to U) \quad \geq \quad NSE(U \to \mathcal{L})$$

with equality if and only if $\mathcal{L}$is a path. The equivalent statement about conditional entropy is

$$H(Y|X) \text{ on } \mathcal{L} \to U \quad \leq \quad H(Y|X) \text{ on } U \to \mathcal{L}$$

with equality if and only if $\mathcal{L}$is a path.

The intuition as to why theorem 5.12 implies theorem 5.11 is that it essentially shows that moving the subnetwork with loops to towards the sink by one is always favorable. Therefore if we shift it once towards the sink for every node in the path subnetwork containing the sink, the NSE will eventually be minimized once the path the containing the sink is completely eliminated. in section **??**. The structure of our proof conveys where most of the entropy is gained and shows why the only time equality holds is when $\mathcal{L}$ is a path.

5.9.9 NME and NSE as negative scalar multiples of conditional entropies

We are able to consider the total mixing entropy and the total sending entropies as negative constants times types of conditional entropy. For two random variables X and Y with joint

distribution $\mathbb{P}(X = x, Y = y)$ the conditional entropy is defined:

$$\begin{aligned} H(Y|X) &= \sum_{x,y} -\mathbb{P}(X = x, Y = y) \log \mathbb{P}(Y = y|X = x) \\ &= \sum_{x} \mathbb{P}(X = x) \sum_{y} -\mathbb{P}(Y = y|X = x) \log \mathbb{P}(Y = y|X = x). \end{aligned}$$

We show how the NSE is a negative constant times a conditional entropy. Consider a flow network q_{ij}. Recall NSE $= \sum_i f_i \text{NSE}_i = -\sum f_i H(\mathcal{P}_i)$where H denotes the entropy of a discrete random variable and for this discussion we take $\mathcal{P}_i(j) = \frac{\tilde{q}_{ij}}{\sum_j \tilde{q}_{ij}}$. Then we can normalize f_i to a probability of picking a node which we denote $\phi_i := \mathbb{P}(X = i) = \frac{f_i}{\sum_j f_j}$ where we represent the random variable of choosing a node by X. Then the distributions at each node can be considered as a conditional distribution $\mathbb{P}(Y = j|X = i) = \mathcal{P}_i(j)$ where we represent drawing a receiver by Y. We construct the joint distribution $\mathbb{P}(X = i, Y = j) = \mathbb{P}(X = i)\mathbb{P}(Y = j|X = i)$. Then NSE $= -\left(\sum_l f_l\right) H(Y|X)$. To see the calculation for this:

$$\begin{aligned} \text{NSE} &= \sum_i f_i \left[\sum_{j:\tilde{q}_{ij}>0} \frac{\tilde{q}_{ij}}{\sum_j \tilde{q}_{ij}} \log \left(\frac{\tilde{q}_{ij}}{\sum_j \tilde{q}_{ij}} \right) \right] \\ &= \left(\sum_l f_l \right) \sum_i \frac{f_i}{\sum_l f_l} \left[- \sum_{j:\tilde{q}_{ij}>0} -\mathcal{P}_i(j) \log \left(\mathcal{P}_i(j) \right) \right] \\ &= -\left(\sum_l f_l \right) \sum_i \mathbb{P}(X = i) \left[\sum_{j:\tilde{q}_{ij}>0} -\mathbb{P}(Y = j|X = i) \log \left(\mathbb{P}(Y = j|X = i) \right) \right] \\ &= -\left(\sum_l f_l \right) H(Y|X). \end{aligned}$$

We can also express the NME as a negative constant times a conditional entropy. Let Z represent the distribution of drawing a signal so that $\mathbb{P}(Z = j|X = i) = \frac{\tilde{q}_{ji}}{\sum_j \tilde{q}_{ji}}$. By a similar computation, NME $= -\left(\sum_l f_l\right) H(Z|X)$.

In the next section we use $H(Y|X)$ rather than NSE when we prove theorem 5.12

because interpretation as a conditional entropy allows us to use basic information theoretic manipulations to put $H(Y|X)$ into a form which sheds light on the proof. We are aided in this by the fact that $\sum_l f_l$ is the same for both $U \to \mathcal{L}$ and $\mathcal{L} \to U$.

5.9.10 Proof of theorem 5.12

Before we continue with the proof we introduce definitions and notation for added clarity. Let $\mathcal{L}$ be a flow network with a single source and a single sink with total flow 1. Suppose that $\mathcal{L}$ has M nodes and the nodes are labeled $1, 2, \ldots, M$ where 1 is the source and M is the sink. Let $U = \{u\}$ be a flow network on one node.

- Let $L_{\mathcal{L}} = \sum_{l \in \mathcal{L}} f_l$ denote the average path length from source to sink, i.e. the average number of nodes a path following the markov chain $T_{ij} = \frac{q_{ij}}{\sum_k q_{ik}}$ starting at the source and ending at the sink of network $\mathcal{L}$. To see why this is the average path-length recall that (cross-ref) $f_i = P_{\text{source}i}$ for a network. And so $\sum_{l \in \mathcal{L}} f_l = \sum_{l \in \mathcal{L}} P_{\text{source of } \mathcal{L},l}$, which is equal to the average number of nodes a random walk takes from source to sink.

- Let ϕ_i be the probability of choosing a node in either $U \to \mathcal{L}$ or $\mathcal{L} \to U$ based on the flows f_i. That is $\phi_i = \frac{f_i}{\sum_{l \in U \cup \mathcal{L}} f_l}$.

- Let ψ_i be the probability of choosing node i in the network $\mathcal{L}$ based on the flows f_i. That is, $\psi_i = \frac{f_i}{\sum_{l \in \mathcal{L}} f_l}$.

- Define $\mathcal{P}_{\mathcal{L}i}(j)$ to be the conditional probability of choosing a receiver $j \in \mathcal{L}$ after choosing a node $i \in \mathcal{L}$ for the flow network $\mathcal{L}$. That is $\mathcal{P}_{\mathcal{L}i}(j) = \frac{\tilde{q}_{ij}}{\sum_{k \in \mathcal{L}: \tilde{q}_{ik} > 0} \tilde{q}_{ik}} = \frac{P_{ij}}{\sum_{k \in \mathcal{L}: P_{ik} > 0} P_{ik}}$.

- For $i \in \mathcal{L}$ let $\ell_i = \sum_{j \in \mathcal{L}, P_{ij} > 0} P_{ij}$. This is the average number of nodes a random walk that starts at i sees, including i, until it reaches the sink of $\mathcal{L}$.

- Choosing nodes in $\mathcal{L}$ is a random variable with probability of node i being ψ_i and we denote the entropy of this distribution as $H(\text{nodes}\mathcal{L}) = \sum_{i \in \mathcal{L}} -\psi_i \log \psi_i$.

We now come up with a general equation for the conditional entropy which is valid for both $\mathcal{L} \to U$ and $U \to \mathcal{L}$. As in secttion 5.9.9 let the random variable X be the choosen node and the variable Y be the receiver both chosen from the either of the connected unions $\mathcal{L} \to U$ or $U \to \mathcal{L}$. Let $A, B \in \{U, \mathcal{L}\}$ be set-valued random variables where A represents choosing the chosen node X to be in set A and B represents choosing the receiver to be in set B. The chain rule for entropy states $H(X,Y) = H(Y|X) + H(X)$. This rule can be generalized to an arbitrary set of variables $X_1, X_2, \ldots, X_n$ to be

$$H(X_1,X_2,\ldots,X_n) = H(X_1|X_2,X_3,\ldots,X_n)+H(X_2|X_3,X_4,\ldots,X_n)+\cdots+H(X_{n-1}|X_n)+H(X_n).$$

Then applying this to X, Y and B we get $H(X,Y) = H(Y|B,X) + H(B|X) + H(X)$ and $H(Y|X) = H(X,Y) - H(X)$. Combining these two equations we get a specialized form of the usual chain rule:

$$H(Y|X) \quad = \quad H(Y|B,X) + H(B|X).$$

Written using the definitions of conditional entropies and also conditioning on the set X lies in, this is

$$\begin{aligned} H(Y|X) \quad = \quad & \sum_{x \in U \cup \mathcal{L}} \sum_{B \in \{U,\mathcal{L}\}} \mathbb{P}(X=x)\mathbb{P}(Y \in B|X=x) \sum_{y \in U \cup \mathcal{L}} -\mathbb{P}(Y=y|Y \in B, X=x) \log\left(\mathbb{P}(Y=y|Y \in B\right. \\ + \quad & \sum_{x \in U \cup \mathcal{L}} \mathbb{P}(X=x) \sum_{B \in \{U,\mathcal{L}\}} -\mathbb{P}(Y \in B|X=x) \log\left(\mathbb{P}(Y \in B|X=x)\right). \end{aligned}$$

Since there are only 4 pairs of sets from the partition of the network $U, \mathcal{L}$ we write down the

8 summands:

$$\mathbb{P}(X=u)\mathbb{P}(Y\in U|X=u)\left[-\mathbb{P}(Y=u|Y\in U,X=u)\log\left(\mathbb{P}(Y=u|Y\in U,X=u)\right)\right]+$$
$$\mathbb{P}(X=u)\mathbb{P}(Y\in\mathcal{L}|X=u)\sum_{i\in\mathcal{L}}-\mathbb{P}(Y=i|Y\in\mathcal{L},X=u)\log\left(\mathbb{P}(Y=i|Y\in\mathcal{L},X=u)\right)+$$
$$\sum_{i\in\mathcal{L}}\mathbb{P}(X=i)\mathbb{P}(Y\in U|X=i)\left[-\mathbb{P}(Y=u|Y\in U,X=i)\log\left(\mathbb{P}(Y=u|Y\in U,X=i)\right)\right]+$$
$$\sum_{i\in\mathcal{L}}\mathbb{P}(X=i)\mathbb{P}(Y\in\mathcal{L}|X=i)\sum_{j\in\mathcal{L}}-\mathbb{P}(Y=j|Y\in\mathcal{L},X=i)\log\left(\mathbb{P}(Y=j|Y\in\mathcal{L},X=i)\right)+$$
$$\mathbb{P}(X=u)\left[-\mathbb{P}(Y\in U|X=u)\log\left(\mathbb{P}(Y\in U|X=u)\right)\right]+$$
$$\mathbb{P}(X=u)\left[-\mathbb{P}(Y\in\mathcal{L}|X=u)\log\left(\mathbb{P}(Y\in\mathcal{L}|X=u)\right)\right]+$$
$$\sum_{i\in\mathcal{L}}\mathbb{P}(X=i)\left[-\mathbb{P}(Y\in U|X=i)\log\left(\mathbb{P}(Y\in U|X=i)\right)\right]+$$
$$\sum_{i\in\mathcal{L}}\mathbb{P}(X=i)\left[-\mathbb{P}(Y\in\mathcal{L}|X=i)\log\left(\mathbb{P}(Y\in\mathcal{L}|X=i)\right)\right]$$

We now derive expressions for the conditional entropy on $\mathcal{L}\to U$ and $U\to\mathcal{L}$.

We define

$$\begin{aligned}F_0 &= H(Y|X) \text{ on } \mathcal{L}\to U\\ F_1 &= H(Y|X) \text{ on } U\to\mathcal{L}.\end{aligned}$$

We compute exact equations in terms of the marginal and conditional probabilities of choosing a node and choosing a receiver for the sending entropy. See section 5.9.11 for the direct calculations. Plugging the computations of the conditional probabilities for $U\to\mathcal{L}$

into the expression for $H(Y|X)$ we get

$$F_0 = 0 + 0 + 0 + \sum_{i \in \mathcal{L}} \left(\frac{L_{\mathcal{L}}}{L_{\mathcal{L}} + 1} \right) \psi_i \left(\frac{\ell_i}{\ell_i + 1} \right) \sum_{j \in \mathcal{L}} -\mathcal{P}_{\mathcal{L}i}(j) \log \left(\mathcal{P}_{\mathcal{L}i}(j) \right) + 0 + 0 +$$
$$- \sum_{i \in \mathcal{L}} \left(\frac{L_{\mathcal{L}}}{L_{\mathcal{L}} + 1} \right) \psi_i \left[-\frac{1}{\ell_i + 1} \log \left(\frac{1}{\ell_i + 1} \right) \right] + \sum_{i \in \mathcal{L}} \left(\frac{L_{\mathcal{L}}}{L_{\mathcal{L}} + 1} \right) \psi_i \left[-\frac{\ell_i}{\ell_i + 1} \log \left(\frac{\ell_i}{\ell_i + 1} \right) \right],$$

so

$$F_0 = \left(\frac{L_{\mathcal{L}}}{L_{\mathcal{L}} + 1} \right) \sum_{i \in \mathcal{L}} \left(\frac{\ell_i}{\ell_i + 1} \right) \psi_i \left[\sum_{j \in \mathcal{L}} -\mathcal{P}_{\mathcal{L}i}(j) \log \left(\mathcal{P}_{\mathcal{L}i}(j) \right) \right] +$$
$$\left(\frac{L_{\mathcal{L}}}{L_{\mathcal{L}} + 1} \right) \sum_{i \in \mathcal{L}} \psi_i \left[-\frac{1}{\ell_i + 1} \log \left(\frac{1}{\ell_i + 1} \right) - \frac{\ell_i}{\ell_i + 1} \log \left(\frac{\ell_i}{\ell_i + 1} \right) \right].$$

Plugging the computions of the conditional probabilities for $U \to \mathcal{L}$ into the expression for $H(Y|X)$ we get

$$F_1 = 0 + \left(\frac{1}{1 + L_{\mathcal{L}}} \right) \left(\frac{L_{\mathcal{L}}}{L_{\mathcal{L}} + 1} \right) \left[\sum_{i \in \mathcal{L}} -\psi_i \log \left(\psi_i \right) \right] + 0 + \sum_{i \in \mathcal{L}} \left(\frac{L_{\mathcal{L}}}{1 + L_{\mathcal{L}}} \right) \psi_i \left[\sum_{j \in \mathcal{L}} -\mathcal{P}_{\mathcal{L}i}(j) \log \left(\mathcal{P}_{\mathcal{L}i}(j) \right) \right] +$$
$$\frac{1}{L_{\mathcal{L}} + 1} \left[- \left(\frac{1}{L_{\mathcal{L}} + 1} \right) \log \left(\frac{1}{L_{\mathcal{L}} + 1} \right) \right] + \frac{1}{L_{\mathcal{L}} + 1} \left[- \left(\frac{L_{\mathcal{L}}}{L_{\mathcal{L}} + 1} \right) \log \left(\frac{L_{\mathcal{L}}}{L_{\mathcal{L}} + 1} \right) \right] + 0 + 0,$$

so

$$F_1 = \left(\frac{1}{1 + L_{\mathcal{L}}} \right) \left(\frac{L_{\mathcal{L}}}{L_{\mathcal{L}} + 1} \right) H(\text{nodes } \mathcal{L}) + \left(\frac{L_{\mathcal{L}}}{1 + L_{\mathcal{L}}} \right) \sum_{i \in \mathcal{L}} \psi_i H(\mathcal{P}_{\mathcal{L}i}) +$$
$$\left(\frac{1}{L_{\mathcal{L}} + 1} \right) \left[- \left(\frac{1}{L_{\mathcal{L}} + 1} \right) \log \left(\frac{1}{L_{\mathcal{L}} + 1} \right) - \left(\frac{L_{\mathcal{L}}}{L_{\mathcal{L}} + 1} \right) \log \left(\frac{L_{\mathcal{L}}}{L_{\mathcal{L}} + 1} \right) \right].$$

To prove theorem 5.12 we need to show that $F_0 \leq F_1$ with equality if and only if $\mathcal{L}$ is a path. This is the contents of the next lemma, lemma 5.12. Now we prove some facts that we use in the proof.

Fact 5.7. *For the source of* $\mathcal{L}$, $\mathcal{P}_{\mathcal{L}1}(j) = \psi_i$ *because* $\mathcal{P}_{\mathcal{L}1}(j) = \frac{P_{1j}}{\sum_{k\in\mathcal{L}:P_{1k}>0} P_{1k}} = \frac{f_j}{\sum_{l\in\mathcal{L}} f_l}$. *Thus* $H(\mathcal{P}_{\mathcal{L}1}) = H(nodes\mathcal{L})$.

Fact 5.8. *For node* $i \in \mathcal{L}$,

$$\max_j \mathcal{P}_{\mathcal{L}i}(j) = \mathcal{P}_{\mathcal{L}i}(i) = \frac{1}{\sum_{k\in\mathcal{L}:P_{ik}>0} P_{ik}} = \frac{1}{\ell_i}.$$

We also have the basic lemma giving a lower bound for the entropy of finite probability distributions

Lemma 5.11. *Let* $n \geq 1$ *be a positive integer. If* $P(j) = p_j$ $j = 1, 2, \ldots n$ *is a discrete probability distribution* n *distinct atoms having non-zero probability then* $H(P) \geq \log((\max_j p_j)^{-1})$ *with equality if and only if* $p_i = \max_j p_j = \frac{1}{n}$ *for all* i.

Proof. For element i we have $p_i \log(p_i^{-1}) \geq p_i \log(\min_j(p_j^{-1})) = p_i \log((\max_j p_j)^{-1})$ with equality if and only if $p_i = \max_j p_j$ for all i. Therefore

$$H(P) = \sum_i p_i \log(p_i^{-1}) \geq \sum_i p_i \log((\max_j p_j)^{-1}) = \log((\max_j p_j)^{-1})$$

with equality if and only if p_i is uniform. Since there are n different atoms with non-zero probability, $p_i \equiv \frac{1}{n}$. □

Lemma 5.12. *Let* U *be a flow network consisting of a single node and let* $\mathcal{L}$ *be an arbitrary network with total flow* 1 *and boundary flows consisting of a single source and a single sink. Then the conditional entropy* $H(Y|X)$ *for the network* $U \to \mathcal{L}$, F_1, *is greater than or equal to the conditional entropy* $H(Y|X)$ *for the network* $\mathcal{L} \to U$ *with equality if and only if* $\mathcal{L}$ *is a path.*

Proof. The first part we show is the easiest, that there is equality if $\mathcal{L}$ is a path. Since $\mathcal{L}$ is a path flow network, so are both $\mathcal{L} \to U$ and $U \to \mathcal{L}$ and they are paths of the same length, so $H(Y|X)$ is the same for both networks.

To prove the lemma it suffices to show that $F_1 \geq F_0$ with equality only if $\mathcal{L}$ is a path. We re-state the equations

$$\begin{aligned} F_0 &= \left(\frac{L_{\mathcal{L}}}{L_{\mathcal{L}}+1}\right) \sum_{i \in \mathcal{L}} \left(\frac{\ell_i}{\ell_i+1}\right) \psi_i \left[\sum_{j \in \mathcal{L}} -\mathcal{P}_{\mathcal{L}i}(j) \log\left(\mathcal{P}_{\mathcal{L}i}(j)\right)\right] \\ &+ \left(\frac{L_{\mathcal{L}}}{L_{\mathcal{L}}+1}\right) \sum_{i \in \mathcal{L}} \psi_i \left[-\frac{1}{\ell_i+1} \log\left(\frac{1}{\ell_i+1}\right) - \frac{\ell_i}{\ell_i+1} \log\left(\frac{\ell_i}{\ell_i+1}\right)\right] \end{aligned} \tag{5.9.2}$$

and

$$\begin{aligned} F_1 &= \left(\frac{1}{1+L_{\mathcal{L}}}\right) \left(\frac{L_{\mathcal{L}}}{L_{\mathcal{L}}+1}\right) H(\text{nodes } \mathcal{L}) + \left(\frac{L_{\mathcal{L}}}{1+L_{\mathcal{L}}}\right) \sum_{i \in \mathcal{L}} \psi_i H(\mathcal{P}_{\mathcal{L}i}) \\ &+ \left(\frac{1}{L_{\mathcal{L}}+1}\right) \left[-\left(\frac{1}{L_{\mathcal{L}}+1}\right) \log\left(\frac{1}{L_{\mathcal{L}}+1}\right) - \left(\frac{L_{\mathcal{L}}}{L_{\mathcal{L}}+1}\right) \log\left(\frac{L_{\mathcal{L}}}{L_{\mathcal{L}}+1}\right)\right]. \end{aligned} \tag{5.9.3}$$

To begin the proof, we first break up F_0 into 3 parts. The first part is the first term of the sum in the first line of equation 5.9.2:

$$\begin{aligned} \left(\frac{L_{\mathcal{L}}}{L_{\mathcal{L}}+1}\right) \left(\frac{\ell_1}{\ell_1+1}\right) \psi_1 \left[\sum_{j \in \mathcal{L}} -\mathcal{P}_{\mathcal{L}1}(j) \log\left(\mathcal{P}_{\mathcal{L}1}(j)\right)\right] &= \left(\frac{L_{\mathcal{L}}}{L_{\mathcal{L}}+1}\right) \left(\frac{L_{\mathcal{L}}}{L_{\mathcal{L}}+1}\right) \left(\frac{1}{L_{\mathcal{L}}}\right) H(\mathcal{P}_{\mathcal{L}1}) \\ &= \left(\frac{1}{L_{\mathcal{L}}+1}\right) \left(\frac{L_{\mathcal{L}}}{L_{\mathcal{L}}+1}\right) H(\text{nodes } \mathcal{L}) \end{aligned}$$

by fact 5.7. The second part is the first term of the sum in the second line of equation

5.9.2:

$$\begin{aligned}
&\left(\frac{L_{\mathcal{L}}}{L_{\mathcal{L}}+1}\right)\psi_1\left[-\frac{1}{\ell_1+1}\log\left(\frac{1}{\ell_1+1}\right)-\frac{\ell_1}{\ell_1+1}\log\left(\frac{\ell_1}{\ell_1+1}\right)\right]\\
&=\left(\frac{L_{\mathcal{L}}}{L_{\mathcal{L}}+1}\right)\psi_1\left[-\frac{1}{L_{\mathcal{L}}+1}\log\left(\frac{1}{L_{\mathcal{L}}+1}\right)-\frac{L_{\mathcal{L}}}{L_{\mathcal{L}}+1}\log\left(\frac{L_{\mathcal{L}}}{L_{\mathcal{L}}+1}\right)\right]\\
&=\left(\frac{L_{\mathcal{L}}}{L_{\mathcal{L}}+1}\right)\frac{1}{L_{\mathcal{L}}}\left[-\frac{1}{L_{\mathcal{L}}+1}\log\left(\frac{1}{L_{\mathcal{L}}+1}\right)-\frac{L_{\mathcal{L}}}{L_{\mathcal{L}}+1}\log\left(\frac{L_{\mathcal{L}}}{L_{\mathcal{L}}+1}\right)\right]\\
&=\left(\frac{1}{L_{\mathcal{L}}+1}\right)\left[-\frac{1}{L_{\mathcal{L}}+1}\log\left(\frac{1}{L_{\mathcal{L}}+1}\right)-\frac{L_{\mathcal{L}}}{L_{\mathcal{L}}+1}\log\left(\frac{L_{\mathcal{L}}}{L_{\mathcal{L}}+1}\right)\right].
\end{aligned}$$

The third part is the remaining sums over the nodes in $\mathcal{L}$ from node 2 to the sink:

$$\begin{aligned}
&\left(\frac{L_{\mathcal{L}}}{L_{\mathcal{L}}+1}\right)\sum_{i\neq 1\in\mathcal{L}}\left(\frac{\ell_i}{\ell_i+1}\right)\psi_i\left[\sum_{j\in\mathcal{L}}-\mathcal{P}_{\mathcal{L}i}(j)\log\left(\mathcal{P}_{\mathcal{L}i}(j)\right)\right]+\\
&\qquad\left(\frac{L_{\mathcal{L}}}{L_{\mathcal{L}}+1}\right)\sum_{i\neq 1\in\mathcal{L}}\psi_i\left[-\frac{1}{\ell_i+1}\log\left(\frac{1}{\ell_i+1}\right)-\frac{\ell_i}{\ell_i+1}\log\left(\frac{\ell_i}{\ell_i+1}\right)\right].
\end{aligned}$$

Note that the first and second parts of the expression for F_0 are the first and third sums in

the equation 5.9.3 for F_1. Therefore it suffices to show that

$$\left(\frac{L_{\mathcal{L}}}{L_{\mathcal{L}}+1}\right)\sum_{i\neq 1\in\mathcal{L}}\left(\frac{\ell_i}{\ell_i+1}\right)\psi_i\left[\sum_{j\in\mathcal{L}}-\mathcal{P}_{\mathcal{L}i}(j)\log\left(\mathcal{P}_{\mathcal{L}i}(j)\right)\right]+$$
$$\left(\frac{L_{\mathcal{L}}}{L_{\mathcal{L}}+1}\right)\sum_{i\neq 1\in\mathcal{L}}\psi_i\left[-\frac{1}{\ell_i+1}\log\left(\frac{1}{\ell_i+1}\right)-\frac{\ell_i}{\ell_i+1}\log\left(\frac{\ell_i}{\ell_i+1}\right)\right]$$
$$\leq$$
$$\left(\frac{L_{\mathcal{L}}}{1+L_{\mathcal{L}}}\right)\sum_{i\in\mathcal{L}}\psi_i\left[\sum_{j\in\mathcal{L}}-\mathcal{P}_{\mathcal{L}i}(j)\log\left(\mathcal{P}_{\mathcal{L}i}(j)\right)\right]$$

with equality only if $\mathcal{L}$ is a path. First we rearrange the left side of the inequality by adding zero in the form of $\sum_{i\neq 1\in\mathcal{L}}\frac{1}{\ell_i+1}\psi_i\log(\ell_i)-\sum_{i\neq 1\in\mathcal{L}}\frac{1}{\ell_i+1}\psi_i\log(\ell_i)$ and distributing the terms over the first and second sums. That is

$$\left(\frac{L_{\mathcal{L}}}{L_{\mathcal{L}}+1}\right)\sum_{i\neq 1\in\mathcal{L}}\left(\frac{\ell_i}{\ell_i+1}\right)\psi_i\left[\sum_{j\in\mathcal{L}}-\mathcal{P}_{\mathcal{L}i}(j)\log\left(\mathcal{P}_{\mathcal{L}i}(j)\right)\right]+$$
$$\left(\frac{L_{\mathcal{L}}}{L_{\mathcal{L}}+1}\right)\sum_{i\neq 1\in\mathcal{L}}\psi_i\left[-\frac{1}{\ell_i+1}\log\left(\frac{1}{\ell_i+1}\right)-\frac{\ell_i}{\ell_i+1}\log\left(\frac{\ell_i}{\ell_i+1}\right)\right]$$
$$=$$
$$\left(\frac{L_{\mathcal{L}}}{L_{\mathcal{L}}+1}\right)\sum_{i\neq 1\in\mathcal{L}}\psi_i\left[\left(\frac{\ell_i}{\ell_i+1}\right)H(\mathcal{P}_{\mathcal{L}i})+\left(\frac{1}{\ell_i+1}\right)\log(\ell_i)\right]+$$
$$\left(\frac{L_{\mathcal{L}}}{L_{\mathcal{L}}+1}\right)\sum_{i\neq 1\in\mathcal{L}}\psi_i\left[-\frac{1}{\ell_i+1}\log\left(\frac{1}{\ell_i+1}\right)-\frac{\ell_i}{\ell_i+1}\log\left(\frac{\ell_i}{\ell_i+1}\right)-\left(\frac{1}{\ell_i+1}\right)\log(\ell_i)\right].$$

By fact 5.8, $\max_j \mathcal{P}_{\mathcal{L}i}(j)=\frac{1}{\ell_i}$, and by lemma 5.11, we have that $H(\mathcal{P}_{\mathcal{L}i})\geq\log(\ell_i)$. Applying

this inequality to the first summand above, we obtain the inequality

$$\left(\frac{L_{\mathcal{L}}}{L_{\mathcal{L}}+1}\right)\sum_{i\neq 1\in\mathcal{L}}\left(\frac{\ell_i}{\ell_i+1}\right)\psi_i H(\mathcal{P}_{\mathcal{L}i})+\left(\frac{L_{\mathcal{L}}}{L_{\mathcal{L}}+1}\right)\sum_{i\neq 1\in\mathcal{L}}\psi_i\left[-\frac{1}{\ell_i+1}\log\left(\frac{1}{\ell_i+1}\right)-\frac{\ell_i}{\ell_i+1}\log\left(\frac{\ell_i}{\ell_i+1}\right)\right]$$

$$\leq$$

$$\left(\frac{L_{\mathcal{L}}}{L_{\mathcal{L}}+1}\right)\sum_{i\neq 1\in\mathcal{L}}\psi_i H(\mathcal{P}_{\mathcal{L}i})+\left(\frac{L_{\mathcal{L}}}{L_{\mathcal{L}}+1}\right)\sum_{i\neq 1\in\mathcal{L}}\psi_i\left[-\frac{1}{\ell_i+1}\log\left(\frac{1}{\ell_i+1}\right)-\frac{\ell_i}{\ell_i+1}\log\left(\frac{\ell_i}{\ell_i+1}\right)-\left(\frac{1}{\ell_i+1}\right)\log\right.$$

Now we upper bound the sum:

$$\left(\frac{L_{\mathcal{L}}}{L_{\mathcal{L}}+1}\right)\sum_{i\neq 1\in\mathcal{L}}\psi_i\left[-\frac{1}{\ell_i+1}\log\left(\frac{1}{\ell_i+1}\right)-\frac{\ell_i}{\ell_i+1}\log\left(\frac{\ell_i}{\ell_i+1}\right)-\left(\frac{1}{\ell_i+1}\right)\log\left(\ell_i\right)\right]$$

$$=$$

$$\left(\frac{L_{\mathcal{L}}}{L_{\mathcal{L}}+1}\right)\sum_{i\neq 1\in\mathcal{L}}\psi_i\left[-\frac{1}{\ell_i+1}\log\left(\frac{\ell_i}{\ell_i+1}\right)-\frac{\ell_i}{\ell_i+1}\log\left(\frac{\ell_i}{\ell_i+1}\right)\right]$$

$$=$$

$$\left(\frac{L_{\mathcal{L}}}{L_{\mathcal{L}}+1}\right)\sum_{i\neq 1\in\mathcal{L}}\psi_i\left[\log\left(\ell_i+1\right)-\log\left(\ell_i\right)\right]$$

by the first term, $\left(\frac{L_{\mathcal{L}}}{1+L_{\mathcal{L}}}\right)\psi_1 H(\mathcal{P}_{\mathcal{L}1})$ of the sum $\left(\frac{L_{\mathcal{L}}}{1+L_{\mathcal{L}}}\right)\sum_{i\in\mathcal{L}}\psi_i H(\mathcal{P}_{\mathcal{L}i})$ in the equation for F_1.

Recall that $T_{ij}=\frac{q_{ij}\mathbf{1}_{q_{ij}>0}}{\sum_{k\in n(i):q_{ik}>0}q_{ik}}$ for all $j\in n(i)$. Let node i be a node in $\mathcal{L}\backslash$sink. Then we have the expression

$$\ell_i=\left(\sum_{j\in n(i):q_{ij}>0}T_{ij}\ell_j\right)+1=\sum_{j\in n(i):q_{ij}>0}\left(T_{ij}\ell_j+T_{ij}\right)$$

By concavity of the logarithm, $\log(\ell_i) \geq \sum_j T_{ij} \log(\ell_j + 1)$. Thus

$$\psi_i \log(\ell_i) = \frac{f_i}{\sum_{k \in \mathcal{L}} f_k} \log(\ell_i) \quad \geq \quad \frac{1}{\sum_{k \in \mathcal{L}} f_k} \sum_j f_i T_{ij} \log(\ell_j + 1) = \frac{1}{\sum_{k \in \mathcal{L}} f_k} \sum_j q_{ij} \log(\ell_j + 1)$$

Summing $\psi_i \log(\ell_i)$ over the nodes in $\mathcal{L}$ and using the last inequality, we get

$$\sum_{i \in \mathcal{L}} \psi_i \log(\ell_i) = \sum_{i \neq \text{sink} \in \mathcal{L}} \psi_i \log(\ell_i) \quad \geq \quad \sum_{i \neq \text{sink} \in \mathcal{L}} \left(\frac{1}{\sum_{k \in \mathcal{L}} f_k} \right) \sum_{j : q_{ij} > 0} q_{ij} \log(\ell_j + 1)$$

By swapping the order of the sums we have

$$\begin{aligned}
\sum_{i \neq \text{sink} \in \mathcal{L}} \left(\frac{1}{\sum_{k \in \mathcal{L}} f_k} \right) \sum_{j : q_{ij} > 0} q_{ij} \log(\ell_j + 1) &= \sum_{i \neq \text{sink} \in \mathcal{L}} \left(\frac{1}{\sum_{k \in \mathcal{L}} f_k} \right) \sum_{j \neq 1 \in \mathcal{L}} \mathbf{1}_{\{q_{ij} > 0\}} q_{ij} \log(\ell_j + 1) \\
&= \sum_{i \in \mathcal{L}} \left(\frac{1}{\sum_{k \in \mathcal{L}} f_k} \right) \sum_{j \in \mathcal{L}} \mathbf{1}_{\{q_{ij} > 0\}} \mathbf{1}_{\{i \neq \text{sink} j \neq 1\}} q_{ij} \log(\ell_j + 1) \\
&= \sum_{j \neq 1 \in \mathcal{L}} \left(\frac{1}{\sum_{k \in \mathcal{L}} f_k} \right) \log(\ell_j + 1) \left(\sum_{i \neq \text{sink} \in \mathcal{L}} \mathbf{1}_{q_{ij} > 0} q_{ij} \right) = \sum_{j \neq 1 \in \mathcal{L}} \left(\frac{}{\sum} \right. \\
&= \sum_{j \neq 1 \in \mathcal{L}} \psi_j \log(\ell_j + 1)
\end{aligned}$$

where the condition of not being the sink is removed from domain of the second sum because it is enough that the indicator function $\mathbf{1}_{q_{ij} > 0}$ is zero for all j when i is the sink. Therefore combining $\sum_{i \in \mathcal{L}} \psi_i \log(\ell_i) = \psi_1 \log(\ell_1) + \sum_{i \neq 1 \in \mathcal{L}} \psi_i \log(\ell_i)$ and $\sum_{i \in \mathcal{L}} \psi_i \log(\ell_i) \geq \sum_{i \neq 1 \in \mathcal{L}} \psi_i \log(\ell_i + 1)$ we have by lemma 5.11

$$\begin{aligned}
\psi_1 H(\mathcal{P}_{\mathcal{L}1}) \geq \psi_1 \log(\ell_1) \quad &\geq \quad \sum_{i \neq 1 \in \mathcal{L}} \psi_i \log(\ell_i + 1) - \sum_{i \neq 1 \in \mathcal{L}} \psi_i \log(\ell_i) \\
&= \quad \sum_{i \neq 1 \in \mathcal{L}} \psi_i (\log(\ell_i + 1) - \log(\ell_i)).
\end{aligned}$$

This proves the inequality on the remaining terms in the equation for F_0. Note equality only can occur if $H(\mathcal{P}_{\mathcal{L}1}) = \log(\ell_1)$, which happens only when ψ_i are all equal by lemma 5.11.

This is equivalent to the subnetwork $\mathcal{L}$ being a path. That is $\psi_1 H(\mathcal{P}_{\mathcal{L}1}) = \sum_{i \neq 1 \in \mathcal{L}} \psi_i (\log(\ell_i + 1) - \log(\ell_i))$ only if $\mathcal{L}$ is a path. Therefore, we derive

$$\left(\frac{L_{\mathcal{L}}}{L_{\mathcal{L}}+1}\right) \sum_{i \neq 1 \in \mathcal{L}} \left(\frac{\ell_i}{\ell_i+1}\right) \psi_i H(\mathcal{P}_{\mathcal{L}i}) + \left(\frac{L_{\mathcal{L}}}{L_{\mathcal{L}}+1}\right) \sum_{i \neq 1 \in \mathcal{L}} \psi_i \left[-\frac{1}{\ell_i+1} \log\left(\frac{1}{\ell_i+1}\right) - \frac{\ell_i}{\ell_i+1} \log\left(\frac{\ell_i}{\ell_i+1}\right) \right]$$
$$\leq \left(\frac{L_{\mathcal{L}}}{L_{\mathcal{L}}+1}\right) \sum_{i \in L} \psi_i H(\mathcal{P}_{\mathcal{L}i}).$$

with equality only if $\mathcal{L}$ is a path. This completes the proof that $F_1 \geq F_0$ with equality only if $\mathcal{L}$ is a path. □

In proving the lemma we have also proven the equivalent statement Theorem 5.12. Theorem 0.2 then follows by an iterative application of the argument above. Suppose that we have a combination of subnetworks: $\tau_{n-t} \to \mathcal{R} \to \tau_t$ Then assuming $t \neq 0$, we may rewrite this network as a combination: $\mathcal{L} \to u$, where $\mathcal{L} = \tau_{n-t} \to \mathcal{R} \to \tau_{t-1}$ and u is the sink vertex. Theorem 5.12 then tells us that the NSE for this network is larger than for the network: $u \to \mathcal{L}$; a network in which a node has been removed from τ_t and appended to τ_{n-t}. By iteratively applying this argument, we eliminate all nodes from the sink path, and create a network with lower NSE.

5.9.11 Computation of marginal and conditional probabilities needed for $H(Y|X)$ for networks $\mathcal{L} \to U$ and $U \to \mathcal{L}$

We compute the probabilities of choosing a node for either network :

$$\begin{aligned} \mathbb{P}(X = u) &= \frac{1}{\sum_{l \in U \cup \mathcal{L}} f_l} = \frac{1}{1 + L_{\mathcal{L}}} \\ \mathbb{P}(X = i) &= \left(\frac{L_{\mathcal{L}}}{1 + L_{\mathcal{L}}}\right) \psi_i. \end{aligned}$$

And the probabilities of the chosen node being in either $\mathcal{L}$ or U:

$$\begin{aligned}\mathbb{P}(X \in U) &= \frac{1}{L_{\mathcal{L}}+1} \\ \mathbb{P}(X \in \mathcal{L}) &= \frac{L_{\mathcal{L}}}{L_{\mathcal{L}}+1}.\end{aligned}$$

Consider the flow $\mathcal{L} \to U$. We compute the conditional probabilities to compute equation 5.9.1. The probabilities of the receiver lying in a subnetwork given a chosen node are

$$\begin{aligned}\mathbb{P}(Y \in U|X=u) &= 1 \\ \mathbb{P}(Y \in U|X=i) &= \frac{1}{\sum_{k\in U\cup\mathcal{L}} P_{ik}} = \frac{1}{\ell_i+1} \text{ for } i \in \mathcal{L} \\ \mathbb{P}(Y \in \mathcal{L}|X=u) &= 0 \\ \mathbb{P}(Y \in \mathcal{L}|X=i) &= \frac{\ell_i}{\ell_i+1}.\end{aligned}$$

The conditional probabilities of a receiver given that the receiver lies in a subnetwork for specific chosen nodes are

$$\begin{aligned}\mathbb{P}(Y=u|Y \in U, X=u) &= 1 \\ \mathbb{P}(Y=i|Y \in \mathcal{L}, X=u) &= 0 \text{ for } i \in \mathcal{L} \\ \mathbb{P}(Y=u|Y \in U, X=i) &= 1 \text{ for } i \in \mathcal{L} \\ \mathbb{P}(Y=j|Y \in \mathcal{L}, X=i) &= \frac{\mathbb{P}(Y=j|X=i)}{\mathbb{P}(Y \in \mathcal{L}|X=i)} = \left(\frac{P_{ij}}{\sum_{k\in U\cup\mathcal{L}} P_{ik}}\right)\left(\frac{\ell_i+1}{\ell_i}\right) \\ &= \left(\frac{P_{ij}}{\ell_i+1}\right)\left(\frac{\ell_i+1}{\ell_i}\right) = \frac{P_{ij}}{\sum_{k\in\mathcal{L}} P_{ik}} = \mathcal{P}_{\mathcal{L}i}(j) \text{ for } i,j \in \mathcal{L}.\end{aligned}$$

Consider the flow $U \to \mathcal{L}$. We compute the conditional probabilities to compute equation

5.9.1. The probabilities of the receiver lying in a subnetwork given a chosen node are

$$\begin{aligned}
\mathbb{P}(Y \in U | X = u) &= \frac{P_{uu}}{\sum_{i \in U \cup \mathcal{L}} P_{ui}} = \frac{1}{L_{\mathcal{L}} + 1} \\
\mathbb{P}(Y \in U | X = i) &= 0 \\
\mathbb{P}(Y \in \mathcal{L} | X = u) &= \frac{L_{\mathcal{L}}}{L_{\mathcal{L}} + 1} \\
\mathbb{P}(Y \in \mathcal{L} | X = i) &= 1
\end{aligned}$$

The conditional probabilities of a receiver given that the receiver lies in a subnetwork for specific chosen nodes are

$$\begin{aligned}
\mathbb{P}(Y = u | Y \in U, X = u) &= 1 \\
\mathbb{P}(Y = i | Y \in \mathcal{L}, X = u) &= \frac{\mathbb{P}(Y = i | X = u)}{\mathbb{P}(Y \in \mathcal{L} | X = u)} = \left(\frac{P_{ui}}{\sum_{j \in U \cup \mathcal{L}} P_{uj}} \right) \left(\sum_{k \in \mathcal{L}} \frac{P_{uk}}{\sum_{j \in U \cup \mathcal{L}} P_{uj}} \right)^{-1} \\
&= \left(\frac{f_i}{L_{\mathcal{L}} + 1} \right) \left(\frac{L_{\mathcal{L}} + 1}{L_{\mathcal{L}}} \right) = \frac{f_i}{L_{\mathcal{L}}} = \psi_i \text{ for } i \in \mathcal{L} \\
\mathbb{P}(Y = u | Y \in U, X = i) &= 1 \text{ for } i \in \mathcal{L} \\
\mathbb{P}(Y = j | Y \in \mathcal{L}, X = i) &= \frac{\mathbb{P}(Y = j | X = i)}{\mathbb{P}(Y \in \mathcal{L} | X = i)} = \frac{P_{ij}}{\sum_{k \in U \cup \mathcal{L}} P_{ik}} \left(\sum_{l \in \mathcal{L}} \frac{P_{il}}{\sum_{k \in U \cup \mathcal{L}} P_{ik}} \right)^{-1} \\
&= \frac{P_{ij}}{\sum_{k \in \mathcal{L}} P_{ik}} \left(\sum_{l \in \mathcal{L}} \frac{P_{il}}{\sum_{k \in \mathcal{L}} P_{ik}} \right)^{-1} = \mathcal{P}_{\mathcal{L}i}(j) \text{ for } i, j \in \mathcal{L}.
\end{aligned}$$

CHAPTER 6

Optimization Algorithm

6.1 Optimization description

Our optimization algorithm is comprised of two main parts. One step of the algorithm consists of performing the two parts, one after the other. The number of steps of the algorithm are specified and the algorithm progresses by taking the result of the previous step (which we call the recent network) as the initialization of the current step, performing the two main parts to create what we call the current network. Both a total number of steps and a stopping criterion are specified so that the number of steps used is both thorough and efficient.

The first main part is a discontinuous jump from the last network of conductances to a new network of conductances. This is because the local optimacan each only be accessed from a finite region of conductance values and we need to jump from region of conductance values to another.

The second main part is to initialize a continuous optimization algorithm (e.g. gradient descent) to approximate the optimum when restricted to the current region of conductances. In our case we use an interior point algorithm because it allows staying in the current region by enforcing a set of linear and non-linear constraints.

We describe quite generally how these two parts work, how we choose an optimum, how we initialize and terminate the algorithm, before addressing the specific details of the two parts of the algorithm.

6.1.1 Part 1 of algorithm: The growth step: Adding materal and changing topology

Based on observed behavior of the continuous optimization described in sections 6.3.1 and 6.3.3 and the landscape of flow directions, described in section 6.3.2 we construct a new intial point for continuous optimization based on a current local optima.

Section 6.3.1 describes how continuous optimization finds sparse conductance networks that are local optima, or that approach local optima, of the mixing dissipation function but does not find the global optimum. Therefore, more points need to be tested as initial points before we can conclude that what we have is a good approximation of the globally optimal network. Since the recent network is expected to be sparse, we devise a way to add more conductance to reintroduce vanished edges we call this the network growth step. Another important issue is that short circuits between nodes that in the global network are well separated steer the optimization off course because the continuous optimization, especially when taken together with the flow direction change step, leads to a much shorter network with a worse cost (see section 6.3.3). Therefore the way we add conductance must avoid high conductance pathways between distant nodes.

After we add material, we explore the local landscape of conductances by changing flows. To probe networks that are similar to the recent network, we swap the directions of flow on a few edges within the network (see section 6.3.2). We refer to the directed graph of the flow network disregarding the strength of the flows as the flow topology. We are exploring flow topologies that are close to the initial topology -- that agree except for a small number of flow directions. We do this by solving for how to perturb the conductances to change the topology.

We randomize the growth step by making the degree to which we grow the network, the number of flow reversals and by randomly choosing which flows to swap and which edges to perturb to engender these swaps.

6.1.2 Part 2 of algorithm: Continuous optimization

After adding material and changing topology, we input the result into a continuous optimization method to find the local optima that comes from the previous local optimum and the current jump. To make sure that we sufficiently search different flow topologies, we use an interior point method where we maintain the sign of the physical flows stemming from the conductances q_{ij} as non-linear constraints, to ensure that we do not change the topology of the network. We use the Matlab implementation of interior point method using fmincon. To boost the efficiency of interior point, we compute the gradient of the objective function using lagrange-multipliers as detailed in the Section 6.7.

All of the conductances in the network have to be non-negative. We hardwire non-negativity into the conductances, by writing our function as a precomposition of the exponential function on the set of conductances at each edge via defining the pre-exponential of each conductance $\tilde{\kappa}_{ij}$ so that $\kappa_{ij} = \exp(\tilde{\kappa}_{ij})$. To maintain the building cost of the function it is convenient to formulate the optimization as choosing the best perturbation of the conductances after the steps in part 1 of the algorithm. That is we write $\delta_{ij} \mapsto \kappa_{ij} + \delta_{ij}$ and find the optimal δ_{ij}. We combine this with the exponential precomposition to be define $\tilde{\delta}_{ij}$ to be such that $\delta_{ij} = \exp(\tilde{\kappa}_{ij} + \tilde{\delta}_{ij}) - \exp(\tilde{\kappa}_{ij})$. So the variable we are optimizing over is actually $\tilde{\delta}_{ij}$ the pre-exponential perturbation for each edge.

The last step before plugging the conductances into the mixing-dissipation cost is to normalize the building material $\kappa_{ij} \mapsto \left(\frac{C}{\sum_{(i,j)} \kappa_{ij}^{\gamma}}\right)^{\frac{1}{\gamma}} \kappa_{ij}$.

Now we describe the way both parts are applied, how the optimization routine is initialized how how we choose its termination. Also we talk about how the optimum is updated after each iteration.

6.1.3 Initialization, algorithm iterations and termination criterion

6.1.3.1 Initialization

In most of our numerical experiments we first initialize the κ_{ij} to be chosen independently and uniformly at random from $[0, 1]$ then we apply the rescaling $\kappa_{ij} \mapsto \left(\frac{C}{\sum_{(i,j)} \kappa_{ij}^{\gamma}}\right)^{\frac{1}{\gamma}} \kappa_{ij}$ so that the building cost constraint is maintained from the outset. In cases where the algorithm is unable to find good networks starting from random initial conditions we hand-pick the initialization. In our numerical experiments this happens in the 2-source 2-sink case because optimal networks may have one connected component, butthe optimization algorithm only finds a demonstrably higher cost function network with two connected components. For this we need to choose a network which is connected and we suspect is close to the local optimum and initialize the procedure with that.

6.1.3.2 Choosing the optimum at each step

At each step of the algorithm we perform the randomized growth step followed by random topology change. Then we perform the continuous optimization method. Iteratively, we replace the recent network by the new network if and only if it beats the prior optimum. This makes our algorithm a greedy algorithm with random steps.

It is worthwhile to note a previously used method of that in designing our optimization method we also experimented with using simulated annealing. We used simulated annealing with a different step scheme and the current one. Simulated annealing would guarantee that we leave local optima and eventually choose better local optima [KGV83], but proved to be highly inefficient, in that it spent a lot of time exploring, or testing topologies that offer no advantage in terms of the cost function.

We choose a greedy algorithm because,based on it being successfully able to locate the optimal network when $c = .001$ (at which value dissipation contributes negligibly to the cost function, and for which we are able to show rigorously that the optimal network is a path

that visits every node in the ambient network. This path is sometimes not reached because growing the network can sometimes introduce geometric frustrations, especially in the square grid case. In more than half of our trials, the algorithm converged to this path. Therefore it is reasonable to generate many candidate optima using the greedy algorithm and choose the best ones (see Section 6.8.1). Simulated annealing might be a better choice if in a further study we choose to focus on very large grids and specific levels of c and do in-depth studies of single samples.

6.1.3.3 Number of iterations and stopping criterion

Our growth step requires first picking a single direction in which to grow the network: up, right, down or left. To ensure that every direction is sampled, every 4 iterations we choose a new permutation of $(1, 2, 3, 4)$ representing up, right, down, and left growth directions. To be sure that we are not futily repeating the same step and to draw a clear line on when to terminate the algorithm, we believe it is best to change the direction almost every time. (We say almost because a two different permutation might have the last direction of the first permutation and the first direction of the second permutation line up).

We then count the number of times the difference between the mixing-dissipation cost of the current and previous iteration differs by less than 10^{-2} . When this reaches 4 we terminate the algorithm. Otherwise we allow the algorithm to run for 50 iterations, noting that it rarely if ever reaches this. Usually the algorithm terminates in less than 15 iterations.

It is interesting to note that for the same building cost C and weight on dissipation c besides the intial optimization on the dense network, the run time on the sparse network will be the same or decrease because the network is less tortuously packed into the grid, allowing there to be a larger available surface to add growths to.

6.2 Summary of design of the optimization algorithm

Observations:

We observe the following through attempts of optimization

1. The continuous optimization eleminates many edges. So much so that it finds local optima that are much shorter than the expected local optima.
2. Different topologies have different local maxima. An honest algorithm would hop between these segments of the domain.
3. Short circuits are very unproductive because the algorithm likes to delete the lengthy graph in favor of putting all the material in the short circuit. This is because of the dissipation term.

Therefore we want an algorithm that

1. Because of 1: Adds building material to the network, but without deleting or short circuiting the recent optimal network. That is, the recent optimal network should occur as a subset of the edges in the grown network.
2. Because of 1: Is capable of increasing the length of paths. (But also, must be capable of shortening the path)
3. Because of 2: Deliberately changes flows along edges. Purposefully swaps flows.
4. Because of 3: Avoids adding material where it would short circuit the recent optimal network.

Our algorithm answeres these needs by including the following features:

1. Need 1: The growth algorithm adds material that preserves characteristics of the flow of the network

2. Need 2: The growth algorithm can add extra length by itself in the case of single triangle growth and squares added to sharp turns. Path elongation a lso occurs during the flow swapping step.

3. Need 3: Our flow swapping algorithm is specifically added to meet this need.

4. Need 4: The growth algorithm reduces by one edge either a path with 2 edges to one edge in the case of triangles. Or a path with 3 edges to one edge. It does not bridge say, vertices on longer paths.

6.3 Observations: Cost Function land-scape and basic optimization behavior

In this section we present our method of estimating the minima and the minimizers of $\mathrm{NME}(\kappa_{ij}) + cD(\kappa_{ij})$ under the constraint $\sum_{(i,j)} \kappa_{ij}^{\gamma} = C$ and $\kappa_{ij} > 0$. To understand the way our method is designed it is first necessary to have a detailed description of the land-scape of the cost function we wish to optimize. By land-scape we mean the geometry and smoothness of the graph of the function in relation to regions in its domain. We start by investigating the results of finding local optima using continuous search methods such the interior point method via matlab's fmincon method. We discuss the pitfalls of simply applying this technique. Notably, it is guaranteed to find local optima at best, and it does not include a rigorous search through regions of the domain which might contain alternate local optima. Notably, it is blind to what we refer to as different *flow topologies.* By flow topology we refer to the unweighted directed network given by the directions of the edges in the flow network. We will see that, in the domain of conductances, there are discontinuities in the NME along values of κ_{ij} along which flows of the network switch direction. Even if fmincon produced perfect results, we would not be confident that it is adequately checking the partition of the space of positive conductancs with fixed building costs for every resulting topology of flows. We develop a way to explore this space locally using the Sherman-Morrison

formula to explore the boundaries between regions. Also, to change the number of edges with positive conductances, either increasing or decreasing the number of edges with building material, we provide a method that keeps intact much of the recent flow network while offering neighboring possibilities to add new positive conductances and flows.

The strategy we devise is not only directed to productively navigate the landscape of conductances and flows, it also does not a-priori restrict the space of searchable networks. That is, even though it is tailored to the structure of the problem, it does not assume that the optima are in a restricted set of networks. It could potentially find an optima even if it contradicts our conceptualizations of what an optima would look like.

6.3.1 Results of continuous optimization

The cost function is a function on a smoothly parametrizable domain, the constraint surface $\sum_{(i,j)} \kappa_{ij}^{\gamma} = C$, and this might seem to call for a continuous procedure such as gradient descent or an interior point method. We investigate the results on a 5×5 rectangular grid with total flow 1 and a single source and sink at the bottom left and top right corners. We initialize the optimization with uniform conductances (see Fig. 6.3.1) building cost 24 and perform fmincon using the interior-point method with 100000 max iterations where we provide an analytically computed derivative (see Section 6.7).

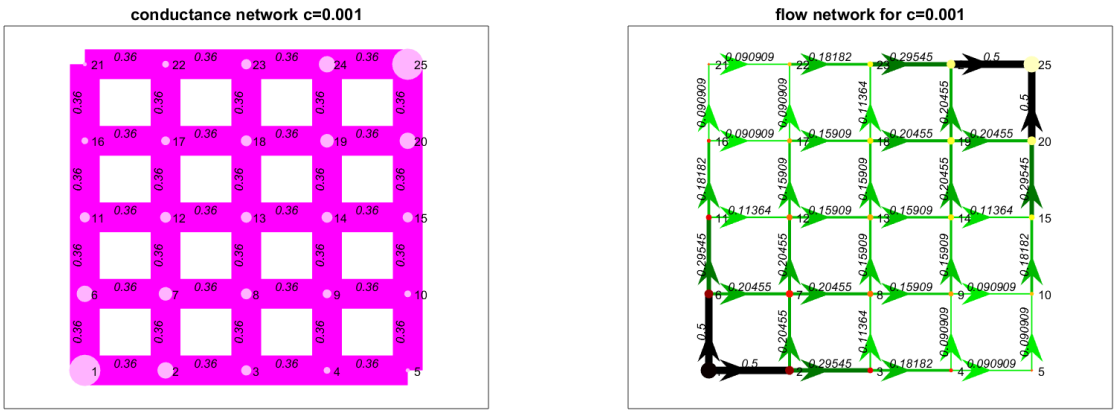


Figure 6.3.1: Left: Uniform conductances, single source single sink boundary flows of magnitude 1. Right: Corresponding physical flow on the square grid of side length 5.

We perform this optimization for mixing dissipation cost with $c = .001$, NME $+ .001D$: with such a small penalty on dissipation we are effectively only optimizing the NME. We include some weight on the dissipation because it is a very small contribution to the value of the function, but it is regularizes the conductances within the path so that they do not become vanishingly small, causing possible errors in the calculation of the cost function and its derivatives or causing the process to be numerically unstable.

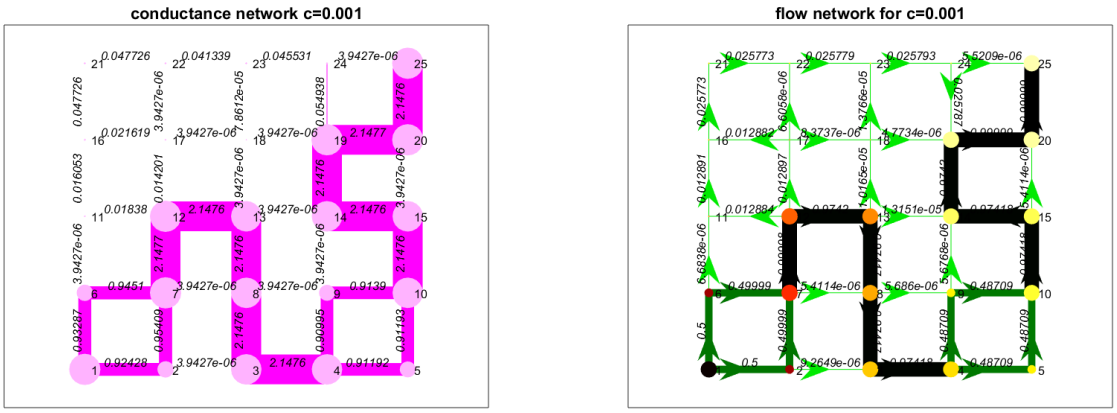


Figure 6.3.2: Conductances and resulting physical flow after fmincon with $c = .001$ and CMD $= -29.5252$.

The result is interesting in that it is not too far off from the actual NME minimizer, which we know is the maximum path going through all of the nodes. Recall that the actual NME minimizer over flows containing 25 nodes is $-\log(25!) = -58.0036$ (see Theorem 5.1. The estimate that our full fledged optimization obtains has a NME $+ .001D = -57.9711$, with NME $= -57.9953$ and is the maximal length path τ_{25} . The error in the NME is small, but is caused by edges which should have very small conductance and flow. Interestingly, there is a path with a loop in the flows on vertices $\{12, 11, 17, 16, 21, 22, 23, 24, 19\}$. A feature like this, loops in the middle of the network, is important for our algorithm to take into consideration, as it transitions to more optimal networks.

On the other hand, a continuous gradient based optimization techniques that accurately cfonverge to dissipation minimizers for 1 source and 1 sink is presented in [CR18]. For multiple sources and sinks, the optimum is a tree [BM07, Dur07] and the optimal tree is found in repeated applications of a continuous relaxation technique. In those cases search over trees needs to be paired with the relaxation technique [BM07]. We re-do this result by optimizing NME $+ 100D$, putting a high cost on dissipation so that effects from the mixing are practically ignored (shown in Fig 6.3.3). The interior point method finds the global optimum which is the shortest path or geodesic connecting nodes 1 and 25 with uniform conductances as proven in Chapter 5 Section and shown in [CR18]. The mixing dissipation cost of this network is 77.5714 while the dissipation of this network is 0.8985. The theoretical minimal dissipation of the geodesic is $\frac{8^{1+\frac{1}{\gamma}}}{(24)^{\frac{1}{\gamma}}} = \frac{8^3}{24^2} = 0.8889$. The slight inaccuracy is due to the non-zero edges not included in the geodesic which are not fully eliminated by the continuous optimization. These inaccuracies are tolerable for our purposes because they are much smaller than the changes in the cost function caused by topology changes.

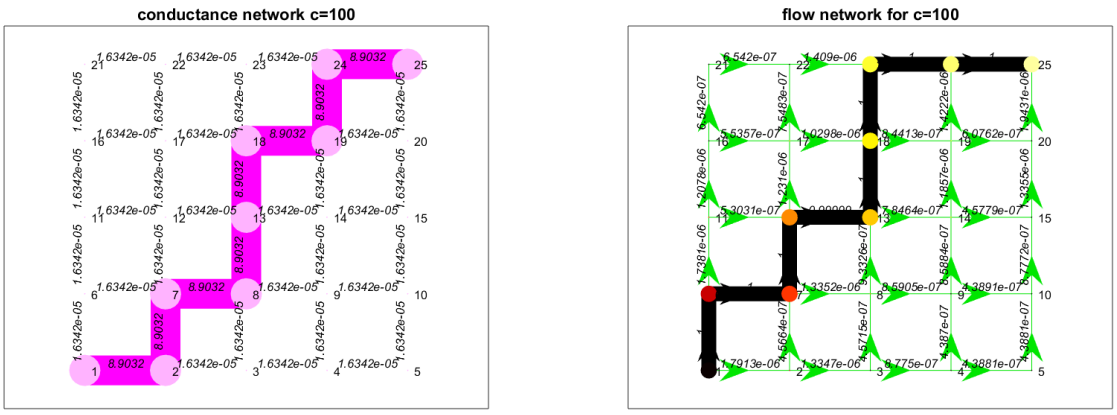


Figure 6.3.3: Conductances and resulting physical flow after fmincon with $c = 100$

Since fmincon alone gives inaccurate, although revealing results, it cannot alone be relied on to accurately estimate minima. We are interested in accurate results produced by as exhaustive of a search as possible. Different initializations of the continuous search algorithm will produce different results, so points in neighborhoods of an extensive collection of local optima need to be sampled. As we will see in Section 6.3.3, to efficiently traverse neighborhoods of different attractor networks we need to deliberately modify the network topologies, without altering the original flow network so much that our algorithm is unable to find local optima. For this reason, we rejected the algorithm of [KSM10] which we found would only find the globally optimal network if initialized in the same region of conductance space as it contains. Next we paint a picture of the geometric structure of the domain which we will use to search over flow topologies.

6.3.2 Flow network topology changes

Through it's definition of our objective function as a function of flows has non-differentiable points built into it. The flow from node i to j $\tilde{q}_{ij}$ is defined using the markov chain, where probability of transitioning frome one node to the toher is the ratio between the positive flow in that direction to the total flow at that node. The total flow of i can be written

$\sum_{j \in n(i), q_{ij} \geq 0} q_{ij}$. The non-differentiable points of this term are when some q_{ij} switches signs because the summand representing the directed edge i to j is constantly zero to the left of 0 and the identity to the right. Seeing that the NME is entirely in terms of total flows and $\tilde{q}_{ij}$, and $\tilde{q}_{ij}$ is the i, j entry in the sum of powers of the markov chain, represented as a finite truncation of the geometric series of matrices $\frac{1}{1-P}$.

Since we are optimizing the mixing-dissipation cost as a function of the conductances, we phrase this as a perturbation of the conductance along an edge causing the direction of flow to swap along an affected edge. To optimize the mixing-dissipation cost the optimum must be searched for in the region of each distinct topology. The reason for this is that along changes in topology there may be hills separating different local optima. This is depicted in the following example.

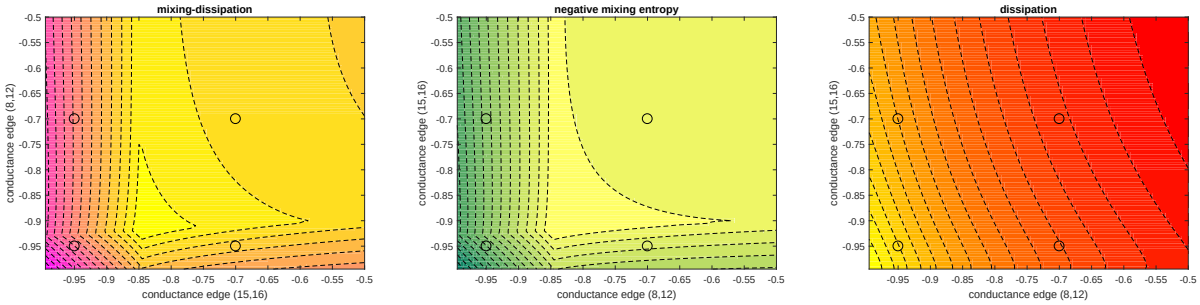


Figure 6.3.4: The contour plot of the mixing-dissipation cost (CMD), the total negative mixing entropy (NME) and the dissipation (D). Points in the regions representing four different flow topologies are shown.

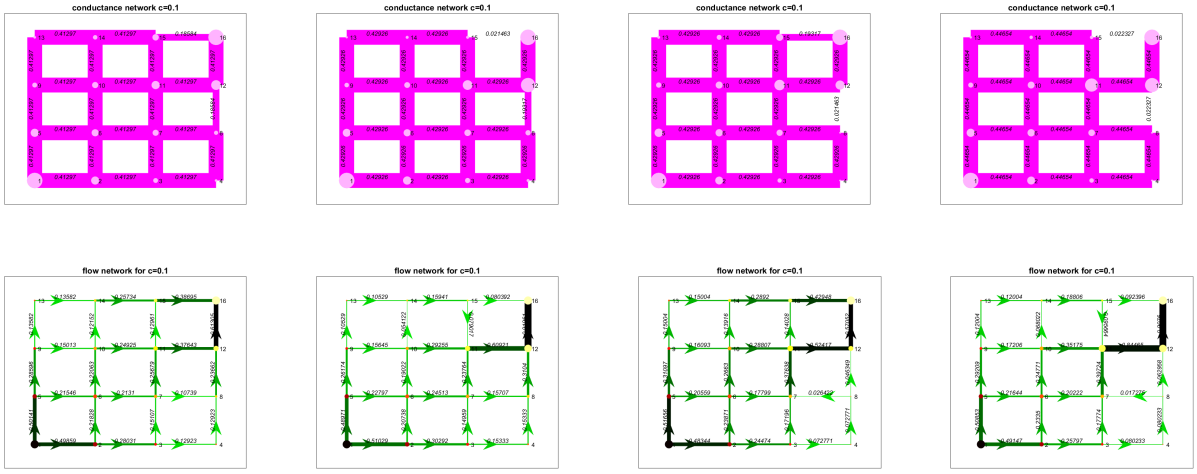


Figure 6.3.5: The conductances (top) and flows (bottom) corresponding to the top-right, left, bottom and bottom left points highlighted in the contour plots.

6.3.3 Short circuits

Due to dissipation preferring trees under the building constraint trees and path-networks are strong attractors for the continuous optimization methods. This is even true at relatively low values of c as in the next figure. Here $c = .2 < c_{25,23}$, which we know implies that if we minimize over only path graphs, then the result should have 25 nodes.

Since the continuous optimization results in sparse conductance networks we are inspired to add material in a fashion which maintains aspects of the flow topology while offering a collection of sub-networks to choose from. One way to do this would be to borrow the idea from [KSM10] to apply a spatial convolution to the conductance edges with a smoothing kernel, such a gaussian. That is, use the spatial coordinates of the grid, which we consider to be $i \mapsto \left(\left\lfloor \frac{i}{N} \right\rfloor + 1, i \mod N\right)$ where the residues of the modulo are taken to be $1, 2, \ldots, N$ and assign the edge a coordinate based on one of its vertices. For the demonstration in figure we use a convolution gaussian $K(a,b)\exp(((a-y)^2+(b-y)^2)/r)$ where $K(a,b) = \frac{1}{\sum_{(x,y)\in\mathcal{N}}\exp(((a-y)^2+(b-y)^2)/r)}$ normalizes the sum over the edges to be 1. Then we rescale the network to obey the building cost constraint. We call this process mollification. In

Figure 6.3.6 we have applied mollification to the recent network and then applied continuous optimization method. Note how the convolution short-circuits the network producting aa much shorter network.

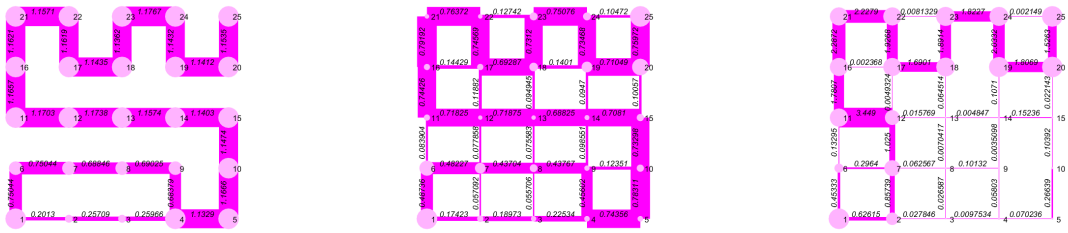


Figure 6.3.6: Recent network, followed by mollifying, followed by continuous optimization. Notice the preference for a short circuit.

Although this approach works for the moving sink and dissipation under removing edges, it is not an efficient method for minimizing the mixing-dissipation cost. After applying mollification, we have noticed that many steps were spent trying networks with a much shorter length due to short-circuiting. At the same time, the solution is not to just make the gaussian more tightly concentrated by making r very small. In that case too little extra building material is added and the network is too similar to the original network the continuous optimization deletes the newly added edges and reverts to the topology of the recent network.

Even worse is if we combine a flow swap to the mollification. We need flow swaps because it is a necessary part of thoroughly traversing the land-scape, as well as breaking out of local optima. But if we use mollification or a method of adding building material which does introduces short circuits then a flow swap might drastically change the topology because it might stop and reverse flows present in the recent network in favor of the short circuit.

Our method of optimization needs to add material in a way that offers new flow topologies while avoiding cutting the shortest low resistance path from the source to the sink down by too large of an amount. Also, it is important that the newly added edges are strong enough

so that the initial input into the continuous optimization does not immediately delete then, We aim to both add new material to expand the possible low resistance sub networks and preserve the progress of the routine by approximately keeping the network of flows from the recent network intact. We keep the idea of convoluting by a smoothing kernel as an inspiration while minding that nodes cannot be connected to downstream nodes.

6.4 How to deal with zero conductance edges

Edges in the ambient conductance network (square or triangular grid) with that are assigned vanishingly small conductances need special treatment. As it has been discussed in section 6.3.1 minimally dissipative networks with the building constraint with exponent $0 < \gamma < 1$ favor loop-less networks and the continuous part of our optimization often removes edges to minimize the number of loops (measured through Betti-number) in the network, making it very sparse. These observations can be useful to make our optimization as efficient as possible, but we proceed with care not to design an algorithm which samples only from a class of sparse networks. Although we want our algorithm to search the collection of paths and trees, which we know contains we know to be good candidates for optima, it must also be capable of finding an optimal network if it lies outside of this class.

The growth step also does not force the network to be treelike even though it does prevent large short circuits. For a path network, the growth step acts like a dilation of the set of edges present in the recent network. It adds material in a way that adds more non-zero flow paths between sources while preserving the topological qualities of the network. The extra attention to make sure that the material added is similar to the nearby material cost is to prevent added paths from being overwhelmingly strong while being good enough avenues of flow to be considered by the continuous optimization.

6.4.1 Considering edge conductances below a tolerance to be 0 (deleting edges)

All of the optima produced by our optimization algorithm assign non-trivial conductances only to a small subset of the edges present in the ambient grid. We therefore consider it a safe choice to disregard edges with very small conductances from computations in order to effectively reduce the dimension of the space that the continuous optimization is working on. We discuss this in detail with regards to the continuous subsection in Section 6.4.2, the network growing method Section 6.5 and we discuss method for finding roots which bring about flow direction switches in Section 6.6

6.4.2 Disregarding low-conductance edges in continuous optimization to reduce dimension of search

As we have discussed in Section 6.3.1 we pre-compose our function by an exponential to enforce positivity off the bat and we also put a lower bound on the logarithm of the positive conductances to be -11 which corresponds to the cutting off all conductances below $e^{-11} = 1.6702 \times 10^{-5}$ because in practice, if we do not then the conductances along connected sub-networks where every node has a degree of 2 (paths) take a long time to converge to a uniform set of conductances. In practice, without a lower bound, the path parts of the network are notably thinner towards the middle.

In our numerical experiments, we have noticed that the continuous optimization has a tendency to eliminate edges, not create edges. This is one of the inspirations for the growth step. Material needs to be added for the continuous optimization to sculpt an optimal network from. We are lead to write the mixing-dissipation cost as a function of edges with a conductance above a certain level because the interior point method has not been observed to revive vanished edges. Noting that for large enough c the optimal network is always the geodesic, and for small c the optimal network is the network going through as many points as possible. Then the size of the search space for an optimization can be reduced by a constant

factor in the case of the maximal path and by $O(N)$ in the case of a geodesic. Even if the final local optima obtained from continuous optimization is a dense graph, for non-zero c it has been seen that intermediate steps usually involve sparse networks ($O(N)$ edges).

Therefore, restricting the function to the high conductance edges reduces the dimension of the optimization for many of the iterations of the algorithm. This reduces the run-time of the interior point method because the size of the search space is reduced in dimension as well as the out-put dimension of the barrier function corresponding to preserving the flow dimensions.

We find the edges with conductance $< 10^{-4}$ deleting them and leaving all other edges. We apply this filter after the growth step and subsequent building cost rescaling but do not apply it again when we are performing flow swaps, because flows after flow swap may be very small and filtering out small conductances may swap the flows back to the recent topology. This way we can find a local optima in the section of space corresponding to a desired flow topology (as in Fig 6.3.4).

The inclusion of the dimension reduction has sped up the algorithm contributing to the feasibility of performing thousands of samples as well as considering very large ambient networks. For example using the intersection c's $c_{m+1,m}$ (or $c_{m+2,m}$ for square grids) we can pick the c which has a path of much shorter length than the ambient grid. For instance we can sample a path of length 25 that has 24 edges (but may be slightly more because of a not previously considered loopy part) in a 9×9 square grid with 144 edges. This effectively reduces the work on a 9×9 grid at this level of c to the amount of work on a 5×5 grid. We do not have a proof, but experimental results suggest that optimizing $\mathrm{NME} + cD$ for fixed value of c is $O(1)$ with respect to N the side length of the ambient grid.

6.5 The growth step

6.5.1 Square grid

Before we describe the algorithm we define what we mean by saying a node is in of the support of the network. A node is in the network if the sum of the weights of neighboring edges is above a certain threshold. In our implementation of the following algorithms, a node i is considered to be in the support if at least one of the conductances of edges adjacent to i is greater than or equal to 2×10^{-2}. An edge is in the support if it's conductance is 2×10^{-2}.

We add new conductance material to an existing network using what we call the *network growth step*. It is written $\text{Grow}_{\text{up}}(\kappa_{ij})$ for a conductance network κ_{ij}, where "up" in the subscript can be replaced with the "right", "down" and "left" to denote the direction of the growth. Here we describe in detail every step of the algorithm. Intuitively speaking, the network growth step can add material in four directions. That is, it can add conductances to edges with vertices above, to the right, below and to the left of nodes in the support. We choose to do this in a single direction at a time because doing so enables us to be sure that we are not inadvertently adding a short circuits to the network. For approximate optimal networks in the collection of results in this paper, we use two steps per iteration of our optimization routine. For both the square and the triangular grid.

We only describe how to perform the growth step in the upwards direction. Then network growth to the right, bottom, and left are all defined through this using di-hedral transformations of the spatial network. That is, for two involution dihedral symmetries termed R and L, we define

$$\begin{aligned} \text{Grow}_{\text{Right}}(\kappa_{ij}) &= L(\text{Grow}_{\text{up}}(L(\kappa_{ij}))) \\ \text{Grow}_{\text{Down}}(\kappa_{ij}) &= R(\text{Grow}_{\text{up}}(R(\kappa_{ij}))) \\ &\text{and} \\ \text{Grow}_{\text{Left}}(\kappa_{ij}) &= LR(\text{Grow}_{\text{up}}(RL(\kappa_{ij}))) \end{aligned}$$

where R is a rotation by π radians, and L is the reflection over the axis going through the bottom left and top right corners. These transformations are specifically chosen, as we will see, so that the diagonals in the triangular grid and to ensure that triangular and square grids are similarly handled in our algorithm.

Growth in the upwards direction for square grids

We start with a network of conductances and we consider the support and it's complement. This is depicted in figure 6.5.1a by nodes in the support having their edges drawn in black and nodes outside of the support having none of their adjacent edges drawn.

1. First we locate edges in the square grid for which their top nodes are outside of the support of the network and their bottom nodes are inside the support of the network.

2. We add positive conductance to these edges as is depicted in figure 6.5.1b. Suppose that i is in the support and $i+N$ is outside of the support. Specifically, the amount we add is the average conductance of edges adjacent to i above a positive threshold. In our numerical experiments we use 2×10^{-2}.That is we add

$$\frac{\sum_{j \in n(i):\kappa_{ij}>.02} \kappa_{ij}}{\#\{j \in n(i) : \kappa_{ij} > .02\}} \tag{6.5.1}$$

   to each new upwards edge. This is so the conductance of the new material has a similar magnitude to the neighboring material in order for the flow down new edges to be a similar size to the flow down old edges. Call the new network $\kappa_{\mathrm{up}ij}$

3. In the support of $\kappa_{\mathrm{up}ij}$ we locate the vertical edges and define the set of *top nodes* to be the nodes which are the tops of any of the vertical edges. The top nodes are circled in figure 6.5.1c.

4. If two top nodes are horizontally adjacent we add positive material if the horizontal edge directly shifted below is in the support of the network. The amount conductance

we add to an edge is the same as a the conductance of the horizontal edge directly below. Note: We do not consider if the edge between two top nodes is in the support and the edge below is in the support. Extra material will be added to this edge. This might be slightly less elegant (although slightly easier to code) and the authors do not believe this alters the optimization quality. The only case when this will occur is if a square with side length 1 is in the support.

5. Note the final network has a different total building cost than the original, but networks are always re-scaled to have the chosen building cost before reaching the continuous optimization routine. Note: As in figure 6.5.1d there may be edges which connect to the support at only one node. These are not eliminated because in our experience, the continuous optimization routine will send their conductance to below the threshold at which it will be filtered out, effectively reconsolidating the material into the rest of the network.

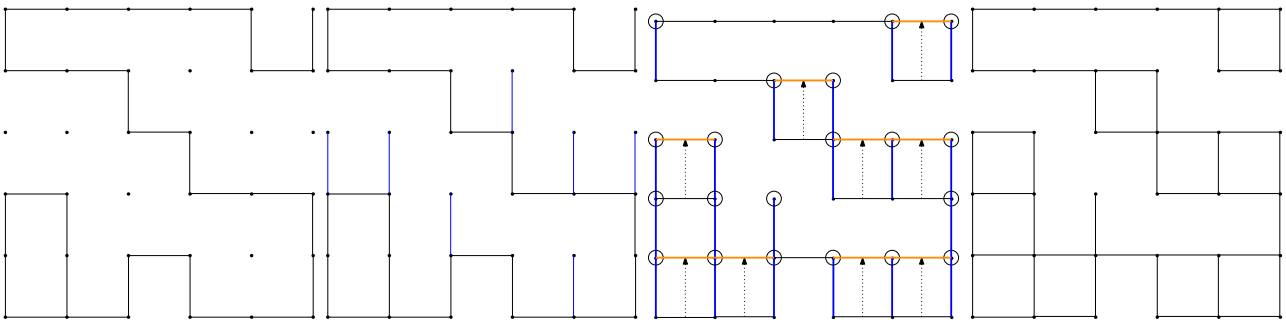


Figure 6.5.1: 6.5.1a. The initial network before growth. 6.5.1b. The network including added conductance along vertical edges from a node in the support of the network to a node outside of the support of the network directly above it. 6.5.1c. The network from step 2. with all vertical edges in the new support highlighed in blue. Nodes at the top node of vertical edges circled. The dotted lines with arrows stem from horizontal edges in the support of the network upwards to newly added edges connected top nodes of two vertical edges. 6.5.1d. The new network with growth in the vertical direction.

6.5.2 Triangular grid

Growth in the upwards direction

We take the definition of the support network as above. The triangular grid is more complex, but an analagous growth step is outlined for networks on the triangular grid. First we draw the triangular network so that its vertices are arranged the same way as the square ambient grid. (To do this we have to alter the lengths of edges within the network). We recycle the notation $\text{Grow}_{\text{up}}(\kappa_{ij})$ where "up" can be replaced by "right", "down" and "left". We use these four orthogonal directions for the triangular grid because they are enough to add material to any edge with vertically, horizontally and diagonally adjacent nodes which lie outside of the support. As in the square case we generate the other growth directions from the definition of the upwards growth by conjugating the results by L, R and LR . It is easy to see that these preserve the adjacency of the nodes by keeping the diagonals between the bottom right and top left nodes of each square. In our experiments we perform the triangle growth step twice, one growth coming directly after the first, in each iteration of the optimization routine.

1. First we locate vertical edges in the triangle grid for which the top node is outside of the support of the network and the bottom node is in the support of the network.

2. Add positive conductance to each of these edges depicted in blue in figure 6.5.2b we call this new network $\kappa_{ij\text{up}}$. The amount of conductance we add is again equal to the average conductance of the edges adjacent to the bottom which are in the support as in equation 6.5.1.

3. In $\kappa_{\text{up}ij}$ locate the nodes which are top nodes of edges in the support. Call this set the top nodes. These are circled in figure 6.5.2c.

4. For every top node i, add conductance to an edge between a horizontal node directly to the left if there is a diagonal edge in the support network such that that diagonal edge, the vertical edge down adjacent to i and the considered new horizontal edge form

a triangle. Let the added horizontal conductance be equal to the arithmetic mean of the vertical and diagonal nodes in this triangle. These are depicted in orange in figure 6.5.2c

5. Not considering the horizontal nodes added in step 4 and only the edges in the support of $\kappa_{\text{up}ij}$ we now add conductance to diagonal edges. For every top node i, if the node directly below it has an adjacent horizontal edge pointing right in the support add a diagonal edge from the top node to the node to the right of the node directly below i. The new diagonal edges are colored dark green in figure 6.5.2c. Add conductance to this edge with an amount equal to arithmetic mean of the conductances of the vertical and horizontal edges in the support directly down and left of this edge.

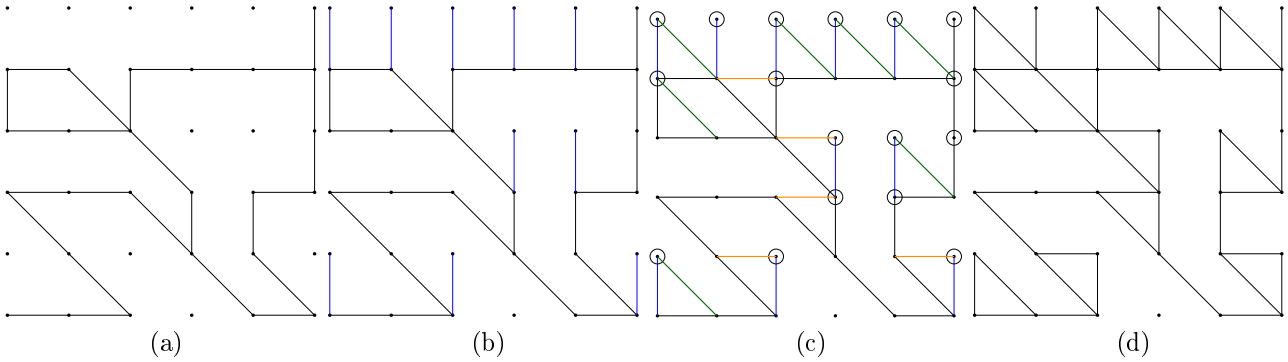
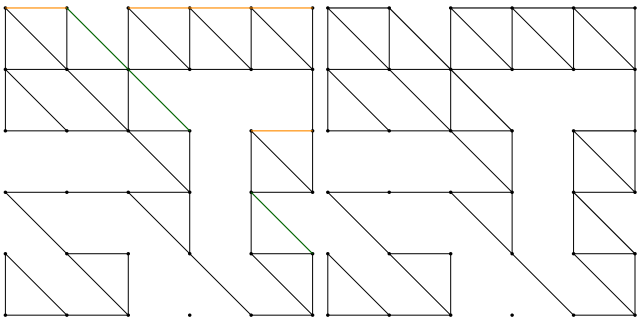


Figure 6.5.2: In figure 6.5.2a a recent triangular network input into the growth algorithm. Note that we have distorted edge lengths so that the vertices are the sasme as the ambient square grid. In figure 6.5.2b conductances are added to the vertical edges connected a node in the support to a node outside of the support of the network in blue. In figure 6.5.2c horizontal and down-right diagonal edges are added stemming from the top nodes such that edges in the support complete triangles with these added edges. The new horizontal edges are in orange and the new diagonal edges are in dark-green. Figure 6.5.2d is the new network obtained from one upwards growth step.



(a) (b)

Figure 6.5.3: Figure 6.5.3a is the network obtained from performing two upwards growth steps on the network in figure 6.5.2a in a row with new horizontal edges in orange and new diagonal edges in dark green. Figure 6.5.3b is the same network with all edges in the new support shown colored black.

6.5.3 Desirable properties of growing the network for optimization

6.5.3.1 Avoiding short-circuits

We only need to consider the upwards growth steps because the same properties hold for the other directions by swapping each mention of a direction with its rotated or reflected counterpart.

In the upwards growth step step 2, adding upwards edges does not connect any two nodes in the support by a new low resistance edge. The only step that does is step 4 when horizontal conductances are added. Still, if two nodes in the support are connected by a new horizontal edge, by the way we have chosen the nodes, they are already connected by a path with exactly 3 edges in the support. The new shorter path is not a very big change to the path length from source to sink. It is true that more distant nodes are connected with paths, as is the case in figure 6.5.1c with the nodes in the bottom right corner. Still, in the cases we have explored and encountered in running our optimization, the far away nodes are connected by edges with very similar conductances to the already existing path through a path nearly as long. Therefore, the new paths are not greatly preferred as lowering the dissipation of the

network, and they do not attract the continuous optimization to sub-optimal local optima.

The upwards growth step for triangle grids is similar in this regard. It only connects nodes in the support which already have a path with two support edges connecting them. The newly shortened paths are not short low resistance avenues for conductance connecting distant nodes in the network. Any new path connecting nodes is of a similar length and conductance as an existing path since we add new paths to adjacent support nodes and the conductances of new edges are similar to the conductances of neighboring support edges. Running the growth direction twice, in our experience only shortens paths between nodes by very small amounts.

6.5.3.2 Incremental growth, decrease in network length and addition of bubbles.

Incremental growth is achieved by using the growth step in tandem with the flow swapping step. The growth step, as we can see offers new longer paths which are adjacent to the existing paths.

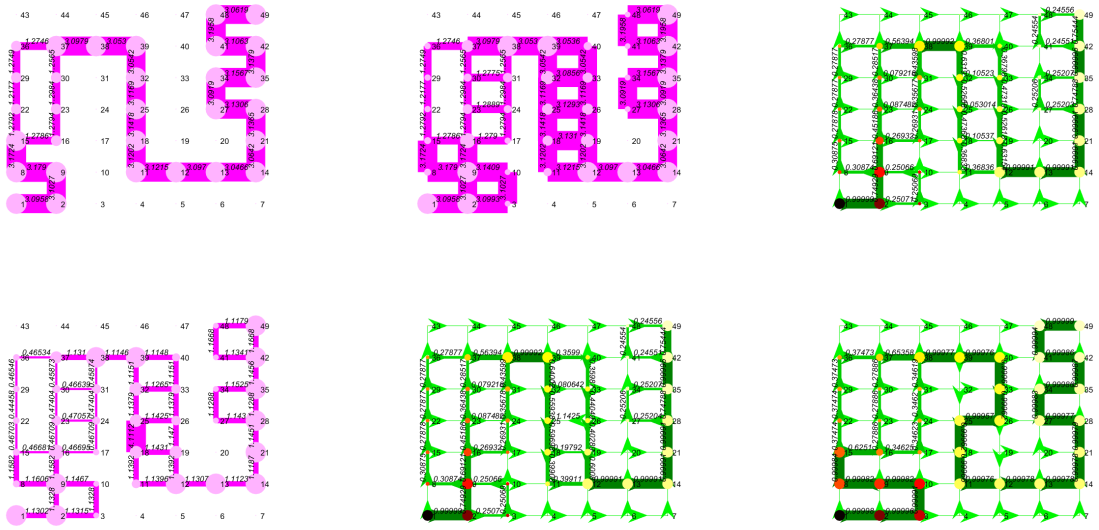


Figure 6.5.4: From left to right top to bottom. The initial netowk. A grow network step to the right the corresponding flows **note the direction of (25,26)** a conductance network with a perturbed edge to swap the flow (**18,25**). The corresponding flow network. Lastly, the flow network of the continuously optimized network. Note the new turn at **25,26** , adding length to the network. Also the grow network step added more loops to the network.

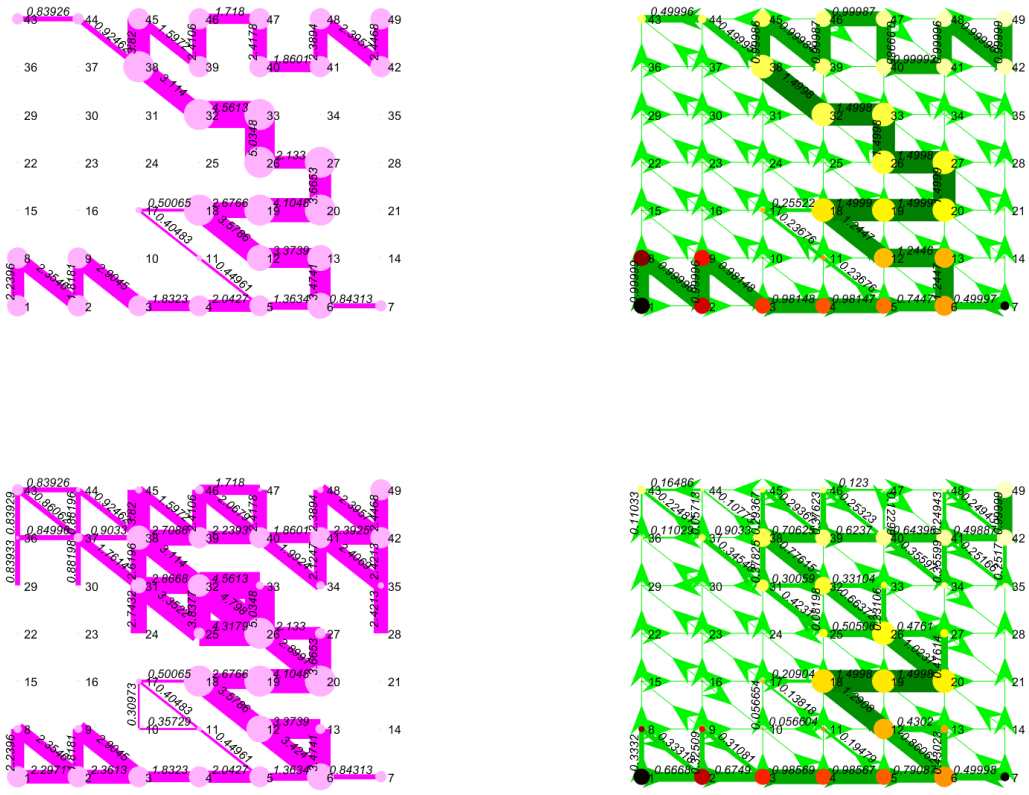


Figure 6.5.5: Two downwards triangular growth steps. The top two images are the conductances ad flows of the original networks. The bottom two images are the conductances and flows of two new networks. Note how the topologolical properties are preserved and there are no high conductance paths connecting nodes which are very far apart in the original network. Note how the conductances added are very similar to their neighboring conductances, a feature built into the algorithm.

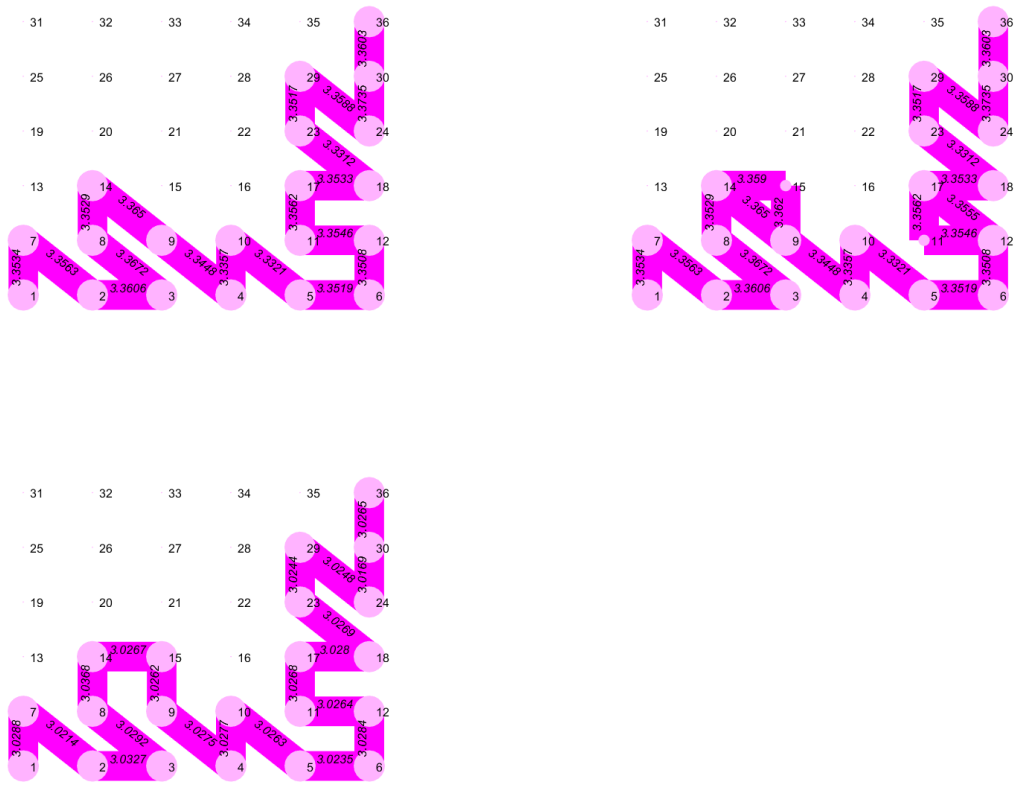


Figure 6.5.6: A path network on the triangular lattice. It undergoes a growth step in the upwards direction, followed by continuous optimization. There is no topology swapping this time. Often times a growth step alone is very productive in increasing the length of networks.

6.5.4 Overview

As we have seen in Section 6.3.1 the different possible flow topologies on the network carve the space of all possible conductances into distinct regions. Between these regions are hells on which the mixing-dissipation cost becomes high. At the boundary between the regions the cost is not differentiable. Continuous optimization algorithms that incrementally decrease the function value from traversing. Therefore, continuous optimization alone cannot be expected to explore conductances and physical flows corresponding to a variety of different

flow-direction graphs.

For reference it is good to look at the landscape of the cost function corresponding to four different flow directions given by perturbing two edges in Fig 6.3.4. Theses contour plots are a depiction of the cost function restricted a two dimensional subspace of the set of conductances. This suggests that different flow direction networks are separated by intersecting manifolds, most likely of co-dimension one (if the gradient of the flow along each edge as a funcction of conductances is non-zero). To be fully certain that every flow direction network was considered, we would have to define these boundary sub-manifolds and curves crossing through them. This is a difficult, possibly intractable, problem, but the graph also suggests that changing one flow direction at a time is an effective way to locally explore the topologies similar to one under consideration.

6.5.4.1 Choosing the flow direction network to test for minimum

For this reason we find a closed formula for the perturbations which will arise in exactly one flow change. In our flow direction changing routine (fully described in section 6.6) we first randomly chose, up to a predefined limit a number (2 in the applications presented in this paper) of flow reversals. Then we define a new network by randomly choosing one of the possible perturbations of conductance along a single edge which swaps the flow in exactly one other edge and adding it to the network. We keep track of both the edges whose flows we reversed and the edges we perturbed. We repeat the process by randomly choosing a pair of possible edge to swap flows (affected edge) and conductance (causal edge) perturbation giving rise to the flow change, excluding all other previously flow direction changed edges and perturbed edges from consideration. We stop once the set limit of flow-reversals is met or the network runs out of flows reversible by single edge perturbation. We then take this network as the starting point of the continuous optimization routine, applying the added non-linear constraints that the direction of the flows must be the same as that of the starting point. We do this by computing the pressure differences between all pairs of adjacent nodes in the

original network and recording this list of signs, and enforcing that the pressure differences between adjacent nodes is > 0 if the pair has a positive value in the list and < 0 if the pair has a negative value in the list.

We compare this to the recent local optimum, and choose the new local optimum if it has a lower mixing-dissipation cost. At the beginning of the next step we initialize the set of perturbed conductances and flow change pairs to be empty, randomly pick a new number of flows to reverse and repeat the process.

Note that we remove both the causal edge and its affected edge from consideration even though perturbations in a single edge may lead to flow reversals in multiple edges. Our goal is only to find topologies close to the current one, so we search for the smallest positive and negative perturbations at which a flow reversal occurs. We don't want to recover our starting topology by reversing flow in the same edge twice. In practice, sincewe consider only a small set of flow reversals, and we want to spread the reversals out over the network as much as possible. Also, it is desirable to have direct control how much the network is altered in each step. The opportunity to perturb one edge twice to swap a flow is not completely missed because it might be an available flow reversal in the next round of the optimization routine.

6.6 Formula for finding perturbations swapping flows.

As in the ambient networks we study in our computational experiments, the square and triangular grids with side length N, let the nodes be referred to by an indexing $1, 2, \ldots N^2$. (This section is applicable to any type of network, it is for convenience that we use the same definitions throughout the computational part of the paper.) Throughout this section, we assume that the network of conductances has one connected component in the sense that removal of the edges with zero conductance do not produce a disconnected graph. Our method can be modified for situations (such as the two source-two sink configurations shown in Fig. 4.4.2) where the graph has more than one connected component. Then we know

that the Laplacian of the conductance matrix defined in Chapter 2 Section 3 $\Delta_{\kappa_{ij}}$ has rank N^2-1. We form a new matrix which is a version of $\Delta_{\kappa_{ij}}$that is invertible on the space of all sets of pressures whose restriction to the sub-space $\{0\} \times \mathbb{R}^{N^2-1}$ (that is $p_1 = 0$ and p_i are free $i \neq 1$) is equivalent to $\Delta_{\kappa_{ij}}$ which we refer to as the invertible Laplacian and denote it by $\tilde{\Delta}_{\kappa_{ij}}$. If d_{ij} is the i,j entry of $\Delta_{\kappa_{ij}}$ then define $\tilde{\Delta}_{\kappa_{ij}}$ to be the $N^2 \times N^2$ real matrix with entries $\tilde{d_{ij}}$ such that

$$\tilde{d_{ij}} \quad = \quad \begin{cases} d_{11}+1 & i=1, j=1 \\ d_{ij} & \text{otherwise} \end{cases}.$$

To see that $\tilde{\Delta}_{\kappa_{ij}}$ is invertible let p be an N^2 dimensional vector (list of pressures) such that $\tilde{\Delta}_{\kappa_{ij}} p = 0$. Then applying the original Laplacian to p we find that $\Delta_{\kappa_{ij}} p$ is the vector with $-p_1$ in the first entry, and all other entries zero. For all columns of $\Delta_{\kappa_{ij}}$, summing the entires is zero, implying that the sum of the entries in $\Delta_{\kappa_{ij}} p$ must be 0 therefore $-p_1 = 0$ and along with all of the other entries, showing that the kernel is 0. To see that its restriction to the subspace $\{p_1 = 0\}$ is equivalent to the Laplacian restricted to the same subspace let p be such that $p_1 = 0$. Then $\tilde{\Delta}_{\kappa_{ij}} p$ is the sum of columns 2 through N^2 of the $\tilde{\Delta}_{\kappa_{ij}}$ multiplied by p_2 through p_{N^2} respectively. Since these columns agree with columns 2 through N^2 of $\Delta_{\kappa_{ij}}$, we have equality $\tilde{\Delta}_{\kappa_{ij}} p = \Delta_{\kappa_{ij}} p$.

The initialization of the optimization routine is that each edge is a random variable distributed uniformly i.i.d and then subsequently re-scaled to obey the building cost, and the network is connected with probability very close to 1. However, as the network evolves under the optimization algorithm, the algorithm will cause κ_{ij} to be approximately disconnected at some step. That is, removal of edges with very low conductance, for example $\kappa_{ij} < 10^{-4}$, disconnects the graph. In our numerical experiments, even though many conductances are brought to be close to 0, they never become exactly 0, and the Laplacian $\tilde{\Delta}_{\kappa_{ij}}$ is invertible and the way we compute the inverse is numerically stable enough for our purposes. That

is $\max \left|\text{inv}(\tilde{\Delta}_{\kappa_{ij}})\tilde{\Delta}_{\kappa_{ij}} - I_{N^2}\right|$ is often less than 10^{-6} and has not been observed to be bigger than 10^{-5}. Therefore, since we only use the inverse of the Laplacian to compute the flow changes (we use the Matlab linear solver in the computation of the objective function and gradients for the interior point method, and this computes the Morse Penrose inverse when the Laplacian becomes over-determined, effectively assigning pressure p=0 to any nodes that are disconnected from the source, which is an unproblematic way to handle disconnection of the graph) it is suitable for our optimization even when many edges are deleted. There are several ways to fix this in a mathematically rigorous sense, such as writing the code to be more modular to exclude nodes and edges from the node and edge list, effectively deleting rows and columns from the adjacency matrix and Laplacian at each step if the total weight is small. Our algorithm accurately finds flow reversal with a small number of missteps such as when a flow is weak, it sends it closer to zero, instead of changing its sign, so we have elected not to make these extra improvements.

6.6.1 Closed formula for flow reversals

We use the following formula for the inverse of the sum of a matrix and the outer product of two vectors introduced by Jack Sherman and Winifred J. Morrison [SM50] in our calculation.

Theorem 6.1. *(Sherman-Morrison) Suppose that A is an invertible $n \times n$ matrix and u and v are two n dimensional vectors which are presented as single column matrices. Then $1 + v^T A^{-1} u$ is non-zero if and only if $A + uv^T$ is invertible and when this holds,*

$$\left(A + uv^T\right)^{-1} = A^{-1} - \frac{A^{-1}uv^T A^{-1}}{1 + v^T A^{-1} u}. \tag{6.6.1}$$

For $1 \leq i \leq N^2$ let e_i represent the i^{th} standard basis vector, that is e_i has 1 for its i^{th} entry and 0 for all other entries. Let κ_{ij} be a network of conductances. Let t represent a perturbation to the network of conductances along the edge connecting nodes a, b. Then if we also refer to the adjacency matrix for the conductance network as κ_{ij}, then this is

equivalent to changing entry κ_{ab} to $\kappa_{ab} + t$. We refer to the conductance network and adjacency matrix obtained from κ_{ij} from adding t to the conductance on edge (a, b) as the perturbed conductances and denote it

$$\tilde{\kappa}_{ij} = \begin{cases} \kappa_{ab} + t & i = a, j = b \\ \kappa_{ij} & \text{otherwise} \end{cases}.$$

The Laplacian for the perturbed conductance network is the same as the Laplacian for the original network except that we add $-t$ to entries a, b and b, a and we add $+t$ to the diagonal entries a, a and b, b. The Laplacian for $\tilde{\kappa}_{ij}$ is equal to

$$\tilde{\Delta}_{\tilde{\kappa}_{ij}} = \tilde{\Delta}_{\kappa_{ij}} + t(e_a - e_b)(e_a - e_b)^T.$$

Then by equation 6.6.1 we have

$$\begin{aligned} \tilde{\Delta}^{-1}_{\tilde{\kappa}_{ij}} &= \left(\tilde{\Delta}_{\kappa_{ij}} + t(e_a - e_b)(e_a - e_b)^T\right)^{-1} \\ &= \tilde{\Delta}^{-1}_{\kappa_{ij}} - \frac{\tilde{\Delta}^{-1}_{\kappa_{ij}} t(e_a - e_b)(e_a - e_b)^T \tilde{\Delta}^{-1}_{\kappa_{ij}}}{1 + (e_a - e_b)^T \tilde{\Delta}^{-1}_{\kappa_{ij}} t(e_a - e_b)}. \end{aligned}$$

Let u, v be another edge in the network. We wish to find a perturbation of a, b such that the flow along u, v is reversed. This can only happen when $p_u - p_v = 0$. We solve for the root and see that the pressure drop from u to v is of the form $r + q(t - s)^{-1}$ so the sign of the pressure drop swaps from one side of the root to the other. Let Q be the vector of boundary flows with Q_i the boundary flow at i. We have the pressures of the perturbed conductance network with boudnary flow Q

$$\tilde{p} = \left(\tilde{\Delta}^{-1}_{\kappa_{ij}} - \frac{\tilde{\Delta}^{-1}_{\kappa_{ij}} t(e_a - e_b)(e_a - e_b)^T \tilde{\Delta}^{-1}_{\kappa_{ij}}}{1 + (e_a - e_b)^T \tilde{\Delta}^{-1}_{\kappa_{ij}} t(e_a - e_b)}\right) Q.$$

Let R_i be the i^{th} row of $\tilde{\Delta}^{-1}_{\kappa_{ij}}$, which, since $\tilde{\Delta}^{-1}_{\kappa_{ij}}$ is symmetric, is the transpose of the i^{th}

column of $\tilde{\Delta}_{\kappa_{ij}}^{-1}$ and d_{ij} be the i,j entry of $\tilde{\Delta}_{\kappa_{ij}}$. Then the pressure at node u in the perturbed conductance from with the boundar flows Q is

$$\tilde{p}_u = R_u Q - \frac{t\left(d_{au} - d_{bu}\right)\left(R_a - R_b\right)Q}{1 + t\left(d_{aa} - d_{ab} - d_{ba} + d_{bb}\right)},$$

and so

$$\tilde{p}_u - \tilde{p}_v = \left(R_u - R_v\right)Q - \frac{t\left(d_{au} - d_{av} - d_{bu} + d_{bv}\right)\left(R_a - R_b\right)Q}{1 + t\left(d_{aa} - d_{ab} - d_{ba} + d_{bb}\right)}.$$

Therefore the pressure drop is a monotonic function of t, so the the zero of this equation is where the pressure reverses. Setting the left side to 0 we get the perurbation of a, b t_{abuv}

$$\begin{aligned} t_{abuv} &= \frac{\left(R_u - R_v\right)Q}{\left(\left(d_{au} - d_{av} - d_{bu} + d_{bv}\right)\left(R_a - R_b\right)Q - \left(d_{aa} - d_{ab} - d_{ba} + d_{bb}\right)\left(R_u - R_v\right)Q\right)} \\ &= \frac{p_u - p_v}{\left(d_{au} - d_{av} - d_{bu} + d_{bv}\right)\left(p_a - p_b\right) - \left(d_{aa} - d_{ab} - d_{ba} + d_{bb}\right)\left(p_u - p_v\right)}. \end{aligned}$$

Using this formula we store the list of pairs of causal edges (a, b) and affected edges (u, v) and their roots t_{abuv}. This can be done using for-loops but we take advantage of singleton expansion in Matlab to store these roots as a $N^2 \times N^2 \times N^2 \times N^2$ array where the first two dimensions correspond to nodes on a causal edge and the last two dimensions are the nodes on the affected edges. After that we filter out perturbations which violate the positivity constraint on κ_{ij} that is, we only keep t_{abuv} such that $t_{abuv} + \kappa_{ij} > 0$. For our application we add some slack to avoid making the conductance along a cut set too small by removing roots from consideration with $5 \times 10^{-3} - \kappa_{ij} > t_{abuv}$: the main function of this filter is to prevent the algorithm from getting stuck engineering and then re-engineering flow reversals on edges that already have very low conductances. We also disregard roots that change flows along edges which have too low conductance because we do not want to change the flow along edges which are considered deleted.

6.7 Continuous Optimization Step

6.7.1 Enforcing positivity and the building cost constraint

For each fixed set of flow directions, we wish to find the conductances and corresponding magnitudes of flows which are optimal in terms of cost of mixing and dissipation. We do not have sufficient hypotheses to guarentee a unique minimum for a given direction of flows. Therefore the continuous optimization routine is at best able to ensure a local a local optimum.

Let κ_0 be our initial set of conductances represented as an adjacency matrix. Let $\tilde{\kappa}_0$ be the adjacency matrix of the logarithm applied to each conductance–the log conductances. Let $\tilde{\delta}$ be a perturbation to the log-conductances. Then the resulting additive perturbation on the conductances is $\delta = \exp(\tilde{\kappa}_0 + \tilde{\delta}) - \exp(\kappa_0)$.

Given a set of conductances, optimizing the mixing-dissipation cost under the constraints is equivalent to optimizing the log-perturbation. The log-perturbation is chosen because

$$e : \tilde{\delta} \mapsto \exp(\tilde{\kappa}_0 + \tilde{\delta})$$

maps the set of all real adjacency matrices on the underyling grid to the set of all positive adjacency matrices. We choose this -recomposition because even the interior point algorithm may go outside of the feasible region, and negative conductances are absolutely prohibited and may even lead to some of our coded functions not being able to be evaluated.

To enforce the building constraint, after a perturbation of the log-conductance we re-scale the perturbed conductance (written as an additive perturbation of conductance) via

$$r : \kappa_0 + \delta \quad \mapsto \quad \left(\frac{C}{\sum_{(i,j)} (\kappa_{0ij} + \delta_{ij})^\gamma} \right)^{\frac{1}{\gamma}} (\kappa_0 + \delta).$$

If κ_0 is the network after the growth process and flow swaps then to perform the

continuous optimization we first take the adjacency matrix formed by $\log(\kappa_{0ij})$ for all $\kappa_{0ij} > 0$ then we run the interior point method on

$$\text{NME} \circ r \circ e(\tilde{\delta}) + cD \circ r \circ e(\tilde{\delta}).$$

We maintain the directions of flows by first computing the signs of q_{ij} at the initial point (perturbation of 0, $\delta = \tilde{\delta} = 0$ at conductances κ_0) for each ordered pair i, j of neighboring nodes. We write this $q_{ij}(\kappa_0)$. Recall $p_i(e(\tilde{\delta}))$ denotes the pressure at each node i computed for the conductance matrix $e(\tilde{\delta})$. Then we maintain the non-linear constraints using barrier functions in the interior point method maintaining for all neighbors i and j

$$\text{sgn}(q_{ij}(\kappa_0))(p_i(e(\tilde{\delta})) - p_j(e(\tilde{\delta}))) \quad \geq \quad 0.$$

In the language of optimization we write this

$$\begin{aligned} \text{minimize} \quad &: \quad \text{NME} \circ r \circ e(\tilde{\delta}) + cD \circ r \circ e(\tilde{\delta}) \\ \text{subject to} \quad &: \quad \text{sgn}(q_{ij}(\kappa_0))(p_i(e(\tilde{\delta})) - p_j(e(\tilde{\delta}))) \geq 0 \text{ for all } (i,j) \in \mathcal{E} \\ &\quad\quad \tilde{\kappa} + \tilde{\delta} > -11 \end{aligned}$$

the second constraint is so that too much effort is not expended in shrinking edges. We note that applying this constraint leads to close to uniform conductances on path parts of the network.

We are able to speed up the run-time of this algorithm via including a closed form computation of the derivatives of the objective function $\text{NME} \circ r \circ e(\tilde{\delta}) + cD \circ r \circ e(\tilde{\delta})$.

6.7.2 Computing the derivatives

In this section we give the details of how we compute $\frac{\partial}{\partial \kappa_{ij}}\text{NME}(\kappa_{ij})$, $\frac{\partial}{\partial \kappa_{ij}}D(\kappa_{ij})$,$\frac{\partial}{\partial \tilde{\delta}_{ij}}r$, $\frac{\partial}{\partial \tilde{\delta}_{ij}}s$ and via the chain rule, the derivatives of the composition $\frac{\partial}{\partial \tilde{\delta}}\text{NME} \circ r \circ e + cD \circ r \circ e$. The total derivative is then the matrix product of the derivatives of $\frac{\partial r}{\partial \kappa_{ij}}$ and $\frac{\partial s}{\partial \tilde{\delta}_{ij}}$. Taking on the most difficult and interesting derivative first we compute the derivative of the NME with respect to the conductances κ_{ij}.

6.7.3 Computing the derivative of NME with respect to κ_{ij}

Before we compute the derivative of the NME we note that we write κ_{ij} as $|\mathcal{E}|$ dimensional vector, instead of as a matrix for our contributions. Thus $\kappa_{ij} \in \mathbb{R}^{|\mathcal{E}|}_{\geq 0}$. Recall the definition of the negative mixing entropy, NME is

$$\text{NME}(\kappa_{ij}) \quad = \quad \sum_i f_i \sum_j \frac{\tilde{q}_{ji}}{N_i} \log\left(\frac{\tilde{q}_{ji}}{N_i}\right).$$

Even though κ_{ij} does not explicitly appear anywhere in this expression, the negative mixing entropy for a set of boundary flows can be seen as a function of κ_{ij} because together with boundary flows, the κ_{ij} conductance network results in a unique set of physical flows. In total, we can write the composition of function that yields the negative mixing entropy graphically as

$$\kappa_{ij} \longmapsto p_i \longmapsto f_i \Longmapsto T_{ij} \longmapsto P_{ij} \Longmapsto \tilde{q}_{ij} \longmapsto N_j \Longmapsto \text{NME}.$$

Where a single arrow $\longmapsto$ represents a function of the immediately preceding variable and $\Longmapsto$ represents a function of more than one of the variables to the left. We recapitulate each of these definitions and mappings in 7 function compositions listed here.

1. p_i are the pressures at each node i and can be found by solving $Q_i - \sum_{j \in n(i)} \kappa_{ij}(p_i - p_j)$

for all nodes i simultaneously. This is the same as finding a solution to matrix vector product $\Delta_\kappa \boldsymbol{p} = \boldsymbol{Q}$.

2. Then $q_{ij} = \kappa_{ij}(p_i - p_j)$ is a function of κ_{ij} and p_{ij} and so is $f_i = \sum_{j \in n(i)} q_{ij} \mathbf{1}_{q_{ij}>0} + |Q_i| \, \mathbf{1}_{Q_i<0}$.

3. Then the Markov chain is formed via $T_{ij} = \frac{\mathbf{1}_{q_{ij}>0} q_{ij}}{\sum_j \mathbf{1}_{q_{ij}>0} q_{ij}}$ is a function of the flows.

4. The hitting probabaility matrix is a function of T_{ij} via $\boldsymbol{P} = \frac{1}{1-\boldsymbol{T}}$.

5. The flow from i to j $\tilde{q}_{ij}$ can be viewed as a function only of P_{ij} via considering $f_i = \sum_{a \text{ source}} P_{ai}$ but we consider f_i as its own variable, so $\tilde{q}_{ij}$ is a function of T_{ij} and f_i computed $\tilde{q}_{ij} = f_i T_{ij}$.

6. Recall N_i is the normalizing constant of the distribution of species arriving at node i which is the sum $N_i = \sum_{j:\tilde{q}_{ji}>0} \tilde{q}_{ji}$, so it is a function just of the flows from one node to another.

7. Lastly the NME, as written immediately above is a function of $f_i, \tilde{q}_{ij}$ and N_i.

This list of compositions makes it tempting to compute the derivative using the chain rule. Without any cleverness applied besides the concept of products of multi-dimensional arrays, this is computationally inefficient theoretically in terms of space and in practice in terms of time. This is mostly we have matrix valued functions of matrices, and the most natural way to store their total derivatives is as a 4 dimensional array. We actually have successfully carried this out using the tensor package from Sandia National labs for Matlab, but the computation was too slow for our application due to the spatial complexity of the process– even with using their sparse arrays. Therefore we borrow methods of back-propagation, as applied to similar biologically inspired fluid flow problems in [CR18] and tailor the algorithm to the complicated function we wish to optimize.

The basic idea is to consider the intermediate variables as free variables and apply their definitions as the images of functions via Lagrange multipliers. Then we compute the values

of the intermediate variables, take the derivatives with respect to the free variables, finally we set the constraint functions equal to zero and solve for the Lagrange multipliers. Then the derivative with respect to κ_{ij} is the derivative of the cost function extended to include the solved multipliers. Solving the Lagrange multipliers ends up being a sequential process where we first solve for the one "closest" to the objective function and then proceed backwards based on their dependencies.

We have the lagrangian (where $\mathcal{N}$ is the set of all nodes)

$$\begin{aligned}
\Theta &= \sum_{i\in\mathcal{N}} f_i \sum_{j:\tilde{q}_{ji}>0} \frac{\tilde{q}_{ji}}{N_i} \log\left(\frac{\tilde{q}_{ji}}{N_i}\right) - \sum_{i\in\mathcal{N}} \alpha_i \left(N_i - \sum_{j:\tilde{q}_{ji}>0} \tilde{q}_{ji}\right) - \sum_{i,j\in\mathcal{N}} \gamma_{ij}\left(\tilde{q}_{ij} - f_i P_{ij}\right) \\
&- \sum_{ij\in\mathcal{N}} \mu_{ij}\left(\delta_{ij} - \left(P_{ij} - \sum_{l\in\mathcal{N}} T_{il}P_{lj}\right)\right) - \sum_i \sum_{j\in n(i)} \lambda_{ij}\left(T_{ij} - \frac{q_{ij}\mathbf{1}_{q_{ij}>0}}{f_i}\right) \\
&- \sum_{i\in\mathcal{N}} \beta_i \left(f_i - \sum_{j\in n(i)} q_{ij}\mathbf{1}_{q_{ij}>0} + |Q_i|\,\mathbf{1}_{Q_i<0}\right) - \sum_i \nu_i \left(Q_i - \sum_{j\in n(i)} \kappa_{ij}(p_i - p_j)\right).
\end{aligned}$$

After formulating this function, first compute the values of the intermediate variables, $p_i, q_{ij}, f_i, T_{ij}, P_{ij}, \tilde{q}_{ij}$ and N_i. Then take the derivative of Θ with respect to N_i to obtain an expression for α_i in terms of $N_i, f_i, \tilde{q}_{ji}$. Then working through the following sequence in order top to bottom, observe that we can solve for the Lagrange multipliers on the right using the Lagrange multipliers on the left and the equation setting a derivative of Θ equal to 0

$$\begin{aligned}
\alpha &\xrightarrow{\frac{\partial\Theta}{\partial\tilde{q}_{ab}}=0} \gamma \\
\gamma &\xrightarrow{\frac{\partial\Theta}{\partial P_{ab}}=0} \mu \\
\mu &\xrightarrow{\frac{\partial\Theta}{\partial T_{ab}}=0} \lambda \\
(\lambda,\gamma) &\xrightarrow{\frac{\partial\Theta}{\partial f_a}=0} \beta \\
(\lambda,\gamma,\beta) &\xrightarrow{\frac{\partial\Theta}{\partial p_a}=0} \nu.
\end{aligned}$$

We proceed to do this. For variables with one index let the bold version represent the column vector for the variable, (e.g. $\boldsymbol{p}$ for p_i) and for variables with two indices let the bold version represent the matrix indexed by those variables, (e.g. $\boldsymbol{\mu}$ for μ_{ij}).

First lets solve for α by isolating ther terms dependent on N_a they are.

$$f_a \sum_{j:\tilde{q}_{ja}>0} \frac{\tilde{q}_{ja}}{N_a} \log \frac{\tilde{q}_{ja}}{N_a} - \alpha_a N_a.$$

The derivative with respect to N_a is

$$\frac{\partial \Theta}{\partial N_a} = \sum_i \left(-\frac{\tilde{q}_{ia}}{N_a^2} \log \left(\frac{\tilde{q}_{ia}}{N_a} \right) - \frac{\tilde{q}_{ia}}{N_a^2} \right) - \alpha_a = 0.$$

Therefore solving for α we only need to multiply the flow matrix $\tilde{\boldsymbol{q}}$ by the diagonal matrices $\text{diag}(\boldsymbol{N^{-2}})$ or $\text{diag}(\boldsymbol{N^{-1}})$, take the logarithm applied to positive entries and sum down the columns.

We then solve for γ. First we extract all of the terms being summed that depend on $\tilde{q}_{ab}$. We get

$$f_b \frac{\tilde{q}_{ab}}{N_b} \log \left(\frac{\tilde{q}_{ab}}{N_b} \right) - \alpha_b \left(N_b - \tilde{q}_{ab} \right) - \gamma_{ab} \tilde{q}_{ab}.$$

Then the derivative of Θ with respect to $\tilde{q}_{ab}$ is

$$\frac{\partial \Theta}{\partial \tilde{q}_{ab}} = \frac{f_b}{N_b} \log \left(\frac{\tilde{q}_{ab}}{N_b} \right) + \frac{f_b}{N_b} + \alpha_b - \gamma_{ab} = 0$$

Thus

$$\gamma_{ab} = \frac{f_b}{N_b} \log \left(\frac{\tilde{q}_{ab}}{N_b} \right) + \frac{f_b}{N_b} + \alpha_b.$$

Now we solve for Lagrange multiplier μ_{ab}. The terms which depend on P_{ab} are

$$-\sum_{ij\in\mathcal{N}}\mu_{ij}\left(\delta_{ij}-\left(P_{ij}-\sum_{l\in\mathcal{N}}T_{il}P_{lj}\right)\right)-\sum_{i,j\in\mathcal{N}}\gamma_{ij}\left(\tilde{q}_{ij}-f_iP_{ij}\right)$$

$$\mu_{ab}P_{ab}-\sum_i\mu_{ib}T_{ia}P_{ab}+\gamma_{ab}f_aP_{ab}$$

Then the derivative with respect to P_{ab} is

$$\frac{\partial\Theta}{\partial P_{ab}} \;=\; \mu_{ab}-\sum_i\mu_{ib}T_{ia}+\gamma_{ab}f_a=0$$

and solving for γ_{ab} we get

$$\gamma_{ab} \;=\; \left(\sum_i\mu_{ib}T_{ia}-\mu_{ab}\right)/f_a$$

Then in terms of matrix multiplication, $\boldsymbol{\gamma}$ can be expressed as where $\text{diag}(\boldsymbol{f}^{-1})$ is the diagonal matrix with $\frac{1}{f_a}$ on the diagonal for $a\in\mathcal{N}$

$$\boldsymbol{\gamma}=\text{diag}(\boldsymbol{f}^{-1})\times(\boldsymbol{\mu}-\boldsymbol{T}'\boldsymbol{\mu}).$$

Now we solve for λ_{ab}. For this we need the terms depending on T_{ab} where $(a,b)\in\mathcal{E}$. These are

$$-\sum_{j\in\mathcal{N}}\mu_{aj}T_{ab}P_{bj}-\lambda_{ab}T_{ab}.$$

Then the derivative with respect to T_{ab} is

$$\frac{\partial\Theta}{\partial T_{ab}} \;=\; -\sum_{j\in\mathcal{N}}\mu_{aj}P_{bj}-\lambda_{ab}=0$$

therefore, in matrix form this is $\boldsymbol{\lambda}=-\boldsymbol{\mu}\boldsymbol{P}'$.

Next up is to find β in terms of λ and γ. For this we find only the terms depending on f_a. These are

$$f_a \sum_{j:\tilde{q}_{ja}>0} \frac{\tilde{q}_{ja}}{N_a} \log\left(\frac{\tilde{q}_{ja}}{N_a}\right) + \sum_{j\in\mathcal{N}} \gamma_{aj} f_a P_{aj} + \sum_{j\in n(i)} \lambda_{aj} \frac{q_{aj}\mathbf{1}_{q_{aj}>0}}{f_a} - \beta_a f_a.$$

Taking the derivative of Θ with respect to f_a we get

$$\frac{\partial\Theta}{\partial f_a} = \sum_{j:\tilde{q}_{ja}>0} \frac{\tilde{q}_{ja}}{N_a} \log\left(\frac{\tilde{q}_{ja}}{N_a}\right) + \sum_{j\in\mathcal{N}} \gamma_{aj} P_{aj} - \sum_{j\in n(i)} \lambda_{aj} \frac{q_{aj}\mathbf{1}_{q_{aj}>0}}{f_a^2} - \beta_a = 0.$$

Solving for β_a we get

$$\beta_a = \sum_{j:\tilde{q}_{ja}>0} \frac{\tilde{q}_{ja}}{N_a} \log\left(\frac{\tilde{q}_{ja}}{N_a}\right) + \sum_{j\in\mathcal{N}} \gamma_{aj} P_{aj} - \sum_{j\in n(i)} \lambda_{aj} \frac{q_{aj}\mathbf{1}_{q_{aj}>0}}{f_a^2}.$$

This can be simplified by viewing some of these terms as matrix multiplications. The first term is the matrix $\sum_{\text{down columns}} \tilde{\boldsymbol{q}} \times diag(\boldsymbol{N^{-1}}) \cdot \log(\tilde{\boldsymbol{q}} \times diag(\boldsymbol{N^{-1}}))$ where $diag(\boldsymbol{N^{-1}})$ is the diagonal matrix formed by N_a^{-1} on the diagonal. The second term is the diagonal of the matrix $\boldsymbol{\gamma P'}$. The third term is the diagonal of the matrix $\boldsymbol{\lambda(q_{>0})'} diag(\boldsymbol{f^{-2}})$ where $\boldsymbol{q_{>0}}$is the matrix for q_{ij} except negative terms are made to be zero. This summarizes how we can compute β_a using matrix operations.

Lastly, we solve for ν_a. We write down the terms which depend on p_a. They are

$$\sum_j \kappa_{aj} p_a(\nu_a - \nu_j) + \sum_{i\in n(a)} (\beta_a q_{ai}\mathbf{1}_{q_{ai}>0} + \beta_i q_{ia}\mathbf{1}_{q_{ia}>0}) + \sum_{i\in n(a)} \left(\lambda_{ai}\frac{q_{ai}\mathbf{1}_{q_{ai}>0}}{f_a} + \lambda_{ia}\frac{q_{ia}\mathbf{1}_{q_{ia}>0}}{f_i}\right).$$

We know that $\frac{\partial}{\partial p_a} q_{ai} = \kappa_{ai}$ and $\frac{\partial}{\partial p_a} q_{ia} = -\kappa_{ia}$. Thus the derivative with respect to p_a of Θ is

$$\frac{\partial\Theta}{\partial p_a} = \sum_j \kappa_{aj}(\nu_a - \nu_j) + \sum_{i\in n(a)} (\beta_a \kappa_{ai}\mathbf{1}_{q_{ai}>0} - \beta_i \kappa_{ai}\mathbf{1}_{q_{ia}>0}) + \sum_{i\in n(a)} \left(\lambda_{ai}\frac{\kappa_{ai}\mathbf{1}_{q_{ai}>0}}{f_a} - \lambda_{ia}\frac{\kappa_{ai}\mathbf{1}_{q_{ia}>0}}{f_i}\right).$$

Note that $\sum_j \kappa_{aj}(\nu_a - \nu_j) = \Delta_\kappa \boldsymbol{\nu}$. The first term of the second sum amounts to masking the

adjacency matrix for positive flows, multiplying across rows by β_a and summing across the rows. The second term is the mask of the adjacency matrix of the conductances for negative flows with the matrix product by the column vector $\boldsymbol{\beta}$. The third sum can be expressed in a similar fashion. Therefore, solving for $\boldsymbol{\nu}$involves solving for the laplacian of κ_{ij} on terms which are summing across the rows of a matrix or matrix-vector multiplication. This is the format we use for coding the derivative in Matlab.

Finally, we solve for the derivative of the NME with respect to κ_{ab}. The terms depending on κ_{ab} are

$$\lambda_{ab}\left(\frac{q_{ab}\mathbf{1}_{q_{ab}>0}}{f_a}\right)+\lambda_{ba}\left(\frac{q_{ba}\mathbf{1}_{q_{ba}>0}}{f_b}\right)+(\beta_a q_{ab}\mathbf{1}_{q_{ab}>0}+\beta_b q_{ba}\mathbf{1}_{q_{ba}>0})+(\nu_a-\nu_b)\kappa_{ab}(p_a-p_b)$$

and taking the derivative with respect to κ_{ab} we get

$$\begin{aligned}\frac{\partial\Theta}{\partial\kappa_{ab}} &= \lambda_{ab}\frac{\mathbf{1}_{q_{ab}>0}\,(p_a-p_b)}{f_a}+\lambda_{ba}\frac{\mathbf{1}_{q_{ba}>0}(p_b-p_a)}{f_b}(\beta_a(p_a-p_b)\mathbf{1}_{q_{ab}>0}+\beta_b(p_b-p_a)\mathbf{1}_{q_{ba}>0})\\ &+ (\nu_a-\nu_b)(p_a-p_b).\end{aligned}$$

Since we have solved for ν_i,β_i and λ_{ij} we have an expression for the derivative with respect to κ_{ab}.

6.7.4 Derivative of the dissipation

To get the derivative with respect to κ_{ij} of the dissipation we can solve it in a similar method to the above to essentially prove what we cite as Cohn's theorem. The derivative is $\frac{\partial D}{\partial\kappa_{ij}}=-\frac{q_{ij}^2}{\kappa_{ij}^2}$ also written $\frac{\partial D}{\partial\kappa_{ij}}=-\left(p_i-p_j\right)^2$.

6.7.5 Derivative of the pre-exponential and the re-scaling.

For the derivative of the re-scaling we need to consider derivatives with respect to the conductance we are currently re-scaling and derivatives with respect to conductances that

are not the ones being multiplied. We get

$$\begin{aligned}\frac{\partial}{\partial \kappa_{kl}}\frac{C\kappa_{kl}}{\sum \kappa_{ij}^{\gamma}} &= \frac{C\sum \kappa_{ij}^{\gamma} - C\gamma\kappa_{ij}^{\gamma}}{\left(\sum \kappa_{ij}^{\gamma}\right)^2}.\\ \frac{\partial}{\partial \kappa_{xy}}\frac{C\kappa_{kl}}{\sum \kappa_{ij}^{\gamma}} &= \frac{C^2 - C\kappa_{kl}\gamma\kappa_{xy}^{\gamma-1}}{\left(\sum \kappa_{ij}^{\gamma}\right)^2} \text{ for } x, y \neq k, l.\end{aligned}$$

The derivative with respect to the pre-exponential is just to multiply by $\exp(-\tilde{\delta})$ elementwise.

6.8 Method of sampling and curation of global optima

We use Theorems 5.4, 5.6 in Chapter 5 as a guideline for our sampling procedure. Our mathematical result from that section guarantees the existence of a γ_ϵ such that for $\gamma < \gamma_\epsilon$ loopy networks with $-\log((m+2)!) + \epsilon < \text{NME} < -\log(m!) - \epsilon$ do not contribute to the graph of θ restricted to $c_{m+d,m} \leq c \leq c_{m,m-d}$, where d is the permissible increment between path lengths between two points (recall it is 1 for a triangular grid and 2 for a square grid). It does not directly give insight into the graphs at different values of γ nor does it tell us what the collection of optimal networks will be for $c_{m+d,m} \leq c \leq c_{m,m-d}$. Still, we use the idea that the graph of θ and the optima of the mixing-dissipation cost have a different characteristic on each interval $c_{m+d,m} \leq c \leq c_{m,m-d}$ as a heuristic.

By Theorem 5.4 minimizing over tree graphs always leads to shortening by the minimal permissible path size difference as we pass from $[c_{m+d,m}, c_{m,m-d}]$ to $[c_{m,m-d}, c_{m-d,m-2d}]$ where the optima over the first of these intervals is a path of length m, i.e. τ_m and the optima over the second of these intervals is a path of length $m-d$, i.e. τ_{m-d}. In practice running the optimization algorithm of the paper and by the investigations in [BM07, Dur07, Cor10, KSM10] optimization of *minimal dissipation* networks with a given set of sources and sinks while maintaining a fixed material cost with exponent $0 < \gamma < 1$ gives a tree structure, and combining other types of functions produces loopy but with betti number decreasing for

decreasing γ. Our algorithm has yielded networks which are similar to paths with a loopy "head" sub-network at the source set, so the discontinuous changes in the path length over trees as the weight is taken to be in different intervals in the partitions over intersection points does appear to affect the minima at different levels of γ. By this observation it seems important to sample each interval $[c_{m+d,m}, c_{m,m-d}]$ with equal thoroughness.

Our strategy is as follows: first compute the values of c for the intersections between lines from paths of neighboring lengths τ_m and τ_{m-d}, $c_{m,m-d}$ for $2N-1 < m < N^2$ where N is the side length of the grid. We also take a left end point *book end* of .01 to be close to 0. This is preferable to choosing $c = 0$ because the NME only depends on flows, and an arbitrarily irregular set (in that some may be very large and very small) of conductances which may lead to numerical instability can produce the same set of flows. For the right-most end point *book-end* we use $c_{2N-1+d,2N-1} + x$ where $x = 1$ or 3 depending on the setting. In our cases for both triangular and square grids it is $x = 1$ for $\gamma = .5, .2$ and the latter $x = 3$ for $\gamma = .8$. We include the intervals $[0.01, c_{N^2}]$ and $[c_{2N-1+d,2N-1}, c_{2N-1+d,2N-1} + x]$ in our sampling. Then between consecutive points, $c_{m+d,m}, c_{m,m-d}$ or the lower and upper book-end intervals pick T equally spaced points,

$$c_{m+d,m}, c_{m+d,m} + \left(\frac{c_{m,m-d} - c_{m+d,m}}{T}\right), c_{m+d,m} + 2\left(\frac{c_{m,m-d} - c_{m+d,m}}{T}\right), \\ \ldots, c_{m+d,m} + (T-1)\left(\frac{c_{m,m-d} - c_{m+d,m}}{T}\right).$$

And run the optimization algorithm on each of these weights for each pair of consecutive intersection points out of the possible intersection points on the grid once. The optimization algorithm is greedy in nature, and therefore produces only approximates the optimal values of the mixing-dissipation cost. The quality of the approximations are indeed random variables given by the initialization of the optimization routine and the randomly chosen growth directions and topology changes. This may make it seem to be a good idea to sample each of these points many times to get better approximations of the optima. It is actually both

more accurate and simpler to instead increase T the number of sub-divisions.

To see this intuitively, assume for argument that the probability of the optimization routine terminating at the optimal network is non-zero, p. Let $\mathbf{T}$ be the total number of subdivisions of an interval and say we are trying to approximate the function on that interval. If we fix $\mathbf{T}$ and sample each point enough, then the best approximation we can be guaranteed to obtain of θ at *any* c in the interval is a piecewise linear function with $\mathbf{T}$ disjoint pieces. If instead of repeatedly sampling each subdivision we increase $\mathbf{T}$ then for large $\mathbf{T}$ the number of perfectly approximated optima is roughly $p\mathbf{T}$ uniformly sampled points in the interval. Although each local optimum corresponds to a specific value of c, the same network can be used for different values of c, producing an affine cost function (since both dissipation and NME are fixed). Thus our numerical minima give us $p\mathbf{T}$ affine functions to minimize over and this quantity can be increased to our desire. It is also clearer and quicker to implement and adjust because it requires only one for-loop and one value controlling simultaneously the number of subdivisions and sample size. Our computed local optima enable us to approximate θ as the envelope of the affine cost functions of our computed locally optimal networks, $\hat{\theta}$. This envelope is defined as follows:

Definition 6.1. For $\mathbf{T} > 0$ samples of positive real numbers c_i and the corresponding networks κ_i, total negative mixing entropies NME_i and dissipations D_i for $1 \leq i \leq \mathbf{T}$ define the envelope of κ_i , $\hat{\theta} : \mathbb{R}_+ \to \mathbb{R}$to be the pointwise minima of the collection of mappings

$$\{c \mapsto \mathrm{NME}_i + cD_i : 0 \leq i \leq \mathbf{T}\}.$$

That is for all $c > 0$

$$\hat{\theta}(c) \quad = \quad \min_{0 \leq i \leq \mathbf{T}} \mathrm{NME}_i + cD_i.$$

The envelope thus defined gives the best mixing-dissipation cost at each value of c, searching over all of the computed local optima.

6.8.1 Global optima

In the spirit of sampling at one point many times, we can subdivide the space again after we perform many samples. Then in each sub-division we find the distance between the cost functions of all of the local optima generated in that interval and the envlope of cost functions defined above. We then identify the network that has the minimum distance from the envelope.

We apply this notion of an optimal network after we compute the mixing-dissipation minima for the sampled c. We take a sequence of points, starting with the minimal c (usually $c = .01$) and ending with the maximal c. After the minimal c, the sequence is $c_{n,n-d}$ where d is taken to be 1 or 2 based on the ambient grid. From the last intersection point, ($c_{11,9}$ or $c_{10,9}$) to the maximal c, we include a sequence subdividing the last interval. The intervals formed by subsequent elements in this sequence are referred to as **search-intervals**.

We then choose a number of **sub-divisions** for each search-interval $[a, b]$. If the number of sub-divisions is 1 we take c and its corresponding optimal network κ_{ij} minimizing

$$\mathrm{NME}(\kappa_{ij}) + cD(\kappa_{ij}) - \hat{\theta}(c). \tag{6.8.1}$$

If the number of sub-divisions is $n > 1$ then for each $m = 0, 1 \ldots, n-1$ we take the c and κ_{ij} minimizing the expression 6.8.1 over the interval $\left[a + m\left(\frac{b-a}{n}\right), a + (m+1)\left(\frac{b-a}{n}\right)\right]$.

6.9 Global optima for each numerical experiment

Every numerical experiment minimized the mixing dissipation cost on ambient grids of side length 5. We also hold the total material cost constant at 24. That is, for all material cost exponents γ used in the numerical experiments, the conductance networks obey the constraints. $\sum_{(i,j)} \kappa_{ij}^{\gamma} = 24$. We do this to fix the slope line generated by the tour τ_{25} as a reference point for the plots of each curve to be compared against. Here we list the search

intervals and the number of sub-divisions for each experiment.

- Square grid with boundary flows of one source and one sink with exponent $\gamma = .5$
  - Search interval end-points: .01, $c_{n,n-2}$ for $25 \geq n \geq 11$, $c_{11,9} + 1$. Number of sub-divisions: 2.
- Triangular grid with boundary flows of one source and one sink with exponent $\gamma = .5$
  - Search interval end-points: .01,$c_{n,n-1}$ for $25 \geq n \geq 10$, $c_{10,9}$+1. Number of sub-divisions: 1.
- Square grid with boundary flows of one source and one sink with exponent $\gamma = .8$
  - Search interval end-points: .01,$c_{n,n-2}$for $25 \geq n \geq 11$,$c_{11,9} + \{1, 2, 3, 4, 5, 6.5, 7\}$. Number of sub-divisions:1.
- Triangular grid with boundary flows of one source and one sink with exponent $\gamma = .8$
  - Search interval end-points: .01,$c_{n,n-1}$ for $25 \geq n \geq 10$, $c_{10,9} + \{2, 4, 6, 7\}$. Number of sub-divisions:1.
- Square grid with boundary flows of two sources and two sinks with exponent $\gamma = .5$
  - Search interval end-points: .01,$c_{n,n-2}$for $25 \geq n \geq 11$, $c_{11,9}, \frac{c_{11,9}+13.01}{2}, 13.01$. Number of subdivisions: 2.

6.9.1 Sampling details

For one source and one sink, we have found that, in practice, 50 evenly space points between intersection points is enough to get a good approximation of the graph of θ but that if $\gamma \leq .5$, near the intersection points between paths are strongly attractive and twice as many samples are needed to accurately resolve the optima near these points. Therefore, we choose 100 sub-division of each interval and apply this uniformly. Similarly, even at smaller values

of γ, if there are multiple sources and sinks in the network, a more in depth investigation of the nature of the networks than is possible for this thesis would be necessary to be sure that we are finding the true global optima.

We note that, some of our numerical experiments use even finer sampling of c values. For the 2 source 2 sink square grid at $\gamma = .5$ with one half of the evenly spaced points are sampled twice and the interval has between 4 and 6 times as many sampled c values than our experiments with single source, single sink networks. This is because we consolidated data from multiple trials, where we attempted to partition the sampling over several machines because running many experiments simultaneously resulted in running out of memory. We do not eliminate these data points because they only increase the robustness of the results. Summarizing, the majority of intervals in the experimental results have 100 evenly spaced samples and some have close to 200.

6.9.2 Optimization tuning parameters

For the 5×5 triangular grid with source inflow of 1 at 1 and sink outflow of -1 at 25 for each step of the optimization routine we choose the number of iterations of the growth step bernoulli-$\frac{1}{2}$ from $\{1, 2\}$ and after choosing the number of growths, we choose the number of topology changes greater than or equal to 0 and less than or equal to the number of growths. Two is chosen as the max possible number of growths because one growth step is only guaranteed to add at least a cycle with 3 sides to it, while 2 growth steps may add several conjoined squares with their diagonals, thus allowing for larger loops. For 5×5 grids it seems like more growth steps is not necessary. From the initial conditions alone, loops may be included in the network, and their size can be elongated or shrunk by the growth step, so it is believed that large enough loops are taken into consideration. Since our optimal networks contain loops with no more than two extra nodes on each side of the loop, and our growth algorithm allows larger loops to be made, we believe that two steps of growth are sufficient to fully explore the space of allowed networks. Growth steps add material

uniformly, and perturb the network away from any optima that is has found, so there are tradeoffs between having enough growth steps to move the optimization method away from a poor local optimum, while still allowing it to reconsolidate material back to a strong global optimum if it is perturbed away from one.

6.9.3 Post-processing the results

We sample points according to the method described in Section 6.9.1 after first weakening the already weak edges of the network. Specifically any edge with conductance $\kappa_{ij} < 1 \times 10^{-3}$ is assigned the much smaller conductance $\kappa_{ij} = 1 \times 10^{-9}$. Running the interior point network on the new support, and allowing it 5000 iterations gives a very small but noticeable improvement on the approximation of θ using the envelope $\hat{\theta}$. We check the validity of this filtering by comparing our putative global optima from Section 6.9.1 with the results of Section **??**. Without this pos-processing our approximation of the optimal networks is only mildly worsened. In particular, the algorithm does not allocate material so uniformly along the edges of any paths within the network, as we would expect it to for minimization of dissipation.

CHAPTER 7

Discussion

7.1 Changes to the model

We have defined two notions of mixing entropy both of which are tied to the assumptions that networked organisms' responses to stimuli emerge from responses occuring at every node in the network, and that each such node receives signals from other nodes through advective currents. The negative mixing entropy models the uniformity signals that each node receives from other nodes in the network. We can think of it as representing the diversity of signals arriving at each location in the network. On the other hand the negative sending entropy models the diversity of locations within a network that signals originating from each node may reach. Both measures of mixing carry biological importance: A flow network that did not evenly disperse information (i.e. with low sending entropy) may be deficient in terms of its ability to marshal an organism-wide response to new information, such as the presence of new food sources, or localized fungivory. A flow network in which individual nodes do not receive a diversity of signals, may hinder the ability of those nodes to make decisions about the cues that the entire network is subject to, as well as being vulnerable to resource plundering, if the signals in consideration are nucleotypes, and too many mutant nucleotypes congregate at a single point within the network.

Our model makes the most basic assumptions about node importance, that every nodes signal or response is proportional to the amount of flow it controls. To model biological networks, the weighted importance of nodes or strengths of signals should be informed by data. For instance, the signals from the spatial periphery of a foraging network might take

higher importance than interior regions which are less likely to discover new food sources. On the other hand interior regions are important for collecting information about risk such as toxins or light sources, and as the primary sites for nuclear division, may generate new nucleotypes also.. Another consideration is modelling the generation of molecular cues or nuclear divisions as a random variable. In our model signals occur at each node and are transerred downstream with a rate proportional to the flow. Since nodes occur every one unit of length, this is akin to assuming external stimuli are generated at a constant nonrandom rate per unit length or area. An example further direction is to formulate a model where information is instead generated by a poisson point process on a planar region with parameters informed by experiments.

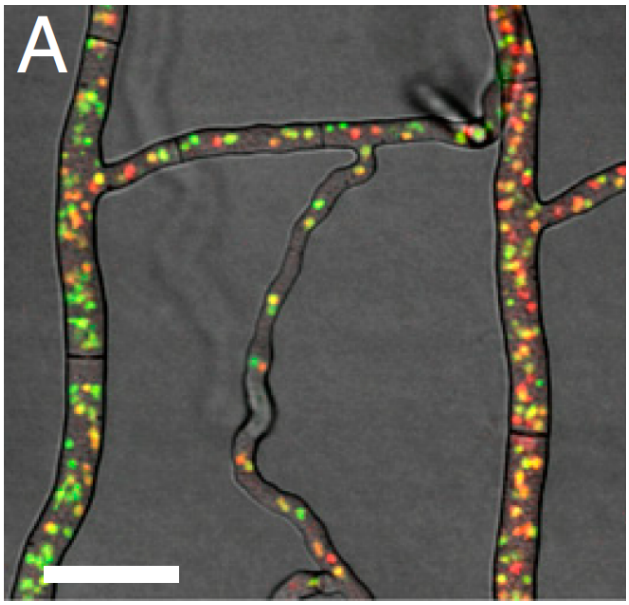


Figure 7.1.1: Image of multi-nucleate mycelium from [RSH13] with two fluourescent tagged nucleotypes. Note how, together, the different nucleotypes form the entire population of nuclei being transported by the flow.

Our mixing entropy is defined with the assumption that advected signals do not interact. That is their concentration is much smaller than the total volume that is transported. This assumption may be appropriate for molecular cues, but is not a good model for experiments involving of red and green fluorescent tagged nuclei in *N. crassa* [RSH13] in which all flowing

nuclei must be assigned to one or another population (Fig. 7.1.1). The consequence of non-interacting signals is that the ratios of the signals in the normalized probability distribution are proportional to the flow originating at the signalling node and arriving at the destination. If, on the other hand, signals consume a significant portion of the volume along each conduit, their summed strength is equal to the flow along one conduit and they divide this flow up, that is, rather than normalizing probabilities of nuclei arriving at a particular point in the network by the total transmission probability of getting from each node to the point, we must normalize by the flows arriving via the conduits that the node is connected to.

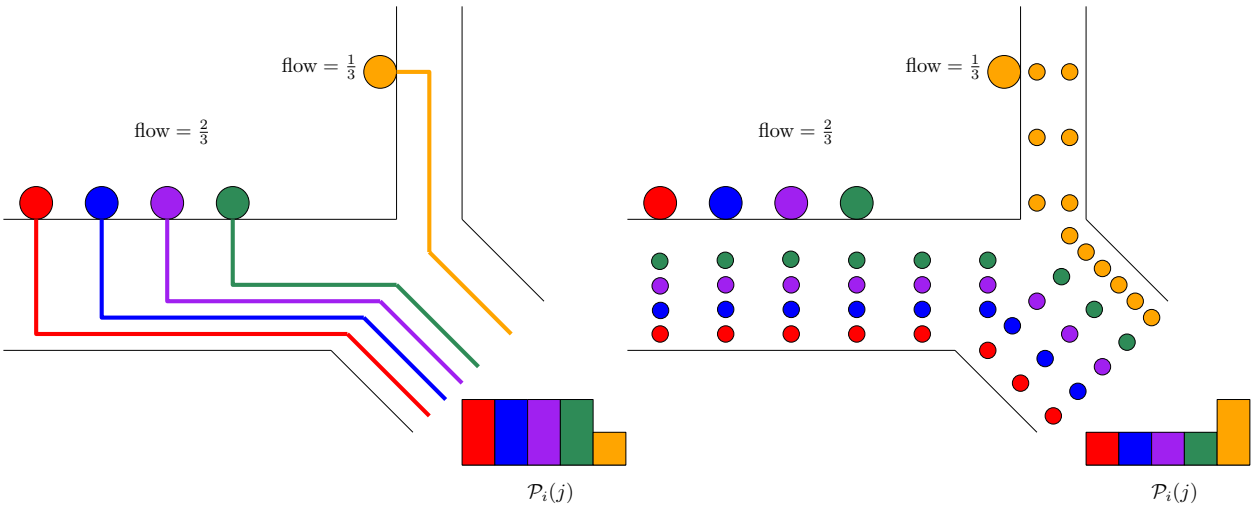


Figure 7.1.2: Left: Diagram of signal strengths and distribution for computation of the NME presented in this paper. Right: A diagram of signal strengths and distribution for computation of a mixing entropy model whose development is for further research.

The left panel of Fig. 7.1.2 represents the model that is the subject of this paper and does not assume interaction between signals. Since the along the tubule with the first four signals has twice the intensity of the second inflowing tubule, each signal probability from the first tubule is twice as large as the probability from the second tubule. Therefore the probability of each signal from the stronger flow is $\frac{2}{9}$ and the probability of the signal from the weaker flow is $\frac{1}{9}$.

The right panel of Fig. 7.1.2 represents a different model from the one studied with the

assumption that signals take up the total volume of flow. The total volume of signals per tubule is represented by a collection of disks. Note that the incoming flow of magnitude $\frac{2}{3}$ is transferring twice as many disks as the flow of magnitude $\frac{1}{3}$. Since there are 4 different types of disks, the occurence of each type of disk from the larger flow is $\frac{1}{4} \times \frac{2}{3} = \frac{1}{6}$ whereas the occurence of the orange disk from the weaker flow is $\frac{1}{3}$.

7.2 Further directions for improving numerical simulation

Our ambient grids have realistically bounded degree and the nodes are a dense approximations of points in space if we increase the grid length and re-scale conductances to produce a mesh with bounded total material. This does not imply that the sets of paths will become closer approximations to all paths in eulidean space. For square grids the graph distance between points whose vector difference makes an angle $\pm\frac{\pi}{4}$ from the horizontal line is always a factor of $\sqrt{2}$ larger than the euclidean distance, because of the city block paths that flows must take between the two vertices. Similarly, for triangular grids the graph distance between points whose vector differences make an angle $\frac{\pi}{6} + n\frac{\pi}{3}$, $n = 0, 1, \ldots, 5$ is always factor of $\frac{3}{2}$ larger than the euclidean distance. With our ambient grids, we are bounded away in accuracy from approximating arbitrary continuous paths of any arc-length. Since real biological transportation networks are not bound to any particular ambient grid, to realistically model how organisms construct optimal 2-dimensional conductance networks for any function [KSM10, TTS10, Cor10, BM07] we must consider how well the ambient graph approximates the euclidean distance [Che89, KG92]. Another approach to choosing an ambient grid is to mimic the properties of biological networks. Mycelia such as those of *N. crassa* tend to branch from a food source with bounded degree and each branch has its own hyphal branching and some cross-linking, thus allowing networks to be approximated by lattices that are less robust than geometric spanners [KG92]. Slime mold networks have a phase that can be approximated by a random lattice from which pruning may occur to leave a mature presumably optimal network [TTS10, Ali18, NYU00].

We know single source single sink networks with a material cost exponent $\gamma = .5$ will contain more loops if simulated on larger grids. But direct numerical optimization of networks on large ambient networks carries high computational costs. Changes to our model such as those suggested in Section 7.1 may affect the topology such as altering the assumption that the signal receivers and senders occur once per unit of length. Research into improving the efficiency of our algorithm would provide further insight into how such networks would look in a continuous setting by successively subdividing them into regions of fixed length. Since our research shows that optimal networks naturally subdivide into subnetworks (such as the loop-path structure identified in Section **??**) this suggest that the complexity of the optimization can be reduced by dividing the ambient grid and optimizing each portion separately.

Although real biological transportation networks have features that resemble our mathematically calculated optimal networks, it is not clear how real biological transportation networks find optimal structures. Structural adaptation is a mechanism by which networks adjust the radii of the flow conduits in response to the flows in the network. For example increasing conduit size in conduits with high flow and otherwise decreasing tubule leads to shortest paths becoming high volume cords, and to the reliable construction of minimally dissipative networks [Ali18, AKF17, TTS10]. Is there an algorithm for structural adaptation that enables networks to optimize their mixing?

Our analysis of optimal mixing networks has highlighted the balance between loop formation and paths created by tradeoffs between minimizing dissipation and maximizing mixing. Our analysis shows that gamma plays a role in controlling this balance. But for real systems, gamma is likely fixed, since it reflects the cost of building network conduits. To explore loop formation independently of gamma, we propose to consider a new cost function that adds a penalty term for average path length $L(q_{ij})$ multiplied by a scalar d, may lead to networks with loops at $\gamma = .5$, especially if $\hat{c}$ is a weight on dissipation known to produce paths of intermediate length (e.g. 19 on ambient grids of side length 5). For example it would

be interesting to numerically minimize NME $+ dL + \hat{c}D$ under the material cost constraint $\sum_{(i,j)} \kappa_{ij}^5 = 24$ and $d \geq 0$. Increasing d will decrease the average path length, and the fact that shorter paths than 19 are not favored by the mixing-dissipation cost may result in a loopy subnetwork appearing near the origin. As we discuss in the next paragraph, we expect our optimization algorithm to be appropriate for this new cost function.

7.3 Other uses of our optimization algorithm

Our optimization algorithm iteratively improves the conductance network, and is considerably more robust than gradient descent when the function of interest is not differentiable, or when it becomes non-differentiable when any of the conductances in the network approach 0. Although it was designed with our function in mind, it is well suited for other functions, especially those optimizing a cost function on conductance networks of the form $F(q_{ij}) + D(\kappa_{ij})$ where F is a function defined in terms of a flow network and D is the dissipation of the physical flow given a conductance network. Our algorithm iteratively adjusts the topology of the flow network and finds local optima of F within the region of the set of conductances giving rise to this topology. The second uses a continuous optimization algorithm such as gradient descent or, in our application, the interior point method

The function F may depend on the topology of the flow network in terms of its directions. At the same time continuously optimizing dissipation promotes sparse, tree-like networks. Therefore to search for a new local optimum, extra conductance material needs to be added to the network.

We then solve for a new conductance along a randomly chosen edge that swaps the flow along one edge. Special consideration needs to made to preserve the progress in finding the optimal topology of the flows. If too many new edges are included in the conductance network, the random choice of a conductance swap may close off a large swath of paths in the flow network from the previous step. We preserve the topology of the network by avoiding

the introduction of short-cuts when adding material. The design of this algorithm is not specifically geared to optimize the NME and could be used for any function on the flows. One example is $\sum_i f_i \sum_{j \in n(i)} \frac{q_{ij}}{\sum_{k \in n(i)} q_{ik}} \log \left(\frac{q_{ij}}{\sum_{k \in n(i)} q_{ik}} \right)$, which is equivalent to the entropy of the distributions of paths from the sources to a node i [TT93]. This has a similar form to the NME in that it is a continuous function of flows at each node weighed by the flow through the node, guarenteeing it is continuous on the hyperplanes where both nodes and edges are erased from the network. Including a penalty term for average path length might also lead to loops in the $\gamma = .5$ setting.

7.4 Further directions for theoretical results: 1. Network size scalings

A subtlety in our mathematical analysis and the proof of Chapter 5 Theorem 5.6 and Theorem 5.7 is that it relies on the size of the ambient grid. We do not have any statement about fixed c and larger ambient grids with the same connectivity. A real biological transportation network can expand the size of its ambient grid by generating features on smaller and smaller length scales. By contrast, the numerical results in 4.3 have bounded Betti number. This along with the visualization of the conductance networks in Fig. 4.3.2 Fig. 4.3.8 reveal that the parameters lead to optima with a bounded set of edges with medium conductance and flow.

It is not known whether optima for similar values of c on larger networks will have the same number of edges. Is there a network size after which a phase transition occurs, that a small number of medium strength edges becomes sub-optimal to many very low conductance edges? The connection between ambient grid connectivity, size and optimal networks should be investigated to see if optimal networks do not depend on their ambient grid. Nonetheless, even if there is a dependence on the size of the ambient grid, real biological networks can not generate features down to arbitrarily small length scales, since the physical diameter of conduits is constrained by the granularity of protoplasm: a *N. crassa* hypha must be large

enough for a nucleus or vacuole to easily pass through it.

There are other functions that are dependent on the ratios of total flows at nodes for which Theorem 5.8 holds. One example is flow entropy as defined by Tanyimboh and Templeman [TT93]. This is defined for a flow network q_{ij} as $\sum_i f_i \sum_{j \in n(i)} \frac{q_{ij}}{\sum_{k \in n(i)} q_{ik}} \log \left(\frac{q_{ij}}{\sum_{k \in n(i)} q_{ik}} \right)$. Can a similar theory be built about the cost function penalizing dissipation with the flow entropy? As we have explained above, our numerical methods seem to be suited to optimize such a function, and the optimal networks on a fixed number of nodes might be able to be ascertained this way. Also, the asymptotics of functions on flow such that the intersection points $c_{m,m-1}$ are decreasing with increasing m defined for any function can be discussed. Would it hold for flow entropy? The fact that the theorems hold allow a convenient way to select for networks by their size using the importance placed on dissipation.

7.5 Further directions for theoretical results: 2. Understanding loop placement and complexity

As we have noted in Chapter 4 Section 4.3, the globally optimal networks in Fig. 4.3.2 and Fig. 4.3.8 indicate that if there are loops in an optimal network, they tend to occur with smaller loops closer to the source and larger loops further from the source. A further direction would be to run numerical simulations on larger grids or to attempt to state and prove theorems in the style of loop placement theorem (Chapter 6 Section **??** Theorem 5.10) that compare loops of different sizes and interconnectedness perhaps characterized by average path length of the loops, occuring closer to the source when optimizing the NME. We see in our experiments optimizing mixing-dissipation cost for networks with many sources that contribution to average path length shifts from paths with a flow of 1 to paths with less flow occuring towards the sources of the network. Is there a single trend that contains the trends of combining of flows from many sources and placement of loops?

Hierarchical, self similar loops has been studied from the point of view of resilience [KM12,

Cor10] and also as promoting dispersal of signals [MAA16]. The results from optimizing the mixing dissipation cost with many sources and one sink **??** indicate that flows originating from multiple sources combine hierarchically, with separate flows originating from a partition of the sources combining into separate paths, a partition of those paths combining into a set of paths with larger flows and so on. Both of these observations could result in the large network limit in self-similarity. Analysis of the structure of loops using an approach such as that described in [KM12] or persistent homology [KM12, EH08] could shed light on whether the flows obey either a hierarchical or self-similar nature.

7.6 Connection to biological fluid flow networks

The first organism that led us to study mixing, *N. Crassa* is tree-like, extending its mycelia in many directions. It is not highly reticulated in comparison to some of the optimal networks we have obtained, yet experimental evidence demonstrates that it mixes different nucleotypes throughout its network using advective currents. What are then the morphological features that lead to the capacity to uniformly integrate nuclei across its network? One observation that may shed light on this question is that *N. crassa* is known to posess hyphal links between up-stream locations along its conduits [RET11]. There is evidence that fungal mycelia follow Murray's law [HOG12], yet this would imply that any edges besides those needed for flows in the direction of the growing tips is energy inefficient.

A first attempt to model *N. crassa* mycelial networks would be to optimize the mixing dissipation cost on a grid with one source at a corner and 3 sinks drawing $\frac{1}{3}$ of the demand of the flow each. This would match an assumption of three main conduits. From our numerical studies and experience simulating mixing-dissipation cost minimizers we expect the optima to essentially follow a transition from a path that goes through evey node with the sinks placed at the most downstream positions possible. This, or any other negative result, should not necessarily dissuade us from consider NME when studying mycelial flows. The presence

of up-stream hyphal linking is a feature that is actually posessed by the optimal networks $\gamma = .8$.

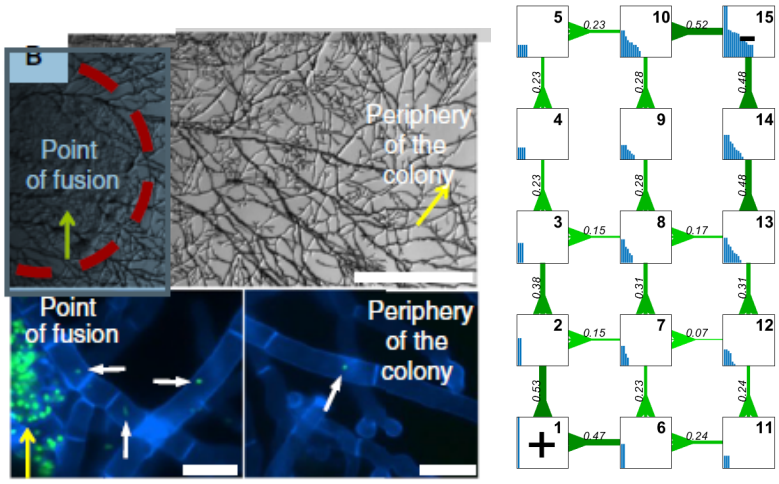


Figure 7.6.1: Left: Pictures of *N. crassa* mycelia from [RET11]. Right: Loopy subnetwork from an optimal flow network on the sqaure ambient grid.

In the left panel of Fig. 7.6.1 we have a loopy subnetwork extracted from an optimal network. Note that there are links running horizontally along the three vertically running vessels. Although this network has only one sink, the small loops near to the source mix the flows as the space between further linking becomes longer. We can conclude that linking postively effects the mixing in this scenario. Perhaps it increases the mixing in more tree-like networks. An added penalty related to average path-length lead to mixing-dissipation optimizers with the 1-source many-sink boundary flow being mostly comprised of the shortest path from the source to the three sinks, but including cross-links to mix the signals. As we have mentioned in Section 7.1 and depicted in Fig 7.1.2, the assumptions of whether signals combine to make up the total volume of the flow should be taken into account when modelling nucleus mixing flows. Perhaps optima under this assumption will resemble mycelial flow network topology more closely.

An average path length penalty makes sense because it has been observed that morphology of mycelia of some fungal species change as the organism is starved, leaving the shortest

cords to connect different regions in the network. Dissipation does promote shorter paths, but average path length is important for constructing a network backbone which serves as high volume transport tubules even after the morphology has gone through a lot of changes. That is, we want to optimize mixing and dissipation for the flows at the current time, but at later times and different scheme flow schemes, a shorter transport path will prove to be of continuous use.

It has been observed experimentally that slime mold in its dendritic form [NYU00] exhibits the highest rate of mixing. This form occurs after pruning network tubes, usually in the order that tubules towards the center are pruned first. The ones that are pruned are thought to be those that are along paths with sub-optimal lengths, and the ones that remain provide low dissipation pathways for flow [NYT01, THA00b]. The characterstic of being "fanned" towards the exterior, with connections to central veins has been noted to promote the mixing and as well as disperal collection of food or molecular cues [AKF17, MAA16]. As evidence that energy considerations for dissipation from friction and network upkeep hold, slime mold exiting food-deprived environment has been shown to obey tubule radii measurements suggesting the network obeys Murray's law [AKF17].

There is semblance between networks built by our numerical experiments and these real life networks. For NME optimal networks, if there are loops, they occur towards the source. For evidence see the results of the mixing-dissipation cost optimized for $\gamma = .8$ in Chapter 4 Section 4.3 and Chapter 6 Section **??** Theorem 5.10 for evidence. Although we are not aware of a physical scenario corresponding to material cost exponent $\gamma = .8$ it still represents a penalty applied to conductances to a higher number of distinct vessels, a realistic constraint to consider when slime mold is in a starved state.

An interesting model inspired by slime mold would be $\frac{1}{2}(\text{NME} + \text{NSE})$. This is can be understood as the ability for the flow to collect and mix signals as well as the flow to disperse signals to a variety of nodes. As proven in Chapter 2 Theorem 2.1, this is the average of the NME of equal strength flows in opposite directions. Since NME optima are reticulated near

to the source, we predict that optimal networks with this function appear to be reticulated towards both the source and sink. If we orient the sources and sink along the boundary, we might expect that conductance networks optimizing this bi-direction NME and dissipation are reticulated around the spatial periphery.

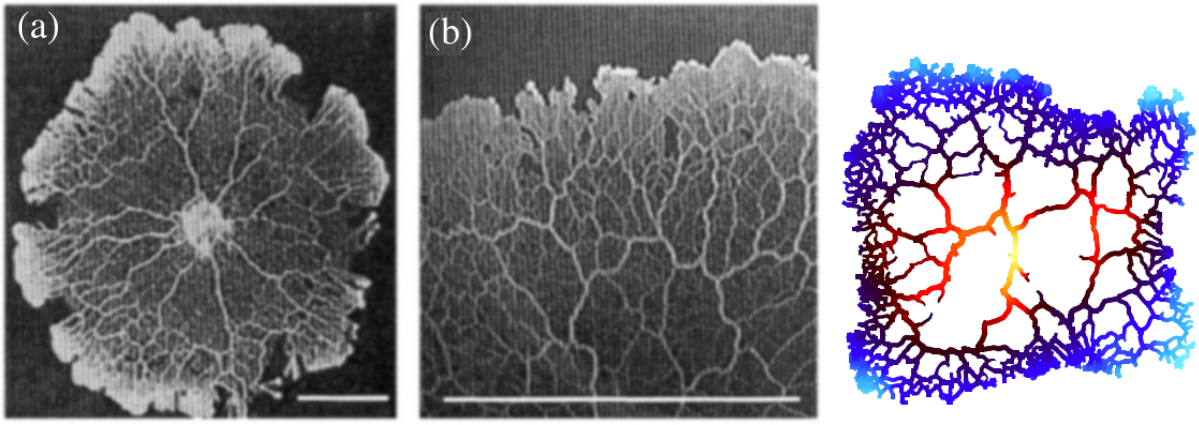


Figure 7.6.2: Image of the dendritic morphology as referred to by Nakagaki et al. in [NYU00] (left). Pruned slime mold network colored by the effective dispersion (Marbach et al) in [MAA16] (right).

Although our model is simple, this has a connection to flows in *P. polycephalum* as they are believed to flow in two opposite directions when caused by a contraction wave spanning the entire network [AAP13]. Since the NSE is a measure of dispersal, it would be an interesting to investigate quantitatively the degree of reticulation towards the periphery in starved slime mold networks [Ali18]. As we have mentioned, we conjecture that optimal networks are more hierarchically reticulated towards the source of the flow (alternating source and sink for both flow directions). Since path length plays a central role in the loop placement proof it seems as though that concept can bridge a connection between the placement not only of a loopy subnetwork, but of loops in general by size. This idea could be especially applied under a material cost constraint. In slime mold foraging, material is continually sent to the expanding fans shown in Fig. 7.6.2. Our modeling and analysis suggest that the fans may ensure that material is maximally dispersed among the possible destinations within a fan.

Another comparison that would be interesting to pursue is how the NME changes as fungi or slime mold inspired adaptive growth models are evolved in time. For instance, the model of flow strengthening tubules which are otherwise decreasing in strength at a constant rate results in formations that contain central cords or bottle necks, similar to our optimal networks [AKF17].

Some mathematical models based on the assumption that networks optimize a function, such as Murray's law, have had success in explaining experimental measurements. This assumption in other cases is too simplistic. We do not purport that networked organisms optimize the mixing-dissipation cost alone because there are many other factors relevant to their survival. Still, the NME would be elevated for networks that appear to have mixing flows. This corresponds less to finding optimal networks and more to computing the dependence of network functions on existing networks.

Another approach would be to compute the correlations of various functions on numerically generated networks. For example: what are the correlations of dissipation with broken links [KSM10], flow entropy and NME? Such a study can illuminate structural features that positively influence various functions of flow networks. Such a study has been done before comparing resilience to network entropy on city water distribution networks [GDS12].

CHAPTER 8

Role Detection in Bicycle-Sharing Networks Using Multilayer StochasticBlock Models

(Joint work with Jane Carlen, Jaume de Dios, Shyr Shea Chang, Stephanie Wang and Mason Porter)

8.1 Introduction

It is useful to view cities as large spatial networks under constant evolution, with intermittent large-scale changes [Bar18, Bar19]. Transportation systems and commuting patterns shape and reveal the functional regions in a city [BBG18], and an increasing amount of evidence suggests that polycentric urban structures tend to emerge from classic monocentric structures [DBC07, ZAH14, RKB11, GBD19]. The combination of a burst in the development of network-analysis methods and the increasing availability of transportation data gives exciting opportunities to improve understanding of urban dynamics. In the present paper, we construct a novel statistical model of networks and aim to uncover the roles of regions in a city through the lens of travel patterns between them.

There is now a deluge of data about transportation and other urban systems, and they offer the potential for numerous fascinating and important insights about cities and human dynamics in urban environments. Local governments release data on many transportation modes, including traffic from buses, trains, and automobiles. Additionally, online databases such as OpenStreetMap [Boe18] and Global Road Network [SGS17] include large-scale road

networks with tens of thousands of streets, covering a total length of tens of millions of kilometers. Subway data in many cities are available from electronic ticketing systems [ZAH14, ZKK18]. However, the infrastructure — such as train lines, bus routes, and highway systems — that underlies these systems adapt very slowly to changing commuting patterns. Pedestrian traffic adapts faster, but it captures only short-distance transportation and is difficult to measure.

Bicycle-sharing systems are an emerging mode of urban transportation that can adapt quickly to the needs of travelers. The number of bicycle-sharing programs worldwide has grown rapidly, from 5 in 2005 to 1571 in 2018 [Sch18]. Over 50 systems were launched in the United States alone from 2010 to 2016, and over 20 bicycle-sharing systems have been launched in France since 2005 [nac18, EL14]. Many existing bicycle-sharing systems are also growing. For example, the number of stations in New York City's "Citi Bike" system has more than doubled since it began in 2013. Docked bicycle-sharing systems follow a general structure: Groups of bikes are parked at "stations" (also called "docks" or "hubs") throughout a coverage area, and users withdraw and return bikes to these stations on demand, with a cost that depends on usage time. A growing portion of bicycle-share systems are dockless (as are the increasingly prominent e-scooters), so users can park bicycles at any location in a coverage area. In the present paper, we analyze docked systems (but we consider how to adapt our models to dockless systems in Section 8.6).

Data from bicycle-sharing systems captures commuting behavior [FWH13], including detailed temporal records and GPS-tracked routes in some cases; and they are widely available from many cities throughout the world [AOS13, MHK18, RZ16, RMZ18, WB18]. These properties make it extremely valuable to analyze bicycle-sharing data increase understanding of urban flows and the properties of human commuting. Bicycle-sharing is used often for "last-mile" transportation, bridging the gap between public transportation and a final destination [GS16]; and insights into the dynamics and function of bicycle-sharing systems can help transit systems evolve to meet the needs of changing cities [NBI19].

In the present paper, we propose two models of temporal network connectivity to capture the functional roles of bicycle-sharing stations. We do this using the lens of mesoscale-structure detection in time-dependent networks [HS12, Hol15, FH16]. We examine trip histories from bicycle-sharing systems in the form of multilayer networks [KAB14, AM19, Por18] in which each layer is a network of trips in a given hour. Edges in each layer represent the total number of trips from one station to another that begin during that hour, and we do not include any interlayer edges.

We aim to partition a network based on a relational equivalence of nodes (a perspective with a rich history in the social-networks literature [LW71, RA15]), rather than on high internal traffic within sets of nodes [MHK18]. Data that has been aggregated over long periods of time can shed light on "community" structure in the latter sense [AOS13] through a partition of the network into contiguous spatial clusters [MHK18]. However, it ignores how bicycle-sharing usage cycles with travel patterns throughout a day [FWH13]. By contrast, our models are designed to detect functional roles of bicycle-sharing systems based on time-dependent behavior. The models that we introduce in this paper are time-dependent extensions of the stochastic block model (SBM) [SN97, NS01, KN11, Pei18]. We include parameters to describe intra-block and inter-block traffic for each hour, but we fix the block assignment of each station over time. That is, we treat a bicycle-sharing network as a temporal multilayer network with fixed node identities across layers [KAB14, AM19]. Although our models are inspired by the analysis of bicycle-sharing systems, they are applicable more generally to multilayer networks where nodes belong to fixed classes.

We introduce mixed-membership and discrete-membership versions of our model, where nodes can be members of multiple blocks or exactly one block, respectively. Both versions are degree-corrected, as we parametrize the time-aggregated degree of each node to avoid conflating block divisions with node activity levels [KN11]. This is especially important for bicycle-sharing networks in which stations have heterogeneous numbers of parking spaces for bicycles and different neighborhoods have different baseline levels of bicycle usage. Our

models are applicable to both directed or undirected networks, although we consider only directed examples in the present paper.

Increasing our understanding of the functional roles of docking stations can aid in the design and maintenance of bicycle-sharing networks, and understanding the usage patterns of stations can help inform the forecasting of their usage in unobserved and partially observed. This is valuable for tasks such as the dynamic restocking of stations with bicycles, which is both challenging and expensive, yet vital to the success of bicycle-sharing systems [SCL13a, CFS18].

A wide variety of community-detection and other clustering methods exist for networks. Such methods include spectral methods, inference using SBMs, optimization of objective functions (such as modularity [NG04]), local methods based on dynamical processes, and others [FH16]. They have different strengths and drawbacks, and some of them are more appropriate for some applications than for others [FH16, POM09, CPS19]. Community detection in multilayer networks is a developing field, with methods applied to diverse applications, such as biology, sociology, and materials science [BJA16, VFP15, CMM15, BDL17, KTB19, PPD18]. Several algorithms for community detection have been generalized to multilayer networks; see [MRM10, PC16, VMG16, SST16, BDL17, YCZ11, ZMN17, JMM17] for examples.

In early work on community detection in multilayer networks, Mucha et al. derived a generalization of modularity optimization for a type of multilayer network known as a "multislice" network [MRM10]. They used it to detect communities that change over time, encompass multiple types of social relationships, or include multiple values of a resolution parameter. Yang et al. introduced a discrete-membership SBM with time-evolving communities and parameters for block-to-block activity that are fixed over time [YCZ11]. Both [XH14] and [MM17] proposed related models (for unweighted (i.e., binary) and weighted networks, respectively), but they relaxed the assumption of fixed block-to-block activity parameters. Similar in spirit, Zhang et al. developed a time-dependent model with degree correction in

which nodes are allowed to switch between blocks which are described by a continuous-time Markov chain [ZMN17]. Mixed-membership SBMs with time-evolving communities have also been developed [XFS10, HSX11]. See [RC18] for a survey of community-detection methods for time-dependent networks. Based on the classification scheme in that paper, our methods belong to the class with "fixed memberships, evolving properties".

Community detection has been applied previously to urban bicycle sharing using various approaches [BAF11, AOS13, RAB09, MHK18, XW18, AMY18, KTB19, HGB19]. Zhu et al. applied k-means clustering to undirected, time-dependent usage data from bicycle-sharing systems and other urban systems in New York City [ZKK18]. Austwick et al. examined modularity optimization with a directed and spatially-corrected null model to identify communities of stations in several cities [AOS13]. However, they detected communities in time-aggregated data, and their discussion pointed out that there are significant limitations to examining community structure while ignoring time-dependent behavior for bicycle-sharing applications. Munoz-Mendez et al. [MHK18] identified communities by hour for bicycle-sharing data from London using an InfoMap algorithm [RAB09] that respects the directed nature of edges in the underlying trip networks. The changes that they discovered in communities over time highlight the importance of time of day in the usage of bicycle-sharing systems.

Closely related to our work, Matias et al. constructed a time-dependent, discrete-membership SBM with fixed blocks over time and applied it to bicycle-sharing networks in London [MRV18]. They detected some functional blocks, but their approach does not incorporate degree correction. Xie and Wang [XW18] employed an approach that does not use an SBM directly, yet they were able to successfully partition a bicycle-sharing network to find home and work roles of bicycle-sharing stations. They used the ratio of in-degree to out-degree at different times to discover home–work splits during peak commute times, similar to the results of our paper. A similarity of their approach to ours is that it corrects for degree; a key difference is that they relied on human supervision to determine peak hours, whereas

our models implicitly increase the weights of more-active time periods in the likelihood function. Etienne and Latifa clustered bicycle-sharing stations in Paris based on their time-dependent usage profiles using a Poisson mixture model [EL14]. They were able to capture time-dependent activity for each group, distinguish between incoming and outgoing activity, control for the overall activity level of a given station (via degree correction), and associate identified groups with their role in the city. A key difference between their approach and ours is that we distinguish activity between blocks, which allows us to detect behavior like last-mile commuting that occurs within blocks.

Our paper proceeds as follows. In Section 8.2, we list our data sources and present basic statistical analysis of the data. In Section 8.3, we introduce the two versions of our time-dependent SBM — discrete and mixed-membership — and we show that they are equivalent up to a constraint. In Section 8.4, we describe the estimation algorithms for our discrete and mixed-membership models. In Section 8.5, we present the results of our models for Los Angeles, San Francisco, and New York City. We discuss the implications of our work for bicycle-sharing systems and suggest areas of further study in Section 8.6. We show some additional details of our work in an appendix. We include code and data to implement our models and replicate the results in our paper as supplementary material (and also at `https://github.com/jcarlen/tdsbm_supplementary_material`).

8.2 Data

We examine United States bicycle-sharing systems in Los Angeles, the Bay Area, and New York City. For Los Angeles, we study only the system's downtown part, which is self-contained; similarly, in the Bay Area, we consider only the San Francisco network. Our three focal systems vary in size and daily usage. Because of this variation and how the data were reported, we study different time periods for each system. We also selected our time periods to exclude days that are likely to be extremely hot or cold. All of the bicycle-sharing

systems that we study have open-data portals, from which we downloaded the following data sets. In summary, after cleaning (see our discussion in the next paragraph), our data consist of the following:

- Downtown Los Angeles: October–December 2016; there are 61 stations and 40,130 trips, of which 73.4% are during weekdays [NYC17].
- San Francisco: September 2015–August 2016; there are 35 stations and 267,412 trips, of which 92.1% are during weekdays [Bay17].
- New York City: October 2016; there are 601 stations and 1,551,692 trips, of which 75.6% are during weekdays [LA 17].

A trip consists of a user checking out a bicycle from a fixed location (a station that includes multiple parking spaces) and returning it to a station. The data for each trip include the starting time; ending time; and starting and ending locations by station ID, latitude, and longitude. Each data set also has a few additional fields about the users; these details include whether they have memberships in the bicycle-sharing system, but we do not use this information in our investigation. We cleaned the data by removing anomalous trips, including extremely short and extremely long trips,[1] and trips to or from a station used for testing or maintenance (as indicated in the data). We also excluded a very small number of stations (two in Los Angeles and six in New York City) that did not have at least one departure and at least one arrival during the given time period. Finally, we excluded one station in New York City that was accessible to other stations only by ferry; it was involved in only nine trips during the given time period. In total, cleaning removed 7.1 % of the trips in Los Angeles, 4.5 % of the trips in San Francisco, and 1.4 % of the trips in New York City. We retain self-edges, which represent trips that start and end at the same

[1]We take extremely short trips to be those that last two minutes or less; we take extremely long trips to be those that last 90 minutes or more in Los Angeles and San Francisco and 120 minutes or more in New York City, given that city's larger coverage area of stations.

station. Although it is common to remove self-edges when analyzing networks [New18], self-edges in the present application represent real trips that we expect to have a very similar data-generating mechanism as trips from stations to geographically nearby stations. As in [KN11], including self-edges also simplifies some elements of parameter estimation.

When fitting our models, we include only weekday trips, as we observe that weekday and weekend activity follow distinct patterns; and weekend activity does not reflect commuting behavior. From the data sets, we construct multilayer networks that are both weighted and directed. In our networks, nodes represent stations, edge values encode the number of directed trips from one station to another that begins in a specified time period, and each of the 24 layers consists of the trips that start in a certain hour.

8.2.1 Preliminary Data Analysis

Our data show clear patterns of bicycle-sharing usage by time of day and day of the week, including heavier use during commuting hours. In Figure 8.2.1, we illustrate usage patterns by plotting the number of trips by starting hour for each city. In New York City and San Francisco, activity spikes during weekday morning and evening commuting hours, whereas weekend trips peak in early afternoon. Similar patterns were observed previously for bicycle-sharing systems in New York and many other cities [AOS13, EL14, ZKK18, XW18, Tai14]. By contrast, in Los Angeles, the number of trips by hour has a mid-day peak on weekdays that is nearly as strong as the morning peak. In New York City and Los Angeles, about one quarter of the trips occur on weekends, but only about eight percent of the trips in San Francisco occur then. This suggests that the San Francisco network covers more commercial areas and fewer residential areas.

For individual stations, the morning and evening peaks for in-degree and out-degree are often unbalanced: one direction has a stronger morning peak, and the opposite direction has a stronger evening peak. This is a key motivation for our time-dependent identification of stations into “home" and “work" types.

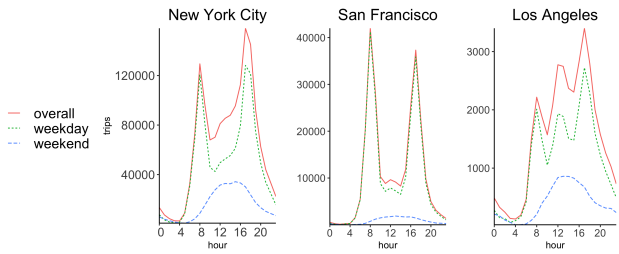


Figure 8.2.1: Total trips by hour for weekdays, weekends, and overall. Hour 0 designates midnight.

To further explore the imbalance between morning and evening activity in each network, we calculate the singular-value decomposition (SVD) of the matrices of in-degree and out-degree for each station by hour. To be explicit, entry i, j of the matrix of in-degrees is equal to the total number of trips that arrive at station i in hour j, and the we constructed the matrix of out-degrees analogously for departing trips. We show results for New York City in Figure 8.2.2 and for Los Angeles (see Figure 8.7.1) and San Francisco (see Figure 8.7.2) in the appendix. The first two principle components either strengthen both observed peaks or weaken one peak while strengthening the other. The first two singular vectors explain at least 88% and as much as 97% percent of the variation in the corresponding matrix, supporting the importance of peak morning and evening commutes for classifying stations.

Another characteristic of our data that we incorporated into the design of our models is the strong positive Pearson correlation coefficient between the total (summed over all time periods) in-degree and out-degree of each station: 0.99 in New York, 0.98 in San Francisco, and 0.91 in Los Angeles. This is an intrinsic feature of docked bicycle-sharing systems, because a bicycle must be returned to a station before a new trip with it can begin. However, the use of trucks to redistribute bicycles in a system can loosen this requirement.

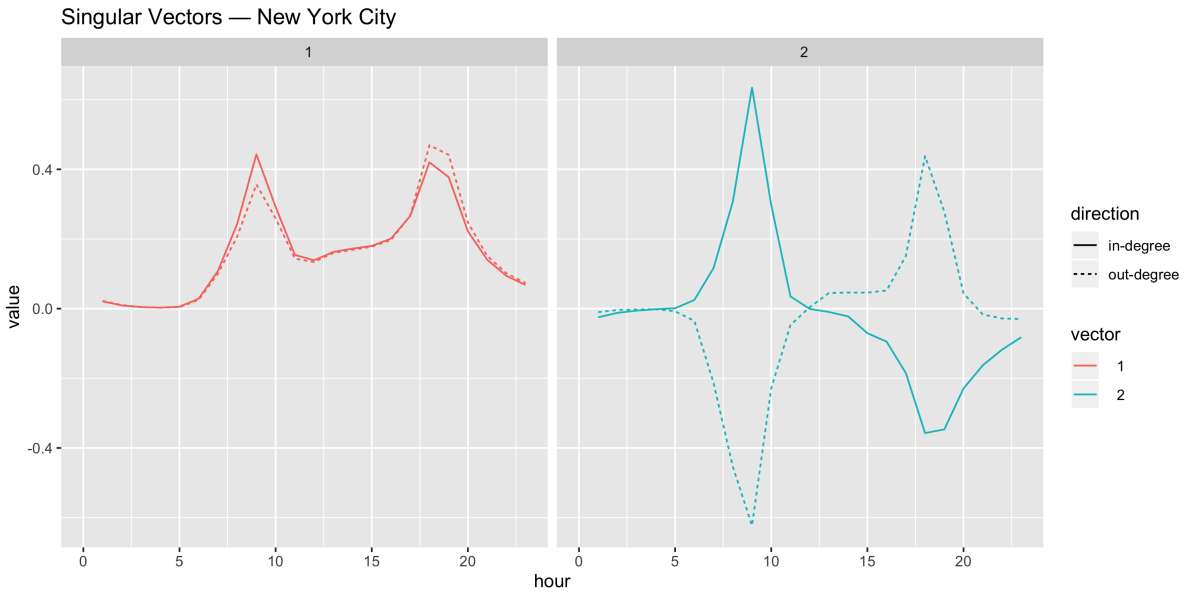


Figure 8.2.2: The first two singular vectors from the New York City bicycle-sharing network.

8.3 Our Stochastic Block Models

In this section, we introduce our time-dependent mixed-membership stochastic block model (TDMM-SBM) and time-dependent discrete stochastic block model (TDD-SBM).

Stochastic block models are a popular class of statistical network model [Pei18]. The motivating principle of SBMs is a type of *stochastic equivalence* in which edges whose endpoints have the same block membership are identically distributed. It is a standard assumption of SBMs that edge values are independent, given the block membership of nodes. More formally, a binary (Bernoulli) random graph Y, with adjacency matrix A, from an SBM with K blocks is defined by

$$P(A_{ij} = 1|G) = \eta_{g_i g_j} , \tag{8.3.1}$$

where G (with components $g_i \in \{1, 2, \ldots, K\}$) is a vector of block assignments for the nodes of Y and η is a $K \times K$ matrix of block-to-block edge probabilities. Note that this definition allows directed graphs, in which η can be asymmetric. For early presentations of SBMs,

see [HLL83, FW92, SN97, NS01]. More recent advances have added flexibility to SBMs. Examples include the mixed-membership SBMs of [ABF08], models with covariates in [IL15], the degree-corrected SBM of [KN11], and the Bayesian implementations reviewed by [Pei18]. Applications of SBM to longitudinal networks include discrete-membership [YCZ11, ZMN17] and mixed-membership [XFS10, HSX11] versions where nodes can switch blocks over time, as well discrete-membership versions with fixed blocks over time but without degree-correction [MRV18, XH14].

8.3.1 Time-Dependent Mixed-Membership Stochastic Block Model (TDMM-SBM)

We now describe the framework for our mixed-membership SBM. Let $i, j \in \mathcal{N}$ (with $|\mathcal{N}| = N$) be nodes, which represent bicycle stations; let $g, h \in \mathcal{K}$ (with $|\mathcal{K}| = K$) be blocks. Our data is a three-dimensional array of size (N, N, T), where T is the number of time slices (i.e., time layers). We consider hourly groupings of the trips based on their starting times. The quantity A_{ijt} is the observed number of trips from station i to station j with starting time greater than or equal to t and less than $t + 1$. Let $\tilde{A}_{ij} = \sum_{t=0}^{23} A_{ijt}$ denote the weights of the associated time-aggregated matrix. Our network is a directed multilayer network, so we count each trip that both starts and ends at a node i during hour t (i.e., self-edges) exactly once in A_{iit}.

For each node i, there is a length-K vector of real numbers $C_{ig} \in [0, 1]$. These numbers represent the mixed-membership block assignment of each node. The block-assignment parameter C_{ig} indicates the "strength" of node i in block g. For each ordered pair g, h of blocks and each time $t \in \{0, 1, \ldots, 23\}$ (where $t = 0$ represents the hour that starts at midnight), there is a parameter ω_{ght}, which we call the "inter-block connectivity" parameter or "block-to-block" parameter, that represents the directed activity from block g to block h during hour t. Note that ω_{ght} need not be equal to ω_{hgt} if the network is directed; this captures any asymmetries in the number of trips with respect to reversing origins and destinations.

We also define the notation $\tilde{\omega}_{gh} = \sum_{t=0}^{23} \omega_{ght}$ for the time-aggregated matrix.

For each pair of nodes, i and j, we assume that the number of trips that depart from i and arrive at j at time t is Poisson-distributed with mean $\mu_{ijt} = \sum_{g,h} C_{ig}\omega_{ght}C_{jh}$. Our use of the Poisson distribution follows [KN11] and [Pei18], facilitates computation, and is standard for modeling count data (although overdispersion is a concern).

For identifiability, we apply the constraint $\sum_i C_{ig} = 1$ for all g. This does not constrain the set of possible models in terms of realizable mean edge activities μ_{ijt}. Consider a model with unconstrained parameters ω_{ght} and C_{ig}. The model with parameters ω'_{ght} and C'_{ig} such that $C'_{ig} = \frac{C_{ig}}{\sum_j C_{jg}}$ and $\omega'_{ght} = \omega_{ght}\left(\sum_j C_{jg}\right)\left(\sum_j C_{jh}\right)$ is an equivalent model, because the means of the distributions of edge weights are equal to the those of the model with unconstrained parameters. That is, $\mu'_{ijt} = \sum_{g,h} C'_{ig}\omega'_{ght}C'_{jh} = \sum_{g,h} C_{ig}\omega_{ght}C_{jh} = \mu_{ijt}$. Because $\sum_i C_{ig} = 1$, we can think of C_{ig} as the proportion of the total activity of block g from the activity of node i; the expected total number of trips at node i is $\sum_g C_{ig} \sum_{h,t}(\omega_{ght}+\omega_{hgt})$. In this light, $\sum_g C_{ig}$ is a measure of the activity of node i in which we do not weight each C_{ig} term by the total activity of the corresponding block. We can interpret C_{ig} relative to C_{ih} as how strongly block g is associated with node i relative to how strongly block h is associated with node i. We use these quantities when visualizing the TDMM-SBM of our data, because they help ensure that we do not overlook blocks with important usage patterns but relatively lower activity. The parameter C_{ig} is analogous to the degree-correction parameter for SBMs that was introduced in [KN11], but we apply it to mixed block membership. We elaborate on this connection in Subsection 8.3.2, where we introduce a model that specifies that nodes have only one block.

We now compute the likelihood function that we will optimize to obtain the maximum-likelihood estimate (MLE). We assume conditional independence between hourly numbers

of trips along each edge, given model parameters, so the likelihood of the data is

$$L(G;\omega,\mathbf{C}) = \prod_{t=0}^{23} \prod_{i,j\in\mathcal{N}} \frac{(\mu_{ijt})^{A_{ijt}}}{A_{ijt}!} \exp\left(-\mu_{ijt}\right) , \tag{8.3.2}$$

where ω and $\mathbf{C}$ give the model parameters (i.e., $\omega = \{\omega_{ght}\}$ and $\mathbf{C} = \{C_{ig}\}$). Note that $\mu_{ijt} = \sum_{g,h} C_{ig}\omega_{ght}C_{jh}$ is a function of these parameters, the set $\mathcal{N}$ of nodes in the network is fixed and pre-determined, and the number K of blocks is also fixed and pre-determined. The unnormalized log-likelihood is

$$\ell(G;\omega,\mathbf{C}) = \sum_{t=0}^{23} \sum_{i,j\in\mathcal{N}} \left[A_{ijt}\log\left(\mu_{ijt}\right) - \mu_{ijt}\right] , \tag{8.3.3}$$

although note that we omit the addition of the constant $-\sum_{i,j,t}\log\left(A_{ijt}!\right)$, because it does not affect maximum of the function.

8.3.2 Time-Dependent Discrete Stochastic Block Model (TDD-SBM)

We derive a single-membership SBM from our mixed-membership SBM by making the extra assumption that, for each node $i \in \mathcal{N}$, we have that $C_{ig} > 0$ for only one block $g \in \mathcal{K}$. (We also call this the "discrete version" of our model.) For our single-membership SBM, we introduce some new notation to aid our description and be consistent with notation in [KN11, ZYM13]. For a given node i, the block g for which $C_{ig} > 0$ is the block g_i that includes node i. Therefore, we use a single parameter $\theta_i = C_{ig_i}$ for each node i to indicate both the strength of i in block g and the membership of node i in block g. We will show that this term is a multilayer extension of the degree-correction term of [KN11]. The mean of the Poisson distribution of the value of an edge from node i to node j at time t is $\theta_i\theta_j\omega_{g_ig_jt} = C_{ig_i}\omega_{g_i,g_jt}C_{jg_j}$. We retain the sum constraints of our mixed-membership model, such that $\sum_{i\in g}\theta_i = 1$ for all g.

We compute optimal values for the parameters ω and $\theta = \{\theta_i\}_{i\in\mathcal{N}}$. As in the TDMM-

SBM, take $\mathcal{N}$ and K to be fixed and pre-determined. Again dropping the constant term $-\sum_{i,j,t} \log(A_{ijt}!)$, the log-likelihood of our single-membership SBM is

$$\ell(G;\omega,\theta) = \sum_t \sum_{g,h} \sum_{i\in g, j\in h} \left[A_{ijt}\log\theta_i + A_{jit}\log\theta_j + A_{ijt}\log\omega_{ght} - \theta_i\theta_j\omega_{ght}\right] .$$

We find explicit formulas for the MLEs of θ_i and ω_{ght}. In the following calculations, removal of t from the subscript of a parameter and addition of a tilde designates a sum over all t. Specifically, we define $\tilde{A}_{ij} = \sum_{t=0}^{23} A_{ijt}$ and $\tilde{\omega}_{gh} = \sum_{t=0}^{23} \omega_{ght}$. We differentiate ℓ with respect to ω_{ght} to yield

$$\frac{\partial}{\partial\omega_{ght}}\ell = \frac{\sum_{i\in g, j\in h} A_{ijt}}{\omega_{ght}} - 1 ,$$

where we have used the block-wise sum constraints on θ_i. Therefore, the MLE for w_{ght} is

$$\hat{\omega}_{ght} = m_{ght} ,$$

where m_{ght} is the sum of weights of edges from nodes in block g to nodes in block h during hour t. That is, $m_{ght} = \sum_{i\in g, j\in h} A_{ijt}$.

We then differentiate ℓ with respect to θ_i to obtain

$$\frac{\partial}{\partial\theta_i}\ell = \frac{\sum_j \tilde{A}_{ij} + \sum_j \tilde{A}_{ji}}{\theta_i} - \sum_h \tilde{\omega}_{g_i h} - \sum_h \tilde{\omega}_{h g_i} .$$

At $\hat{\omega}_{ght}$, the MLE for θ_i is

$$\hat{\theta}_i = \frac{\sum_j \tilde{A}_{ij} + \tilde{A}_{ji}}{\sum_g \tilde{m}_{gh} + \tilde{m}_{hg}} = \frac{k_i}{\kappa_{g_i}} ,$$

where $k_i = \sum_j \left(\tilde{A}_{ij} + \tilde{A}_{ji}\right)$ is the sum of the in-degree and out-degree of i over all time periods, $\tilde{m}_{gh} = \sum_{t=0}^{23} m_{ght}$, and $\kappa_g = \sum_h (\tilde{m}_{gh} + \tilde{m}_{hg})$ is the sum of the in-degrees and out-

degrees of all nodes in block g over all time periods. The term $2\tilde{m}_{gg}$ in the equation for κ_g implies that we count each intra-block edge twice: once for emanating from g and once for arriving at g. Similarly, k_i includes the term $2\tilde{A}_{ii}$, so we count self-edges twice in this term.

Our computation demonstrates that the MLE of the strength of i in block g is the relative proportion of the strength of node i to the total activity of block g. The parameter $\hat{C}_{ig_i} = \hat{\theta}_i$ in the TDD-SBM for modeling directed, multilayer networks is analogous to the degree-correction parameter in the degree-corrected SBM [KN11] for undirected networks with one layer. Indeed, the MLE for the degree correction parameter in the latter model is the proportion of number of edges connected to a node to the number of edges connected to its assigned block. Another similarity between a degree-corrected SBM and our TDD-SBM is that in the MLE of the TDD-SBM, the sum over time of the expectation of the degree of a node i is equal to the degree of node i from the observed data. That is, $\sum_t \sum_j (\mu_{ijt} + \mu_{jit})$, the sum of the mean weights of edges connected to i, is equal to k_i, the degree of node i in the observed data. (See Section 8.7.2 of the appendix for the proof.) For our mixed-membership SBM, we are not aware of such a precise relationship between the data and the expected value of model statistics, although there does appear to be a positive correlation between the time-aggregated node degrees and the sum of the mixed-membership parameters ($\sum_g C_{ig}$ for all i).

We now calculate the MLE of the unnormalized log-likelihood of the TDD-SBM. We obtain

$$\begin{aligned}\sum_t \left[\sum_{i,j} \left(A_{ijt} \log\left(\frac{k_i}{\kappa_{g_i}}\right) + A_{jit} \log\left(\frac{k_j}{\kappa_{g_j}}\right)\right) + \sum_{g,h} m_{ght} \log m_{ght} - \sum_{g,h} m_{ght} \right] \\ = \sum_i k_i \log k_i - \sum_i k_i \log \kappa_{g_i} + \sum_t \sum_{g,h} m_{ght} \log m_{ght} - \tilde{m}\,,\end{aligned}$$

where $\tilde{m}$ is the total number of edges in the network. By a similar calculation as one in

[KN11], we obtain

$$\begin{aligned}
\sum_i k_i \log \kappa_{g_i} &= \sum_t \sum_g \sum_{i \in g} k_{it} \log \kappa_g \\
&= \sum_t \sum_g \sum_{i \in g} (k_{\text{in},it} \log \kappa_g + k_{\text{out},it} \log \kappa_g) \\
&= \sum_t \sum_g \kappa_{\text{out},gt} \log \kappa_g + \sum_t \sum_h \kappa_{\text{in},ht} \log \kappa_h \\
&= \sum_t \sum_g \sum_h m_{ght} \log \kappa_g + \sum_t \sum_g \sum_h m_{ght} \log \kappa_h \\
&= \sum_t \sum_{g,h} m_{ght} \log \kappa_g \kappa_h \,,
\end{aligned}$$

where $k_{\text{in},it}$ and $k_{\text{out},it}$ are the respective in-degrees and out-degrees of node i during hour t, the quantity $\kappa_{\text{in},gt} = \sum_{i \in g} k_{\text{in},it}$ is the number of edges going into g during hour t, and $\kappa_{\text{out},gt} = \sum_{i \in g} k_{\text{out},it}$ is the number of edges that emanate from g during hour t. Including only the terms that depend on block assignments yields the following objective function:

$$\sum_t \sum_{g,h} m_{ght} \log \left(\frac{m_{ght}}{\kappa_g \kappa_h} \right) . \tag{8.3.4}$$

Unlike the directed SBM in [ZYM13], we do not have two strength parameters (representing an in-degree strength θ_i^{in} and an out-degree strength θ_i^{out}) for each station. Nevertheless, our model still captures the directed nature of the data. We can see this by conceptualizing the estimated means as an approximation to the number of trips by hour in both directions at the same time. By representing the 48-dimensional vectors as 2×24 matrices, we see that

$$\theta_i \theta_j \begin{bmatrix} \omega_{g_i g_j 0} & \omega_{g_i g_j 1} & \dots & \omega_{g_i g_j 23} \\ \omega_{g_j g_i 0} & \omega_{g_j g_i 1} & \dots & \omega_{g_j g_i 23} \end{bmatrix} \qquad \text{approximates} \qquad \begin{bmatrix} A_{ij0} & A_{ij1} & \dots & A_{ij23} \\ A_{ji0} & A_{ji1} & \dots & A_{ji23} \end{bmatrix} .$$

This perspective also holds for our mixed-membership SBM, for which

$$\sum_{g,h} C_{ig}C_{jh} \begin{bmatrix} \omega_{gh0} & \omega_{gh1} & \dots & \omega_{gh23} \\ \omega_{hg0} & \omega_{hg1} & \dots & \omega_{hg23} \end{bmatrix} \qquad \text{approximates} \qquad \begin{bmatrix} A_{ij0} & A_{ij1} & \dots & A_{ij23} \\ A_{ji0} & A_{ji1} & \dots & A_{ji23} \end{bmatrix}.$$

The validity of this matrix representation depends on there being a large correlation between the time-aggregated in-degrees and out-degrees of nodes. This is related to the fact that in a given 24-hour period, the number of trips from one station to another is predictive of the number of trips in the opposite direction. This matrix representation is related to the fact that the 24-hour time activities for trips between two stations in one direction are predictive of the activities in the opposite direction. The latter observation, in turn, is related to the axiom of human mobility that for each current of travel, there is a counter current [BBG18]. We observe (as did [ZGW18]), from the above matrix expressions, that maximizing the log-likelihood of both SBMs is equivalent to a form of nonnegative matrix factorization with K^2 48-dimensional basis columns. In a sense, our model is neither an extension of the usual undirected degree-corrected SBM nor one of the usual directed degree-corrected SBMs. Instead, our model's single-layer network analog is a degree-corrected SBM with parameters θ_i and $\tilde{\omega}_{gh}$, except that the $\tilde{\omega}_{gh}$ are not constrained to be symmetric.

8.4 Computations

In this section, we describe the algorithms that we use for both the TDMM-SBM and the TDD-SBM.

8.4.1 Inference using the TDMM-SBM

Let $\omega = \{\omega_{ght}\}$ be the $K \times K \times T$ array that represents the inter-block connectivity parameters, and let $\mathbf{C} = \{C_{ig}\}$ be the matrix that represents the collection of node-strength

parameters. We estimate the model parameters using a two-step gradient descent.[2] First, we move in the direction of the gradient with respect to ω and update the inter-block connectivity parameters. Second, we move along the direction of the gradient with respect to $\mathbf{C}$ and update the node-strength parameters.

In the description of our algorithm, we let $\omega^{(n)}$ and $\mathbf{C}^{(n)}$, respectively, denote the n^{th} update of the inter-block connectivity and node-strength parameters. We initialize the algorithm with random values $\omega^{(0)}$ and $\mathbf{C}^{(0)}$ with components distributed according to $\exp(X)$, where X is a Gaussian random variable with mean 0 and variance 1. (That is, we draw random values from a log-normal distribution.) We denote the mean activity along edge (i, j) with initial parameters $\omega^{(0)}$ and $\mathbf{C}^{(0)}$ by $\mu_{ijt}^{(0)}$. We scale the parameters so that the TDMM-SBM at the starting point of the optimization has the same mean number of trips as the data. Specifically, we multiply the inter-block connectivities $\omega_{ght}^{(0)}$ by $\left(\sum_t \omega_{ght}^{(0)}\right)^{-1} \left(\sum_{i,j,t} A_{ijt}\right) / K^2$ and normalize $\mathbf{C}^0$ to satisfy the constraint $\sum_i C_{ig}^{(0)} = 1$ for each block g. This results in $\sum_{ijt} \mu_{ijt}^{(0)} = \sum_{i,j,t} \sum_{g,h} C_{ig}^{(0)} \omega_{ght}^{(0)} C_{jh}^{(0)} = \sum_{g,h,t} \omega_{ght}^{(0)} = \sum_{g,h} \left(\sum_{i,j,t} A_{ijt}\right) / K^2 = \sum_{i,j,t} A_{ijt}$, the total number of edges in the network. Without this scaling, the initial parameters would have very small magnitudes, such that the mean total number of trips from the TDMM-SBM with these initial parameters is much smaller than the total number of trips in the data. Therefore, it is very likely that the randomly chosen initial parameters will have very small log-likelihoods relative to the MLE log likelihood. Early gradient-descent steps might then dramatically increase the magnitude of the parameters while affecting the relative sizes of individual parameters in unpredictable ways.

To ensure that our estimated parameters are nonnegative, we use the following change of variables: $\exp(\tilde{\omega}^{(n)}) = \omega^{(n)}$ and $\exp(\tilde{\mathbf{C}}^{(n)}) = \mathbf{C}^{(n)}$. We can then write the gradient descent

[2] Although we are maximizing a function and thus technically performing gradient ascent, we refer to this class of method by its more common monicker of "gradient descent".

as

$$\tilde{\omega}^{(n+1)} = \tilde{\omega}^{(n)} + \eta^{(n)} \nabla_{\omega} \ell(\mathbf{C}^{(n)}, \omega^{(n)}) \exp(\tilde{\omega}^{(n)}) \,,$$
$$\tilde{\mathbf{C}}^{(n+1)} = \tilde{\mathbf{C}}^{(n)} + h^{(n)} \nabla_{\mathbf{C}} \ell(\mathbf{C}^{(n)}, \omega^{(n+1)}) \exp(\tilde{\mathbf{C}}^{(n)}) \,,$$

where $h^{(n)}$ and $\eta^{(n)}$ are small positive numbers. From the definitions of $\tilde{C}^{(n)}$ and $\tilde{\omega}^{(n)}$, we write

$$\omega^{(n+1)} = \omega^{(n)} \exp(\eta^{(n)} \nabla_{\omega} \ell(\mathbf{C}^{(n)}, \omega^{(n)}) \omega^{(n)})$$
$$\mathbf{C}^{(n+1)} = \mathbf{C}^{(n)} \exp(h^{(n)} \nabla_{\mathbf{C}} \ell(\mathbf{C}^{(n)}, \omega^{(n+1)}) \mathbf{C}^{(n)}) \,.$$

We take the exponential of a vector to be the result of applying the exponential to each component of the vector. Let $h^{(0)} = \eta^{(0)} = \Delta > 0$ be the fixed initial step size. For our application, we choose $\Delta = 10^{-4}$. We generate two candidate updates for $\omega^{(n+1)}$ for the first step in our algorithm using $h^{(n+1)} = 1.2\, h^{(n)}$ and $h^{(n+1)} = .8\, h^{(n)}$, and we choose the one that gives a $\omega^{(n+1)}$ that yields the larger value of $\ell(\mathbf{C}^{(n)}, \omega^{(n+1)})$. Similarly, we choose the one of $\eta^{(n+1)} = 1.2\, \eta^{(n)}$ or $\eta^{(n+1)} = .8\, \eta^{(n)}$ that gives the $\mathbf{C}^{(n+1)}$ with the larger $\ell(\mathbf{C}^{(n+1)}, \omega^{(n+1)})$.

We compute the gradient of the log-likelihood function ℓ using the chain rule. Recall that we compute the log-likelihood in two parts. One part is the computation of the mean $\mu_{ijt} = \sum_{g,h}^{K} C_{ig} \omega_{gh}^{t} C_{jh}$ of the number of trips from node i to node j at time t. We then insert the expression for the mean into the function $\ell = \sum_t \sum_{i,j} (A_{ijt} \log(\mu_{ijt}) - \mu_{ijt})$. We compute the derivative of ℓ with respect to μ_{ijt} to obtain

$$\frac{\partial \ell}{\partial \mu_{ijt}} = \frac{A_{ijt}}{\mu_{ijt}} - 1 \,.$$

The derivative of μ_{ijt} with respect to C_{kg} is

$$\frac{\partial \mu_{ijt}}{\partial C_{kg}} = \delta_{ki} \sum_{h} \omega_{ght} C_{jh} + \delta_{kj} \sum_{h} C_{ih} \omega_{hgt} \,,$$

and the derivative of μ_{ijt} with respect to ω_{ght} is

$$\frac{\partial \mu_{ijt}}{\partial \omega_{ght}} = C_{ig} C_{jh} \, .$$

Here δ_{ab} is the kronecker delta (i.e. $\delta_{ab} = 1$ if $a = b$ and $\delta_{ab} = 0$ if $a \neq b$). Using the above calculations, we see that the derivatives of ℓ with respect to C_{kg} and ω_{ght} are

$$\begin{aligned}
\frac{\partial \ell}{\partial C_{kg}} &= \sum_{t=0}^{23} \sum_{i,j \in \mathcal{N}} \frac{\partial \ell}{\partial \mu_{ijt}} \frac{\partial \mu_{ijt}}{\partial C_{kg}} \\
&= \sum_{t=0}^{23} \left(\sum_{j \in \mathcal{N}} \left(\frac{A_{kjt}}{\mu_{kjt}} - 1 \right) \sum_{h} \omega_{ght} C_{jh} + \sum_{i \in \mathcal{N}} \left(\frac{A_{ikt}}{\mu_{ikt}} - 1 \right) \sum_{h} C_{ih} \omega_{hgt} \right) \, , \\
\frac{\partial \ell}{\partial \omega_{ght}} &= \sum_{i,j \in \mathcal{N}} \frac{\partial \ell}{\partial \mu_{ijt}} \frac{\partial \mu_{ijt}}{\partial \omega_{ght}} \\
&= \sum_{i,j \in \mathcal{N}} \left(\frac{A_{ijt}}{\mu_{ijt}} - 1 \right) C_{ig} C_{jh} \, .
\end{aligned}$$

We run the gradient descent until four significant digits of the base-10 floating-point representation of the log-likelihood (8.3.3) does not change for 600 steps in a row. For the networks that we examine, this usually takes between 600 and 5000 iterations, with models with more blocks generally needing more iterations to reach this stopping criterion. Because of the non-convexity of the log-likelihood function (8.3.3), we are not guaranteed to reach a global optimum. Most of the time, our method converges to an interesting local optimum (which may also be a global optimum), revealing the existence of functional roles (see Section 8.5). Our results produce recognizable inter-block connectivity parameters ω_{ght} (e.g., home–work commute patterns and leisure-usage patterns) and the parameters C_{ig} indicate known spatial divisions of the stations (e.g., residential versus commercial districts). In some cases, however, our algorithm converges to an uninteresting local optimum; one example is when the block-assignment parameters C_{ig} for each station appear as if they are assigned independently at random. To improve our results, we run our algorithm repeatedly

(specifically, 10 times for each network) and store the estimate with the largest likelihood. We compare the parameters that we obtain from gradient descent versus those that we obtain by running a Hamiltonian Monte Carlo (HMC) sampling method in a Bayesian framework with weak priors (implemented in Stan [CGH17, Sta18]). The log-likelihoods that result from our gradient-descent method are as good or better than those that we obtain with an HMC method. The HMC method is more computationally and memory intensive than our gradient-descent method, although it may be preferable in applications in which one has meaningful prior information about parameters. Improving our optimization method and investigating trade-offs between accuracy and efficiency are worthwhile topics for future work. For instance, it will likely be beneficial to adapt optimization methods for related time-dependent SBMs [XFS10, YCZ11, HSX11, XH14, MRV18] to the optimization of our model.

The supplementary material has our Python implementation of our gradient-descent method, as well as code for our inference in R using Stan.

8.4.2 Inference using the TDD-SBM

To fit our TDD-SBM, we use a Kernighan–Lin-type (KL-type) algorithm [KL70] that we base on the one in [KN11]. Given a number K of blocks, we initialize the algorithm by assigning each node to a block uniformly at random, so each node has a probability of $\frac{1}{K}$ of being assigned to a given block. The algorithm then calculates the best possible block reassignment for any single node with respect to the associated change in log-likelihood (either the largest increase or the smallest decrease). Subsequently, we make the best reassignment for a different node, again chosen uniformly at random, with respect to change in log-likelihood. The algorithm cycles through all nodes; a key feature of the algorithm is that a node that has been reassigned cannot move again until all other nodes have moved. One set of sequential reassignment of all nodes constitutes one step of the algorithm. The algorithm then searches all of the states (with respect to block membership of nodes) that

have occurred during the step, and it selects the state with the maximum log-likelihood of any during the step. This state is the starting point for the next step of the algorithm. A single run of the algorithm is completed when a step does not increase the log-likelihood beyond a preset tolerance value near 0. (In practice, we use 1×10^{-4}.) To find block assignments that are as good as possible, we do many runs of the algorithm for each network. In our examples, we use 50 runs per network. We initialize each run randomly, as described above.

Another key feature of the algorithm is that changes in the block membership of nodes affect only the terms of the objective function that involve the origin and destination blocks of the change. (We see in (8.3.4) that the objective is a sum over block-pair terms over T time slices.) Consequently, we do not need to recalculate the full objective function at each step.

We implement our KL-type algorithm for TDD-SBM in R using **Rcpp** [R C18, EF11]. The back-end calculations are in c++ for speed, and we return results in R to enable visualization and other analyses. Our implementation can also estimate time-independent SBMs, including directed and/or degree-corrected ones. This facilitates comparison of the results of inference from time-dependent and time-independent SBMs. See `https://github.com/jcarlen/sbm` for our R package `sbmt` for parameter estimation for the TDD-SBM. We include code (which uses the `sbmt` package) in supplementary material to replicate our examples in Section 8.5.

8.5 Results

We apply our models to bicycle-sharing networks in Los Angeles, San Francisco, and New York City. (See Section 8.2 for descriptions of these data sets.) The networks that we examine in downtown Los Angeles and San Francisco are relatively small, with 61 stations and 35 stations, respectively, at the time that we collected our data. The stations in these networks are concentrated in downtown areas, where high-rise office and residential buildings

are interspersed. The New York City network is much larger than the other two. It includes about 600 stations at the time of data collection. The stations span most of the lower half of Manhattan and northwestern Brooklyn. They encompass a range of commercial areas, residential neighborhoods, parks, and manufacturing areas.

8.5.1 Downtown Los Angeles

In Figure 8.5.1, we show the mixed-membership (TDMM-SBM) and discrete (TDD-SBM) block assignments of two-block models of the downtown Los Angeles system. For the TDMM-SBM, the we scale the size of a given node i in our plots based on $\sum_g C_{ig}$. We refer to these sums as "C total" values. These values correlate strongly with node degree (specifically, the sum of in-degree and out-degree), which is evident in the similarity of node sizes in the left and right panels of Figure 8.5.1. For both models, we observe that home and work blocks are interspersed geographically. (We will soon describe our method for determining the block labels in Figure 8.5.1.) The TDMM-SBM result reveals a group of stations (which we color in gray) in the left panel of Figure 8.5.1 are neither strongly home-identified nor strongly work-identified; instead, they possess a roughly even mixture of the two types. For this network, the TDD-SBM output is very similar to what we obtain from a discretization of the TDMM-SBM output (which we discretize by assigning each node i to the block with the maximum value of its C_{ig} parameter), but this is not true for all of our bicycle-sharing networks.

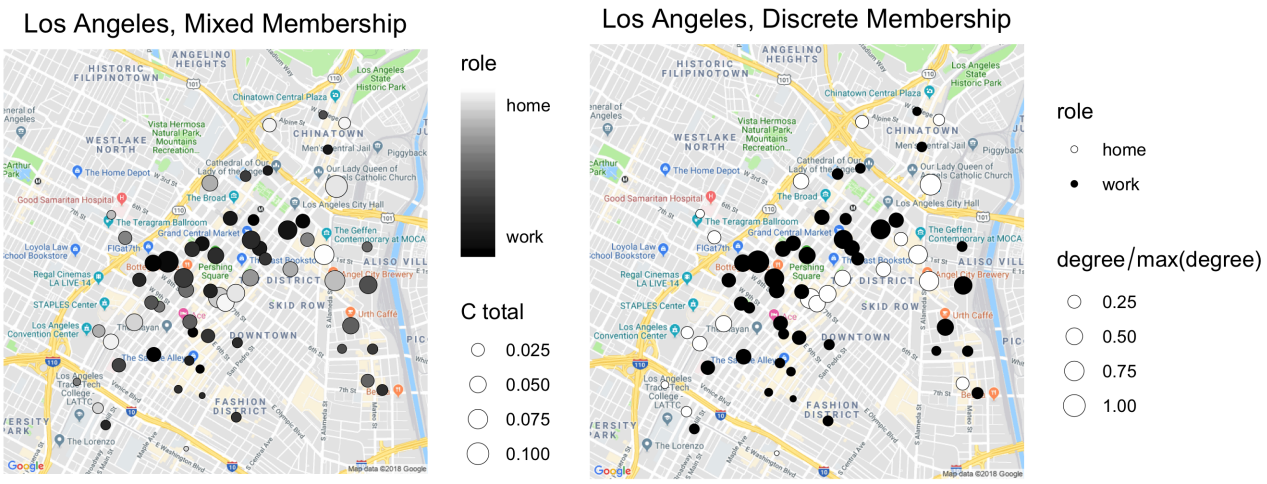


Figure 8.5.1: Downtown Los Angeles bicycle stations classified using (left) a two-block TDMM-SBM and (right) a two-block TDD-SBM. The sizes of the nodes take continuous values. In the left panel, we scale their area based on the value of $\sum_g C_{ig}$; in the right panel, we scale them based on the sum of the in-degree and out-degree (divided by the maximum value of that sum).

Our model does not yield "home" and "work" labels for each block on its own, so we use the time-dependent block-to-block parameter estimates $\hat{\omega}_{ght}$ to assign these labels. We assign the labels heuristically under the assumption that the "home" block is the origin of many trips to the work block in the morning and the "work" block is the origin of many trips to the home block in the evening. Figure 8.5.2, which shows $\hat{\omega}_{ght}$ for each possible value of g and h, with the hour t on the horizontal axis, supports our labeling. Based on our labeling, we observe a clear peak in home-to-work traffic in the mornings and work-to-home traffic in the evenings. We make similar "home" and "work" assignments for San Francisco and New York City. In Los Angeles, the traffic in the work block peaks in the middle of the day. This perhaps represents lunchtime errands, leisure activity, or tourist activity, as there are many tourist attractions in the downtown area. The traffic in the home block has a mild evening peak and has by far the least activity overall.

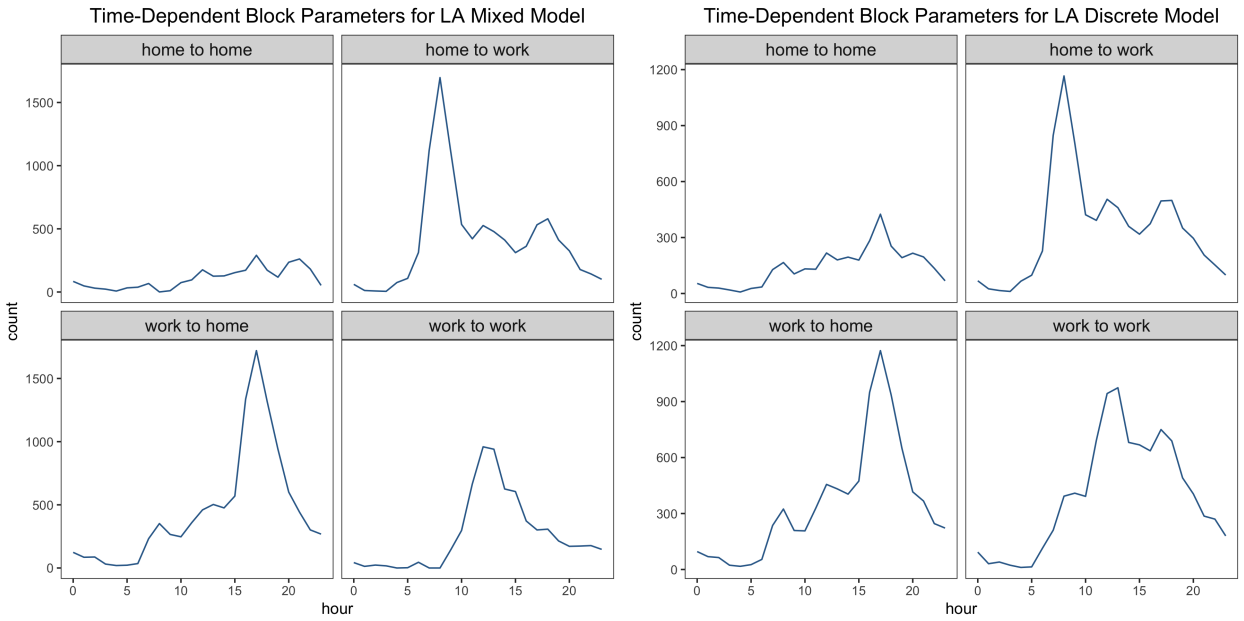


Figure 8.5.2: Estimated time-dependent block-to-block parameters $\hat{\omega}_{ght}$ for the two-block TDMM-SBM and two-block TDD-SBM for downtown Los Angeles.

To further validate our block labels, we use the zoning map for downtown Los Angeles from the Los Angeles Department of City Planning [Dep15].[3] Zoning ordinances determine the allowable uses of city land. They distinguish land that is available for commercial uses, industrial uses, residential uses, park districts, and others. In the background of Figure 8.5.3, we show a simplified version of the underlying zoning map (with a grouping of similar designations). The industrial areas house a mixture of manufacturing and commercial uses. Public facilities include government buildings, public schools, parking under freeways, and police and fire stations [Dep06]. In downtown Los Angeles, manufacturing and industrial areas are split cleanly from residential areas, whereas commercial and residential areas are intermixed across the bicycle-sharing system's coverage area.

[3]Permission for use of these proprietary data is granted by the City of Los Angeles Department of City Planning. Copyright © 2015 City of Los Angeles. All Rights Reserved.

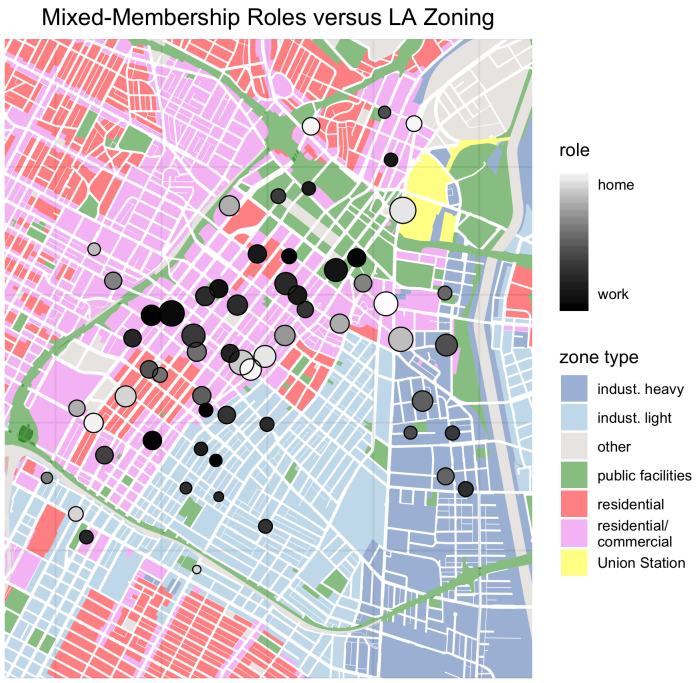


Figure 8.5.3: Mixed-membership (TDMM-SBM) assignments of Los Angeles bicycle-share stations overlaid on a simplified LA zoning map. Industrial blocks include manufacturing and commercial areas. As in Figure 8.5.1, we scale the area of nodes to the value of $\sum_g C_{ig}$.

Figure 8.5.3 illustrates that most stations that are strongly home-identified are in or near zones for pure residential use or mixed residential and commercial use. We find that many stations that are not predominantly home-identified or work-identified align with mixed-use commercial/residential zones. The discrete-role plot (see the right panel of Figure 8.5.1) has a stripe of "home" stations that cut diagonally through the "work" stations. In Figure 8.5.3, we see that this aligns roughly with areas that are zoned for purely residential use. By contrast, industrial and public facility zones tend to host stations that are mostly work-identified (although some of the most strongly work-identified stations are in mixed-used areas).

One station that seems to deviate from the overall pattern is the heavily-trafficked station at Union Station. Although it is adjacent to a public facility zone with many government buildings, it is also strongly home-identified. Although this may seem surprising on its

surface, this is consistent with other home-identified stations, because Union Station is a major transit hub for the Los Angeles metropolitan area. Accordingly, many morning trips originate there, as commuters transition from other forms of transportation, and many evening trips conclude there — an activity pattern that is sensibly associated with home-identified stations. Such idiosyncrasies of transit hubs also arise in our results for San Francisco and New York City.

8.5.2 San Francisco

In Figure 8.5.4, we compare the two-block TDMM-SBM and two-block TDD-SBM for San Francisco. As we saw for Los Angeles, the San Francisco blocks are interspersed geographically, and stations vary from strongly home-identified ones to strongly work-identified ones. The most strongly home-identified station is a major transit hub, the San Francisco Caltrain Station on 4th Street.

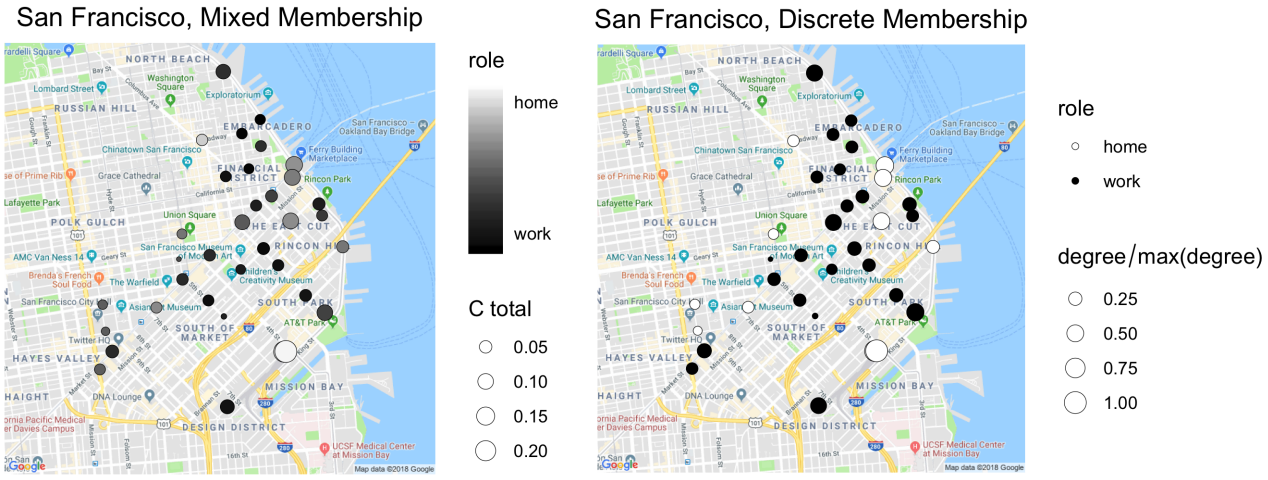


Figure 8.5.4: San Francisco bicycle stations classified using (left) a two-block TDMM-SBM and (right) a two-block TDD-SBM. The sizes of the nodes take continuous values. In the left panel, we scale their area based on the value of $\sum_g C_{ig}$; in the right panel, we scale them based on the sum of the in-degree and out-degree (divided by the maximum value of that sum).

In Figure 8.5.5, we show the estimated traffic between the “home" and “work" blocks

for the TDMM-SBM and TDD-SBM. As in downtown Los Angeles, we observe inter-block commuting. However, unlike in downtown LA, using the discrete model (i.e., TDD-SBM), we observe intra-block morning and evening peaks in both the home and work blocks. This may be due to last-mile commuting, such as using bicycle-sharing facilities to get to or from a train station. Recognizing last-mile usage is important for integrating bicycle sharing with nearby public transportation. One possible reason that we do not observe a similar phenomenon in downtown LA is that San Franciscans are more likely than Angelenos (i.e., inhabitants of Los Angeles) to use public transportation [WK15]. The intra-block morning and evening peaks may also arise from the intermixing of commercial and residential uses of land, such that some travel within blocks may also constitute commuting.

Figure 8.5.5: Estimated time-dependent block-to-block parameters $\hat{\omega}_{ght}$ for the two-block TDMM-SBM and two-block TDD-SBM for San Francisco.

Before presenting our results for New York City, we briefly compare our results from Los Angeles and San Francisco to results for a time-independent SBM fit to these networks, where we have aggregated the data over all time periods. To do this, we calculate the adjacency matrix $\tilde{A}_{ij} = \sum_{t=0}^{23} A_{ijt}$ a time-aggregated network. (The time-independent SBM also has

two blocks and is both directed and degree-corrected.) For downtown Los Angeles, we observe a clear geographically-based division in the results of the time-independent SBM. For San Francisco, however, the differences between the blocks of the time-dependent SBM and time-independent SBM are less noticeable, although they are still present. This confirms that our time-dependent SBMs are detecting behavior that is not evident in the time-aggregated data.

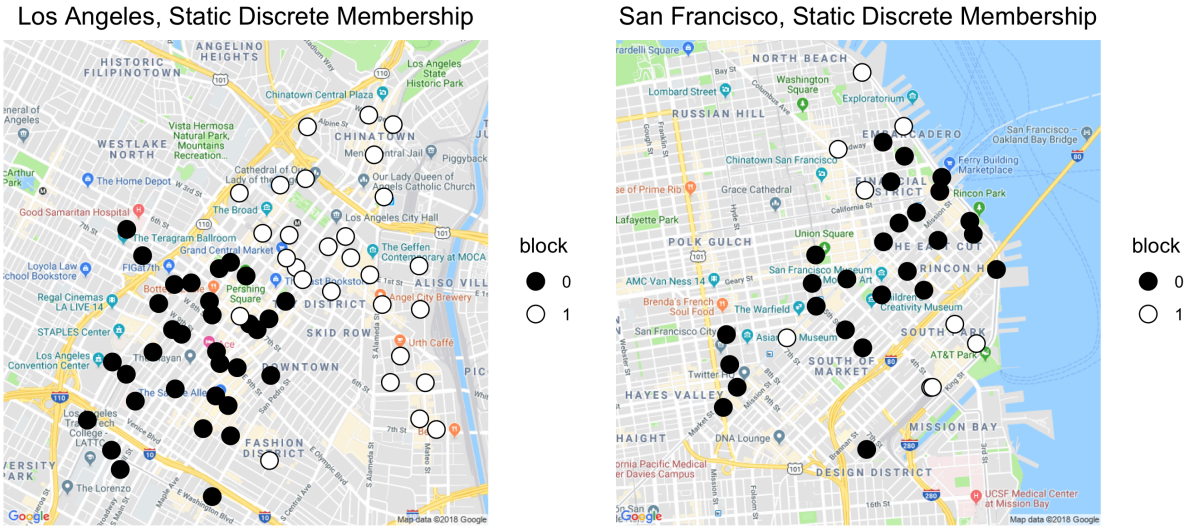


Figure 8.5.6: Estimated blocks of discrete, directed, degree-corrected, time-independent SBM for time-aggregated bicycle-sharing data from Los Angeles and San Francisco.

8.5.3 New York City

In Figure 8.5.7, we compare our results from a three-block TDMM-SBM and a three-block TDD-SBM for New York City. In initial calculations, we found that a two-block TDD-SBM divides the network along the East River into a Manhattan block and Brooklyn block and that the two-block TDMM-SBM divides the network slightly farther north in Lower Manhattan. This suggests a possible limitation to the size of networks for which our time-dependent SBMs can recover functional blocks, as opposed to geographically-based blocks. We will explore this hypothesis further by examining the output of our time-dependent SBMs for the entire New York City bicycle-sharing network and subsequently modeling a subset of the New York City network.

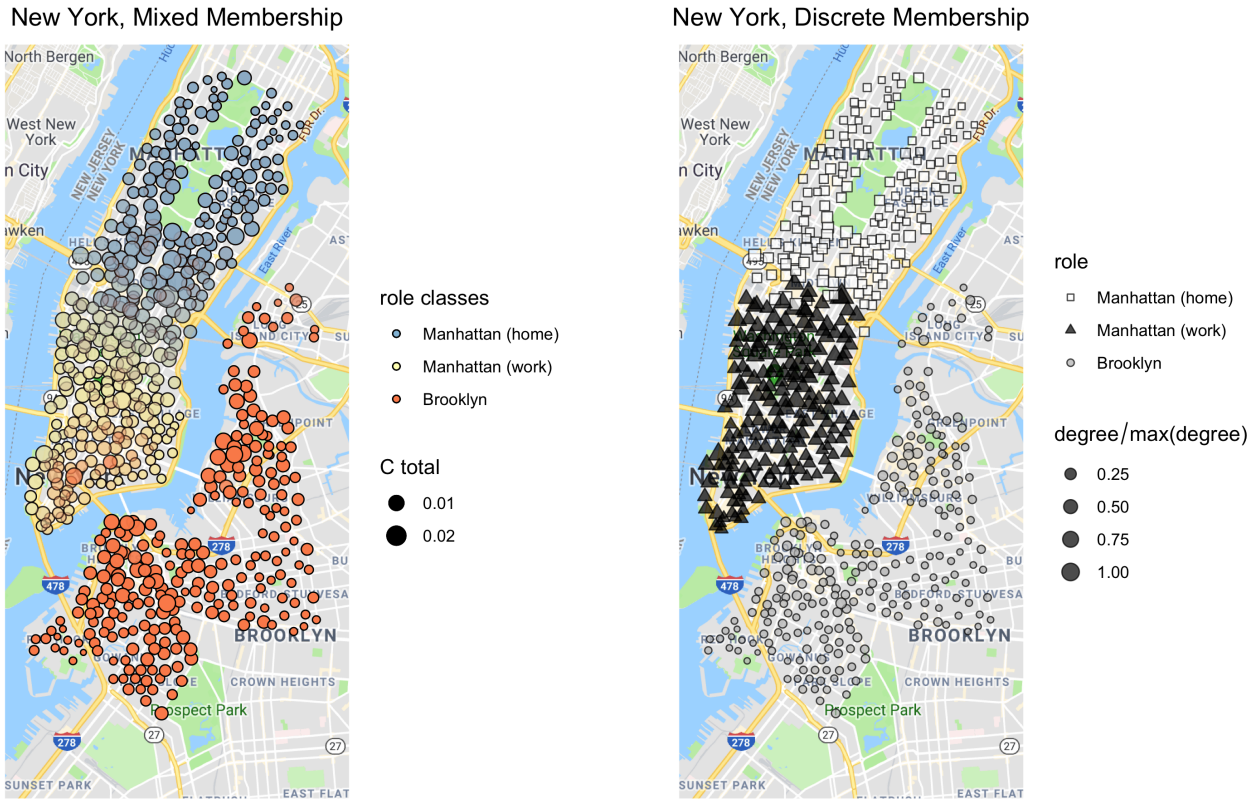


Figure 8.5.7: New York City bicycle stations classified using (left) a three-block TDMM-SBM and (right) a three-block TDD-SBM. The sizes of the nodes take continuous values. In the left panel, we scale their area based on the value of $\sum_g C_{ig}$; in the right panel, we scale them based on the sum of the in-degree and out-degree (divided by the maximum value of that sum).

In Figure 8.5.8, we compare estimated inter-block traffic, as captured by the values of $\hat{\omega}_{ght}$, for the three-block TDMM-SBM and three-block TDD-SBM. We observe prominently that all intra-block traffic has two peaks and much higher hourly trip counts than inter-block traffic. The double peaks are reminiscent of the overall system activity in Figure 8.2.1. This may be due in part to last-mile commuting, as we also suspected in San Francisco. However, for a system that is this large, the double peaks and minimal inter-block traffic suggests that it is useful (and important) to consider each block as its own ecosystem. We also find strong similarity between results from our TDD-SBM and a three-block time-independent SBM for time-aggregated data for New York City (not shown), providing further evidence

that our time-dependent SBMs are not capturing time-dependent roles for New York City. Consequently, we choose the labels of these blocks based on the primary borough and zone type of each block's stations, as indicated in the underlying zoning map for this part of New York City [Dep18] in Figure 8.5.9.

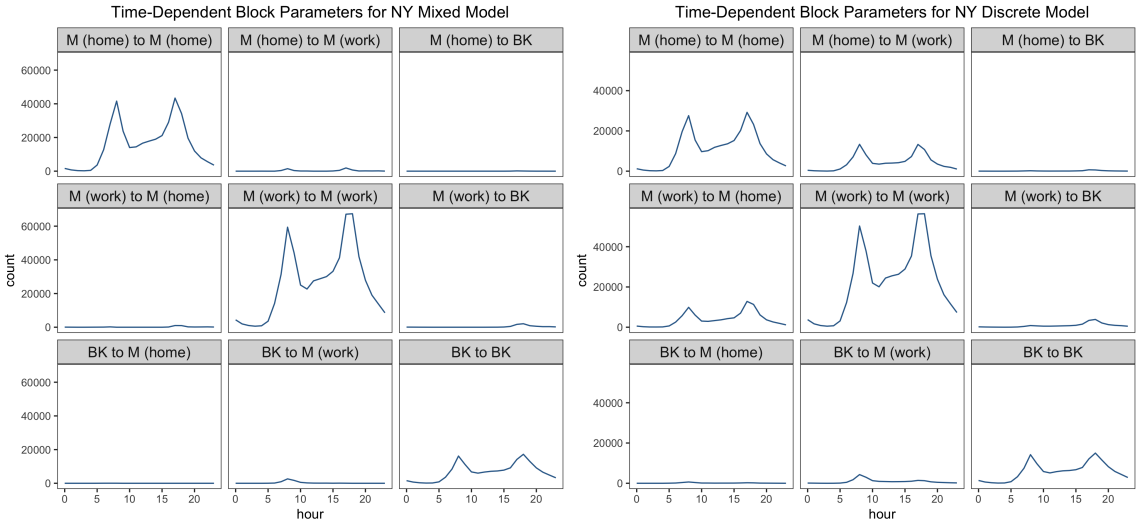


Figure 8.5.8: Estimated time-dependent block-to-block parameters $\hat{\omega}_{ght}$ for the three-block TDMM-SBM and three-block TDD-SBM for New York City. We use "M" to signify Manhattan and "BK" to signify Brooklyn.

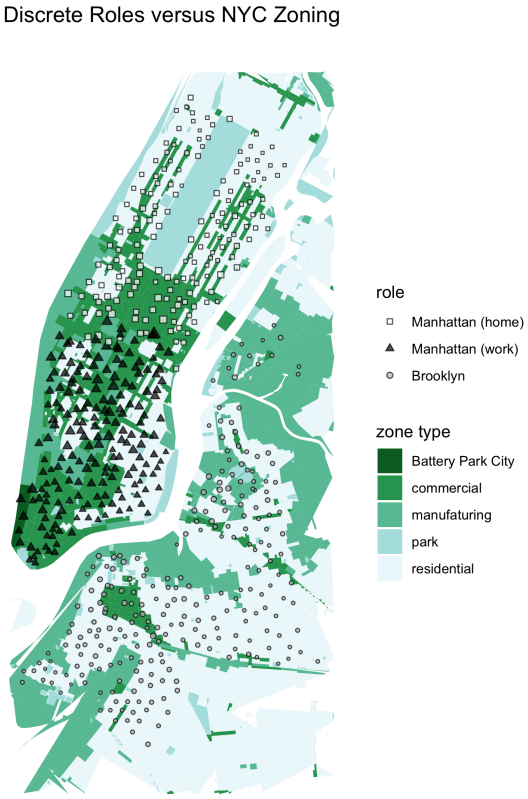


Figure 8.5.9: TDD-SBM station roles versus the coverage-area zoning map of New York City.

In Figure 8.5.9, we illustrate that there is general overlap, although it is far from perfect, between (1) the Upper Manhattan ("home") block and residential areas or parks and between (2) the Lower Manhattan ("work") block and commercial or manufacturing areas. All stations in Brooklyn are in the third block, which contains mostly residential areas. These observations motivate our block labels in this figure, Figure 8.5.7, and Figure 8.5.8. (Although Figure 8.5.9 shows only TDD-SBM-estimated blocks, the same reasoning motivates our labels for the three-block TDMM-SBM.) No block has exclusively commercial or residential areas, reinforcing our conclusion that these blocks represent primarily geographic divisions (with most of the traffic occurring within blocks), as opposed to functionally similar groups of stations.

We examined several time-dependent SBMs for New York City with more than three blocks to try to discover functional blocks, but we found that the blocks were still geographically based. In some cases, TDMM-SBM with larger numbers of blocks were able to find functional divisions within smaller geographic areas (subdividing the blocks in Figure 8.5.7), but neither our discrete model nor our mixed model detected system-wide "home" or "work" blocks. See our supplementary material for our code to fit and visualize time-dependent SBMs of the New York City network with a number of blocks other than three.

8.5.3.1 Manhattan

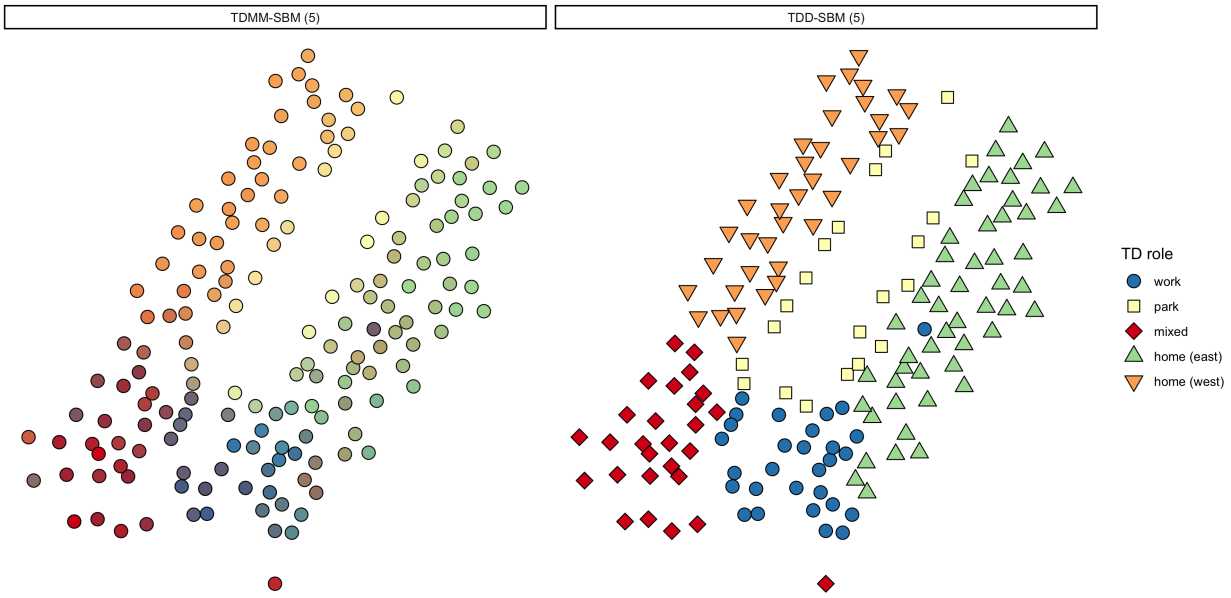


Figure 8.5.10: Comparison of estimated blocks from (left) a five-block TDMM-SBM and (right) a five-block TDD-SBM of the Manhattan (home) block (i.e., the Manhattan subnetwork) of the New York City network (see Figure 8.5.9). In the role labels of the TDD-SBM, we use "W" to represent west and "E" to represent east.

To examine the New York City bicycle-sharing network on a smaller scale, we fit models to the subset of stations and trips within the Manhattan (home) block of the three-block TDD-SBM that we identified above (see Figure 8.5.9); we refer to this subnetwork as the

"Manhattan subnetwork". This subnetwork includes 256,840 trips and 166 stations. In Figure 8.5.10, we present our results for a five-block TDD-SBM and TDD-SBM applied to the Manhattan subnetwork. The area without stations in the middle of each panel of Figure 8.5.10 is Central Park, which has stations on its perimeter but not in its interior.

The estimated blocks of the five-block TDMM-SBM and TDD-SBM (see Figure 8.5.10) of the Manhattan subnetwork outline similar subregions. The mixed-membership block assignments also illustrate how the subregions transition into each other. The models return block-membership parameters that capture the residential and commercial sections of the area much better than the three-block TDD-SBM and TDMM-SBM of the full New York City network; one can see this by comparing the five-block subnetwork results with the underlying zoning map for the area in Figure 8.5.9. The stations in residential zones generally have larger block-membership parameters for "home" blocks than for "work" blocks, and the opposite is true for stations in commercial zones. We label the five detected blocks as (clockwise from top left) "home (west)", "park", "home (east)", "work", and "mixed". We base these labels on the land usage of the underlying areas and the time-dependent block-to-block activity parameters ($\hat{\omega}_{ght}$) that we show in Figure 8.5.11.

We highlight the appearance of the "park" block, which we have not observed in previous models and has distinctive behavior. The park block is similar to a residential block in terms of its spike in morning traffic to the work block and its spike in evening traffic from the work block, but it has distinct intra-block activity that peaks in the afternoon. The intra-block activity resembles weekend activity in the New York City bicycle-sharing system as a whole (see Figure 8.2.1); this reflects leisure use of the bikes. Bicycles near Central Park (which also places them near several major museums) are likely to be used by tourists and other non-commuters during the day for leisure or travel to nearby attractions.

In Figure 8.5.11, we show the values of the block-to-block parameters $\hat{\omega}_{ght}$ for the five-block TDD-SBM and TDMM-SBM. Our estimates of $\hat{\omega}_{ght}$ for these models illustrate important differences in the behavior of different blocks that we can observe only with

a time-dependent model.[4] We see some overlap in the time-dependent behavior of blocks, evidencing potential overfitting. For example, the home (east), home (west), and mixed block have similar traffic with blocks other than their own. However, models of this subnetwork with fewer than five blocks do not cleanly distinguish the “park" block of stations from other residential stations.

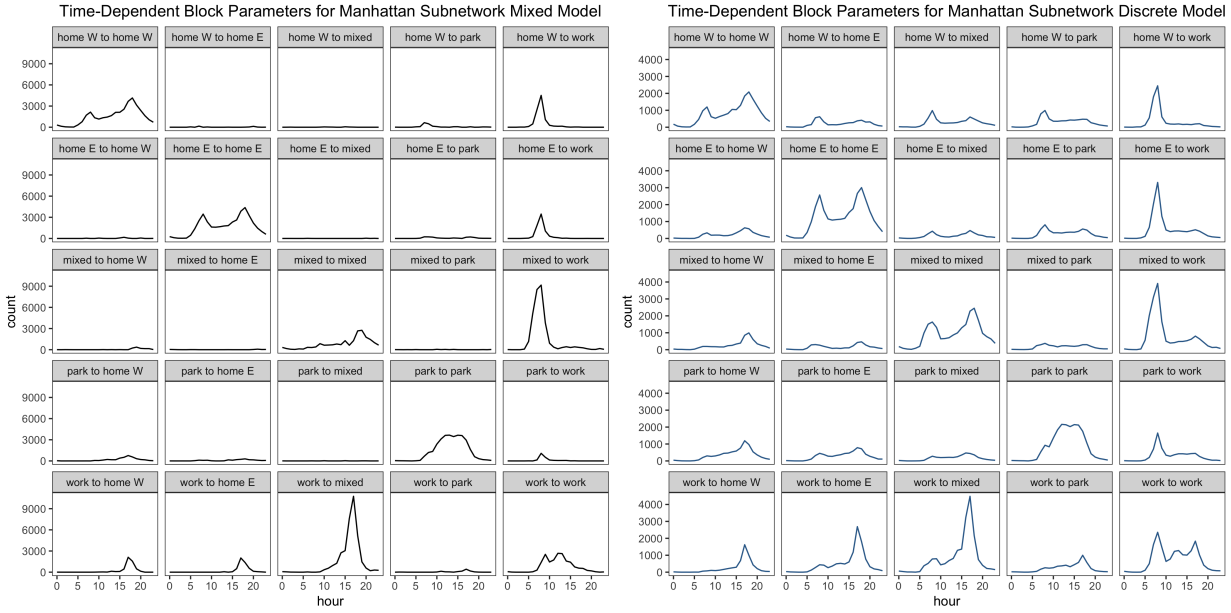


Figure 8.5.11: Estimated time-dependent block-to-block parameters $\hat{\omega}_{ght}$ for (left) a TDMM-SBM with five blocks and (right) a TDD-SBM with five blocks of the Manhattan subnetwork of the New York City bicycle-sharing network.

One reason that our time-dependent SBMs of the Manhattan subnetwork of New York City bicycle-sharing network perform better (with respect to detecting functionally meaningful blocks) than any models that we applied to the entire system is the dependence of station-to-station trip counts on the distance between stations. Although our SBMs correct for the overall activity of each station, they do not normalize expected edge values by the distance between stations. In a small geographic area, such as the coverage areas of the Los Angeles

[4]We obtain similar block identifications for this subnetwork using a discrete, directed, degree-corrected, time-independent SBM as we do from a TDD-SBM with the same number of blocks. We do not show the time-independent SBM results, but they can be produced using the code in our supplementary material.

and San Francisco networks, this is a reasonable choice, as all stations are within “biking distance" of each other. However, when examining a system as large as New York City's, the lack of distance correction weakens the functional groupings that we obtain with our time-dependent SBMs. Intra-block trips dwarf inter-block trips (see Figure 8.5.8), and it seems more reasonable to construe each block as its own ecosystem.

8.5.4 Model Selection

Although statistically rigorous model selection is outside the scope of our paper, we briefly compare the number of parameters in our mixed-membership and discrete SBMs. This is valuable for considering model-selection criteria, such as the Akaike information criterion (AIC) and Bayesian information criterion (BIC), that penalize a model based on its number of parameters. For a network with N nodes, K blocks, and T time slices, the number of parameters for the TDMM-SBM is

$$K \times N - K + T \times K^2 \,, \tag{8.5.1}$$

and the number of parameters for the TDD-SBM is

$$2 \times N - K + T \times K^2 \,. \tag{8.5.2}$$

The first term of (8.5.1) comes from the fact each node in our mixed-membership model has K parameters (C_{ig}, with $g \in \{1, \ldots, K\}$) that express the strength of membership in each block. By contrast, each node in our discrete-membership model has one parameter for block membership and one degree-correction parameter. Therefore, given a value of N, the first term in (8.5.1) increases linearly with the number of blocks, whereas the corresponding term in (8.5.2) is fixed. Otherwise, formulas (8.5.1) and (8.5.2) are equivalent. The $-K$ term in each formula arises from identifiability constraints for each model. As we described in Section 8.3), these constraints are $\sum_i C_{ig} = 1$ for all g for the mixed-membership model

and $\sum_{i \in g} \theta_i = 1$ for all g for the discrete model. The last term in each formula is the total number of ω_{ght} terms in the model (see Section 8.3.1).

In Table 8.5.1, we show the unnormalized log-likelihood and number of parameters (N_p) for TDMM-SBM and TDD-SMB of the Manhattan subnetwork (which has $N = 166$ nodes) with two, three, four, and five blocks.

	TDMM-SBM		**TDD-SBM**	
Number of blocks	N_p	log-likelihood	N_p	log-likelihood
2	426	−260625	426	−270809
3	711	−235162	545	−254779
4	1044	−212295	712	−236198
5	1425	−198489	927	−222539
6	1854	−189670	1190	−216468

Table 8.5.1: Comparison of log-likelihood and number of parameters in models of the Manhattan subnetwork, which has $N = 166$ nodes.

In this example, TDMM-SBM outperforms TDD-SBM with respect to log-likelihood when the two models have the same number of parameters. This result makes sense because of the additional constraint of the TDD-SBM that stations must belong to exactly one block.

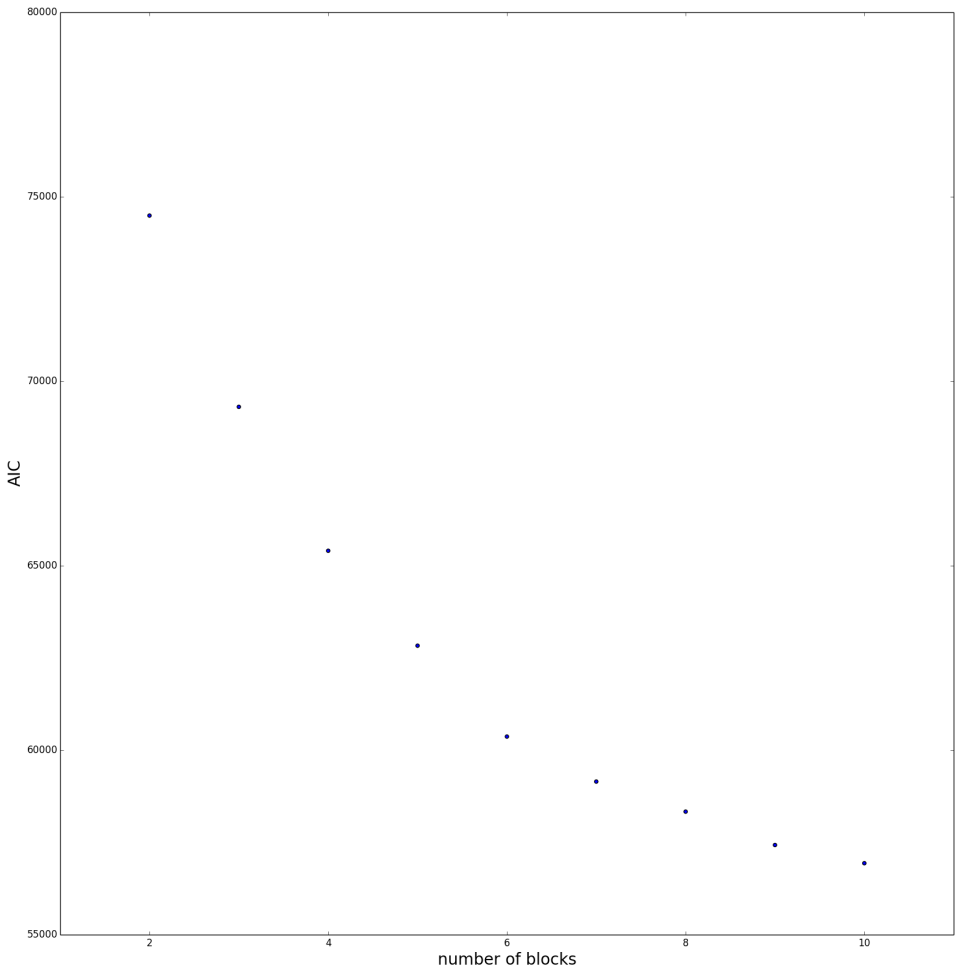


Figure 8.5.12: Akaike information criterion for maximum likelihood TDMM-SBM with 2–10 blocks for the Los Angeles bicycle-sharing network.

Calculating AIC, which is given by [Aka74]

$$\text{AIC} = (2 \times N_p) - (2 \times \text{log-likelihood}) \,,$$

for TDMM-SBM with 2–10 blocks for the Los Angeles bicycle-sharing network selects the TDMM-SBM with the largest number of blocks. The AIC is a cost function for comparing

the relative quality of statistical models; one construes the model with a smaller AIC as the "better" model. The AIC takes into account both the likelihood and how description complexity of a model is in its two summands. The negative log-likelihood is smaller for models with higher likelihood, and one measures the complexity of a model based on its number of parameters. From this perspective, a model with a smaller AIC is better at capturing the trends of a data set while avoiding overfitting. In Figure 8.5.12, we see that AIC decreases as we increase the number of blocks for 2–10 blocks. However, the graphs of the MLE values of ω_{ght} for TDMM-SBM with 7 or more blocks on the Los Angeles network are no more informative than models with fewer blocks. We make this observation for TDMM-SBM with large numbers of blocks using the LA network rather than the Manhattan network, because computing models on data with fewer stations takes less time; and we are confident that the noisy and/or redundant information from using 7 or more blocks on the LA network arise from overfitting.

Although our calculations above are straightforward, choosing appropriate model-selection criteria deserves serious consideration [Yan16, YSJ14]. We leave such an investigation for future work.

8.6 Conclusions and Discussion

We developed time-dependent degree-corrected stochastic block models and used them to analyze daily commute patterns in bicycle-sharing systems in Los Angeles, San Francisco, and New York City. Our SBMs group stations based on their activity over time, allowing us to identify them with home and work roles. Work stations are characterized by inflow from home stations in the morning and outflow to home stations in the afternoon and evening, and residential stations have the opposite characterization. It is also sometimes possible to identify other roles, such as a Manhattan park block that combines residential and leisure/tourist behavior.

We found that many stations in our three focal cities serve a mixture of roles, which we captured with our mixed-membership SBM. However, in some cases, we observed that discrete-membership SBMs that use fewer parameters can provide a clearer picture of the usage patterns that are associated with each. We illustrated through case studies how our discrete and mixed-membership SBMs can provide complementary insights about the bicycle-sharing system behavior. We also demonstrated that applying a time-independent degree-corrected SBM on time-aggregate networks tends to divide stations into contiguous geographic groupings, rather than functional ones.

We evaluated our block labels by comparing them to city zoning maps. The home–work structure that we detected generally aligns well with the underlying zones. However, we found important deviations near major transit hubs. For example, we identified bicycle stations near Union Station in Los Angeles and the San Francisco Caltrain station as home stations because they have high outflow during the mornings and high inflow during the evenings, even though they are not located in residential areas.

It is common to evaluate the results of community detection by comparison with so-called "ground-truth" communities [BDL17, FH16]. (However, it is crucial to encourage caution with respect to such evaluations [PLC17].) The time-dependent commute flows that we detected with our SBMs enabled us to identify and label the functional roles of our blocks. This, in turn, is useful for revealing functional districts without a corresponding zoning map. In the future, it will be worthwhile to compare our detected traffic flows to activity patterns from other mobility data, such as taxis, subway systems, e-scooters, and geo-tagged mobile-app usage [ZKK18].

Developing a deeper understanding of the relationships between station roles and usage patterns throughout a day (and in weekdays versus weekends) can improve the design of bicycle-sharing systems. For example, it can help determine where to place stations and the appropriate sizes of different stations. It can also provide actionable information for efficient redistribution of bicycles, a question that has received much research attention

[SCL13b, PWS14, SSF15, FRT15].

Although our SBMs revealed work and home blocks across different cities, such roles do not describe identical activity patterns. (Accordingly, our labels of "home" and "work" should not be taken as strict classifications of the trips made between blocks, but rather as indicative of dominant activity patterns in a network.) For example, in downtown Los Angeles, the work-to-work activity peaks in the middle of a day, perhaps suggesting that the bicycle-sharing system is being used for errands or leisure activity. In San Francisco and New York City, home-to-home activity has morning and evening peaks, which may be due to the use of the system for "last-mile" transportation or short commutes within mixed-use areas. Such possibilities are reminiscent of recent work on human mobility motifs [SBC13, GHB08]. Using our statistical models, it may be possible to examine footprints of motifs of individual transportation patterns, although it will be necessary for the examined data to include day-by-day movement patterns of individuals to establish a direct connection. Accordingly, we expect it to be fruitful to apply our statistical approaches to the analysis of mobile GPS data.

We now discuss worthwhile future efforts for improving our models and algorithms. In this paper, we presented two types of time-dependent SBMs and used them to reveal interesting urban structures in bicycle-sharing networks. We formulated these SBMs to account for degree heterogeneity and for a balance between cumulative in-degrees and out-degrees of bicycle stations over the course of a day. (The latter feature reflects the classic axiom of Ravenstein [Rav85] that every current of human mobility has an associated countercurrent [BBG18].) However, there is scope for improving our models, just as there is scope for improving SBMs more generally. One area to improve is the fitting of Poisson random variables to the numbers of trips between pairs of stations. Additionally, although it is a convenient simplification to assume independence of edges conditioned on the latent block structure, it would be nice to relax these assumptions. It would also be useful to examine the assumption that the number of trips during each hour are independent random variables.

For example, they are not independent if there are stations that run out of bicycles at some point. As we observed previously, mixed-membership and discrete-membership SBMs can reveal different insights, as can examining SBMs with different numbers of blocks. This helps illustrate why it is important to consider model selection in greater depth.

It would also be useful to explore practical ways that a bicycle-sharing system can exploit our models. Peaks in the computed mean out-degree $\sum_j \mu_{ijt}$ not only indicate expected usage, but they are also suggestive of the main purpose of a docking station (e.g., home, work, or a mixture of the two). A patron who wants to use a bicycle would be very unsatisfied if they walked to a station to pick up a bicycle to ride to work, but they found that no bicycles were available. This suggests that one potentially viable strategy for maintaining the stock bicycles at a station is to ensure that it is above the expected use that is estimated by our SBMs, as our results from both the TDD-SBM and TDMM-SBM identify the most important ways that people are using these stations.

Another important direction for future work is the exploration of different methods for preprocessing data to include only the most significant edges. The two most apparent ways to do this are (1) eliminating insignificant edges by thresholding and (2) choosing time slices that reduce variance. The preferential attachment model of [ZYM13] gives one possible approach for eliminating insignificant edges. The way that one splits the times of a day can improve both accuracy and computational efficiency by reducing the total number of parameters. For example, in the cities that we studied, bicycle trips occurred sporadically between 1 am and 5 am, so it may be desirable lump all of these time slices into one time interval to decrease the number of parameters by 3 and thereby decrease the variance. There exist methods to find suitable ways to segment time periods [CB13], and trying to find the best ones to use in different situations is an active area of research.

Broadening our models to incorporate spatial data is another natural direction to build on our research [Bar18]. The radiation, intervening opportunities, and gravity models have had some success at modeling human mobility over various distances [BBG18]. These models put

more weight on longer trips, and some of them take into account opportunities (the so-called intervening ones) that lie between an origin and destination location. Some of these mobility models also possess statistical justification based on entropy arguments, and it is worthwhile to investigate methods to incorporate them into SBMs. Some of these mobility models have already been incorporated into null models in time-dependent modularity objective functions in the work of Sarzynska et al. [SLC16], who found that radiation and gravity null models perform better than the usual Newman–Girvan null model (which is a variant of a configuration model [New06, FLN18]) for spatial networks. Given that there is an equivalence between a SBM and modularity maximization in a planted-partition model [New16] (and a generalization of this idea arises in multilayer networks [PHL19]), the stage seems to be set for efforts to incorporate spatial information into SBMs. The value of using spatial null models for bicycle-sharing systems was examined in [AOS13, Tai14], so this is a very interesting direction to pursue.

It would be interesting to build on our work using urban spatial null models that go beyond distance and incorporate route difficulty due to traffic or terrain. Some bicycle-sharing researchers are harnessing route information from GPS systems to better understand relationships between station usage and availability of bicycling infrastructure (such as bike lanes) [WB18]. One can also develop models that incorporate station roles. For example, one expects a higher tolerance to distance for traveling between home and work than for a quick bite to eat. Spatial null models can also uncover other types of communities, such as ones that stem from variables (like spoken language and socioeconomic status) that are not directly explainable by spatial data [EEB11, SLC16].

Our time-dependent SBMs are useful for studying many types types of mobility data. For example, it would be interesting to study dockless vehicle-sharing networks, such as e-scooter-sharing programs, using our time-dependent SBMs. If we view the usage of stations as a proxy for a spatial function of demand for bicycles, then data from dockless systems may better approximate such a spatially varying function. One possibility is to partition a

city into a grid (including comparing computations that use different levels of granularity) and measure the usage in each region, taking care to recognize irregularities from transit hubs. Depending on how heavily these systems are used in commuting, we may discover primary functional blocks other than "home" and "work". Looking even further forward, it will also be possible to tailor our methods to analyze multimodal transportation system usage and other urban flows, which are particularly suitable for the setting of multilayer networks [KAB14, AM19, NBI19, GBD19].

Acknowledgements

We thank Brian Karrer and Mark Newman for allowing us to use and share their code for degree-corrected stochastic block models from [KN11]. We thank Susan Handy's lab at UC Davis for useful discussions on contextualizing our work for transportation researchers and planners, David Kempe for introducing us to the work of Rajmonda Caceres [CB13] on choosing a temporal scale for time-dependent networks, and Michelle Feng and others in the networks journal club at UCLA for helpful comments. CM and SSC thank NSF (DMS-1351860) for funding, and SSC also thanks NIGMS (R01 GM126556) and an NIH Ruth L. Kirschstein National Research Service Award (T32-GM008185) for funding. SW thanks NSF (CCF-1422795), ONR (N000141110719, N000141210834), DOD (W81XWH-15-1-0147), Intel STC-Visual Computing Grant (20112360), and Adobe Inc. for funding.

8.7 Appendix

8.7.1 Singular vectors for Los Angeles and San Francisco

We show our singular vectors for the bicycle-sharing networks for downtown Los Angeles in Figure 8.7.1 and for San Francisco in Figure 8.7.2.

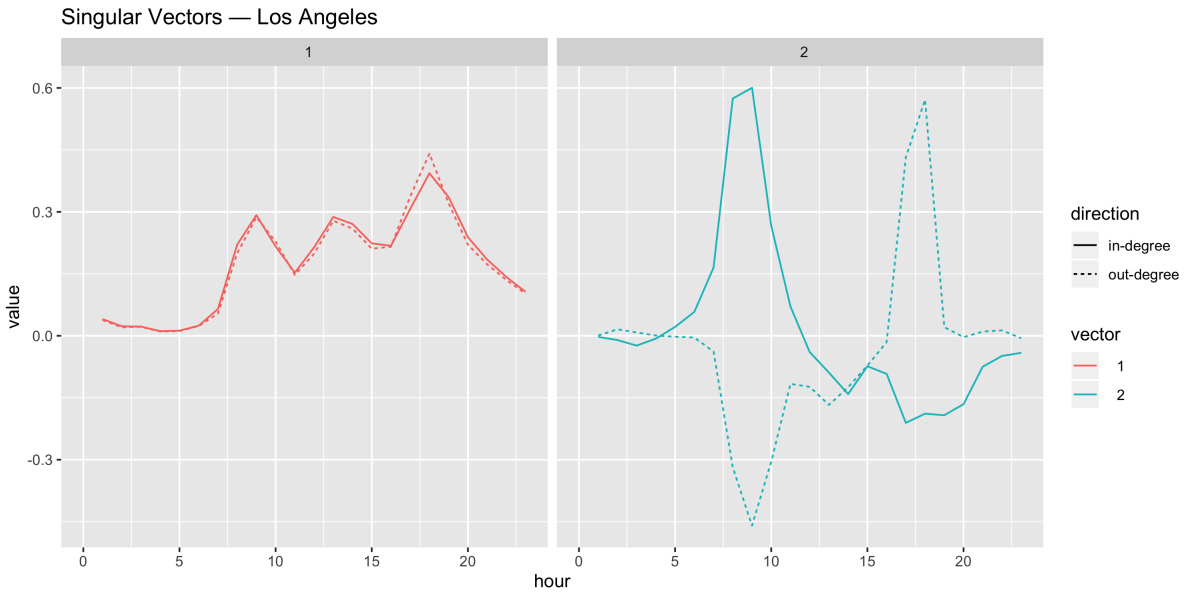


Figure 8.7.1: The first two singular vectors of data for the downtown Los Angeles bicycle-sharing network.

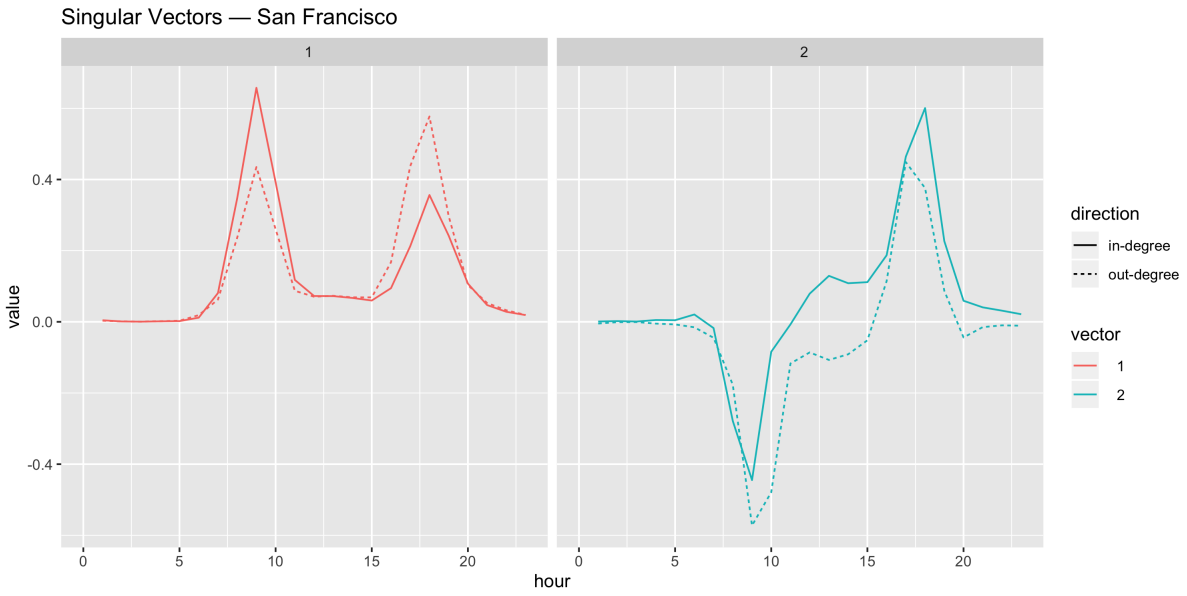


Figure 8.7.2: The first two singular vectors of the data for the San Francisco bicycle-sharing network.

8.7.2 Proof that expected node degrees are the same as node degrees in the data generated from our TDD-SBM

Suppose that we are given a network that is generated by our time-dependent discrete-membership stochastic block model (TDD-SBM). We prove that the expected value of the total degree (i.e., the sum of the in-degree and the out-degree) of a node is the same as the total degree of the node in the observed data.

Let X_{ijt} for each node pair i, j and time $t \in \{0, 1, \ldots, 23\}$ be random edge weights distributed according to the TDD-SBM and inferred from the data A_{ijt}. Recall that g denotes block g in the SBM and that κ_g is the sum of the in-degrees and out-degrees of all nodes in block g over all time periods. For node i, we show that the mean degree of i is equal to the degree of i in the data. That is, $\mathbf{E}\left(\sum_j \sum_{t=0}^{23} X_{ijt}\right) = k_i = \sum_j \sum_{t=0}^{23} A_{ijt}$. We

have

$$\begin{aligned}
\mathbf{E}\left(\sum_{t=0}^{23}\sum_{j}\left(X_{ijt}+X_{jit}\right)\right) &= \frac{k_i}{\kappa_{g_i}}\sum_{t}\sum_{h}\sum_{j\in h}\frac{k_j}{\kappa_h}\left(m_{g_iht}+m_{hg_it}\right)\\
&= \frac{k_i}{\kappa_{g_i}}\sum_{h}\left(m_{g_iht}+m_{hg_it}\right)\\
&= k_i\,.
\end{aligned}$$

We are not aware of a relationship between the expected degrees and degrees of the observed data for our mixed-membership stochastic block model.

BIBLIOGRAPHY

[AAP13] Karen Alim, Gabriel Amselem, François Peaudecerf, Michael P Brenner, and Anne Pringle. "Random network peristalsis in Physarum polycephalum organizes fluid flows across an individual." *Proceedings of the National Academy of Sciences*, **110**(33):13306–13311, 2013.

[AAP17] Karen Alim, Natalie Andrew, Anne Pringle, and Michael P Brenner. "Mechanism of signal propagation in Physarum polycephalum." *Proceedings of the National Academy of Sciences*, **114**(20):5136–5141, 2017.

[ABF08] Edoardo M. Airoldi, David M. Blei, Stephen E. Fienberg, and Eric P Xing. "Mixed Membership Stochastic Blockmodels." *Journal of Machine Learning Research*, **9**:1981–2014, 2008.

[Ach90] David J Acheson. "Elementary fluid dynamics.", 1990.

[Aka74] Hirotugu Akaike. "A new look at the statistical model identification." In *Selected Papers of Hirotugu Akaike*, pp. 215–222. Springer-Verlag, Berlin, Germany, 1974.

[AKF17] Dai Akita, Itsuki Kunita, Mark D Fricker, Shigeru Kuroda, Katsuhiko Sato, and Toshiyuki Nakagaki. "Experimental models for Murray's law." *Journal of Physics D: Applied Physics*, **50**(2):024001, 2017.

[Ali18] Karen Alim. "Fluid flows shaping organism morphology." *Philosophical Transactions of the Royal Society B: Biological Sciences*, **373**(1747):20170112, 2018.

[AM19] A. Aleta and Y. Moreno. "Multilayer networks in a nutshell." *Annual Review of Condensed Matter Physics*, **10**, 2019. Advance online publication, available at `https://doi.org/10.1146/annurev-conmatphys-031218-013259`.

[AMY18] Meisam Akbarzadeh, Syed Sina Mohri, and Ehsan Yazdian. "Designing bike networks using the concept of network clusters." *Applied Network Science*, **3**(1):12, 2018.

[AOS13] Martin Zaltz Austwick, Oliver O'Brien, Emanuele Strano, and Matheus Viana. "The structure of spatial networks and communities in bicycle sharing systems." *PloS One*, **8**(9):e74685, 2013.

[BAF11] Pierre Borgnat, Patrice Abry, Patrick Flandrin, Céline Robardet, Jean-Baptiste Rouquier, and Eric Fleury. "Shared bicycles in a city: A signal processing and data analysis perspective." *Advances in Complex Systems*, **14**(03):415–438, 2011.

[Bar14] Marc Barthélemy. *Spatial networks*. Springer, 2014.

[Bar18] Marc Barthelemy. *Morphogenesis of Spatial Networks*. Springer International Publishing, Cham, Switzerland, 2018.

[Bar19] Marc Barthelemy. "The statistical physics of cities." *Nature Reviews Physics*, **1**:406–415, 2019.

[Bay17] Bay Area Bike Share. "Bay Area Bike Share Open Data.", 2017. Available at `https://s3.amazonaws.com/babs-open-data/babs_open_data_year_3.zip` last checked 2019-05-05; Archived main site: `https://web.archive.org/web/20170303021745/http://www.bayareabikeshare.com/open-data`.

[BBG18] Hugo Barbosa, Marc Barthelemy, Gourab Ghoshal, Charlotte R James, Maxime Lenormand, Thomas Louail, Ronaldo Menezes, José J Ramasco, Filippo Simini, and Marcello Tomasini. "Human mobility: Models and applications." *Physics Reports*, **734**:1–74, 2018.

[BDL17] Pierre Barbillon, Sophie Donnet, Emmanuel Lazega, and Avner Bar-Hen. "Stochastic block models for multiplex networks: an application to a multilevel

network of researchers." *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, **180**(1):295–314, 2017.

[BHD07] Daniel P Bebber, Juliet Hynes, Peter R Darrah, Lynne Boddy, and Mark D Fricker. "Biological solutions to transport network design." *Proceedings of the Royal Society B: Biological Sciences*, **274**(1623):2307–2315, 2007.

[BJA16] Marya Bazzi, Lucas GS Jeub, Alex Arenas, Sam D Howison, and Mason A Porter. "Generative benchmark models for mesoscale structure in multilayer networks." *arXiv:1608.06196*, 2016.

[BM07] Steffen Bohn and Marcelo O Magnasco. "Structure, scaling, and phase transition in the optimal transport network." *Physical review letters*, **98**(8):088702, 2007.

[Boe18] Geoff Boeing. "A multi-scale analysis of 27,000 urban street networks: Every US city, town, urbanized area, and Zillow neighborhood." *Environment and Planning B: Urban Analytics and City Science*, p. 2399808318784595, 2018.

[Cat96] CE Caten. "The mutable and treacherous tribe revisited." *Plant Pathology*, **45**(1):1–12, 1996.

[CB13] Rajmonda Sulo Caceres and Tanya Berger-Wolf. "Temporal scale of dynamic networks." In Petter Holme and Jari Saramäki, editors, *Temporal Networks*, pp. 65–94, Berlin, Germany, 2013. Springer-Verlag.

[CFS18] Hangil Chung, Daniel Freund, and David B. Shmoys. "Bike Angels: An Analysis of Citi Bike's Incentive Program." In *COMPASS*, 2018.

[CGH17] Bob Carpenter, Andrew Gelman, Matthew Hoffman, Daniel Lee, Ben Goodrich, Michael Betancourt, Marcus Brubaker, Jiqiang Guo, Peter Li, and Allen Riddell. "Stan: A Probabilistic Programming Language." *Journal of Statistical Software, Articles*, **76**:1–32, 2017.

[Che89] L Paul Chew. "There are planar graphs almost as good as the complete graph." *Journal of Computer and System Sciences*, **39**(2):205–219, 1989.

[CMM15] Skyler J Cranmer, Elizabeth J Menninga, and Peter J Mucha. "Kantian fractionalization predicts the conflict propensity of the international system." *Proceedings of the National Academy of Sciences of the United States of America*, **112**(38):11812–11816, 2015.

[Cor10] Francis Corson. "Fluctuations and redundancy in optimal transport networks." *Physical Review Letters*, **104**(4):048703, 2010.

[CPS19] Hocine Cherifi, Gergely Palla, Boleslaw K. Szymanski, and Xiaoyan Lu. "On community structure in complex networks: challenges and opportunities.", August 2019. arXiv:1908.04901.

[CR17] Shyr-Shea Chang and Marcus Roper. "A Gradient Descent Method for Optimization of Model Microvascular Networks." *arXiv preprint arXiv:1709.09340*, 2017.

[CR18] Shyr-Shea Chang and Marcus Roper. "Minimal transport networks with general boundary conditions." *SIAM Journal on Applied Mathematics*, **78**(3):1511–1535, 2018.

[CR19] Shyr-Shea Chang and Marcus Roper. "Microvscular networks with uniform flow." *Journal of theoretical biology*, **462**:48–64, 2019.

[CTB17] Shyr-Shea Chang, Shenyinying Tu, Kyung In Baek, Andrew Pietersen, Yu-Hsiu Liu, Van M Savage, Sheng-Ping L Hwang, Tzung K Hsiai, and Marcus Roper. "Optimal occlusion uniformly partitions red blood cells fluxes within a microvascular network." *PLoS computational biology*, **13**(12):e1005892, 2017.

[DBC07] Andrea De Montis, Marc Barthelemy, Alessandro Chessa, and Alessandro Vespignani. "The structure of interurban traffic: A weighted network analysis." *Environment and Planning B: Planning and Design*, **34**(5):905–924, 2007.

[Dep06] Department of City Planning, City of Los Angeles. "Generalized Summary of Zoning Regulations of City of Los Angeles.", January 2006.

[Dep15] Department of City Planning, City of Los Angeles. "City of Los Angeles GIS Data: Zoning.", 2015.

[Dep18] Department of City Planning, New York City. "NYC Zoning Districts.", March 2018.

[Dur07] Marc Durand. "Structure of optimal transport networks subject to a global constraint." *Physical Review Letters*, **98**(8):088701, 2007.

[EEB11] Paul Expert, Tim S. Evans, Vincent D. Blondel, and Renaud Lambiotte. "Uncovering space-independent communities in spatial networks." *Proceedings of the National Academy of Sciences of the United States of America*, **108**:7663–7668, 2011.

[EF11] Dirk Eddelbuettel and Romain François. "Rcpp: Seamless R and C++ Integration." *Journal of Statistical Software*, **40**(8):1–18, 2011.

[EH08] Herbert Edelsbrunner and John Harer. "Persistent homology-a survey." *Contemporary mathematics*, **453**:257–282, 2008.

[EL14] Côme Etienne and Oukhellou Latifa. "Model-Based Count Series Clustering for Bike Sharing System Usage Mining: A Case Study with the Vélib' System of Paris." *ACM Transactions on Intelligent Systems and Technology*, **5**(3):1–21, 2014.

[FH16] Santo Fortunato and Darko Hric. "Community detection in networks: A user guide." *Physics Reports*, **659**:1–44, 2016.

[FLN18] Bailey K. Fosdick, Daniel B. Larremore, Joel Nishimura, and Johan Ugander. "Configuring Random Graph Models with Fixed Degree Sequences." *SIAM Review*, **60**(2):315–355, 2018.

[FRT15] Iris Forma, Tal Raviv, and Michal Tzur. "A 3-step math heuristic for the static repositioning problem in bike-sharing systems." *Transportation Research Part B: Methodological*, **71**:230–247, 01 2015.

[FW92] Katherine Faust and Stanley Wasserman. "Blockmodels: Interpretation and evaluation." *Social Networks*, **14**:5–61, 1992.

[FWH13] Elliot Fishman, Simon Washington, and Narelle Haworth. "Bike share: A synthesis of the literature." *Transport Reviews*, **33**(2):148–165, 2013.

[GBD19] Riccardo Gallotti, Giulia Bertagnolli, and Manlio De Domenico. "Disentangling activity-aware human flows reveals the hidden functional organization of urban systems.", 2019. arXiv:1908.02538.

[GDS12] Roberto Greco, Armando Di Nardo, and G Santonastaso. "Resilience and entropy as indices of robustness of water distribution networks." *Journal of Hydroinformatics*, **14**(3):761–771, 2012.

[GHB08] Marta C. González, Cesar A. Hidalgo, and Albert-Laszlo Barabási. "Understanding individual human mobility patterns." *Nature*, **453**(7196):779, 2008.

[GS16] Greg Phillip Griffin and Ipek Nese Sener. "Planning for Bike Share Connectivity to Rail Transit." *Journal of Public Transportation*, **19**:1–22, 2016.

[Hat02] Allen Hatcher. *Algebraic Topology*. Cambridge University Press, 2002.

[HC13] Dan Hu and David Cai. "Adaptation and optimization of biological transport networks." *Physical review letters*, **111**(13):138701, 2013.

[HGB19] Mark He, Joseph Glasser, Shankar Bhamidi, and Nikhil Kaza. "Intertemporal Community Detection in Bikeshare Networks.", 2019. arXiv:1906.04582.

[HLL83] Paul W. Holland, Kathryn Blackmond Laskey, and Samuel Leinhardt. "Stochastic Blockmodels: First Steps." *Social Networks*, **5**(2):109–137, 1983.

[HOG12] Luke Heaton, Boguslaw Obara, Vincente Grau, Nick Jones, Toshiyuki Nakagaki, Lynne Boddy, and Mark D Fricker. "Analysis of fungal networks." *Fungal Biology Reviews*, **26**(1):12–29, 2012.

[Hol92] John Henry Holland et al. *Adaptation in natural and artificial systems: an introductory analysis with applications to biology, control, and artificial intelligence*. MIT press, 1992.

[Hol15] P. Holme. "Modern temporal network theory: A colloquium." *The European Physical Journal B*, **88**(9), 2015.

[HS12] Petter Holme and Jari Saramäki. "Temporal networks." *Physics Reports*, **519**(3):97–125, 2012.

[HSX11] Qirong Ho, Le Song, and Eric P. Xing. "Evolving cluster mixed-membership blockmodel for time-varying networks." In Geoffrey Gordon, David Dunson, and Miroslav Dudik, editors, *Proceedings of the 14th International Conference on Artificial Intelligence and Statistics*, pp. 342–350, 2011.

[IL15] INRA and Jean-Benoist Leger. *Blockmodels: Latent and stochastic block model estimation by a 'V-EM' algorithm*. INRA, 1.1.1 edition, 2015. R package version 1.1.1.

[Jac77] François Jacob. "Evolution and tinkering." *Science*, **196**(4295):1161–1166, 1977.

[Jin52] John L Jinks. "Heterokaryosis: a system of adaptation in wild fungi." *Proceedings of the Royal Society of London. Series B-Biological Sciences*, **140**(898):83–99, 1952.

[JMM17] L. G. S. Jeub, M. W. Mahoney, P. J. Mucha, and M. A. Porter. "A Local Perspective on Community Structure in Multilayer Networks." *Network Science*, **5**(2):144–163, 2017.

[KAB14] Mikko Kivelä, Alex Arenas, Marc Barthelemy, James P. Gleeson, Yamir Moreno, and Mason A. Porter. "Multilayer networks." *Journal of Complex Networks*, **2**(3):203–271, 2014.

[KG92] J Mark Keil and Carl A Gutwin. "Classes of graphs which approximate the complete Euclidean graph." *Discrete & Computational Geometry*, **7**(1):13–28, 1992.

[KGV83] Scott Kirkpatrick, C Daniel Gelatt, and Mario P Vecchi. "Optimization by simulated annealing." *science*, **220**(4598):671–680, 1983.

[KL70] B. W. Kernighan and S. Lin. "An Efficient Heuristic Procedure for Partitioning Graphs." *Bell System Technical Journal*, **49**(2):291–307, 1970.

[KM12] Eleni Katifori and Marcelo O Magnasco. "Quantifying loopy network architectures." *PloS one*, **7**(6):e37994, 2012.

[KN11] Brian Karrer and Mark EJ Newman. "Stochastic blockmodels and community structure in networks." *Physical Review E*, **83**(1):016107, 2011.

[KSM10] Eleni Katifori, Gergely J Szöllősi, and Marcelo O Magnasco. "Damage and fluctuations induce loops in optimal transport networks." *Physical Review Letters*, **104**(4):048704, 2010.

[KTB19] Teruyoshi Kobayashi, Taro Takaguchi, and Alain Barrat. "The structured backbone of temporal social ties." *Nature Communications*, **10**:220, Jan 2019.

[LA 17] LA Metro. "Metro Bike Share Trip Data.", 2017. Available at `https://bikeshare.metro.net/wp-content/uploads/2017/01/Metro_trips_Q4_2016.zip`; last checked 2019-05-05; the version was last modified on 2018-09-17.

[Lew11] Roger R Lew. "How does a hypha grow? The biophysics of pressurized growth in fungi." *Nature Reviews Microbiology*, **9**(7):509, 2011.

[Luz73] D Luzsa. "X-Ray Anatomy of the Vascular System [in Russian], Acad." *Kiado, Budapest*, 1973.

[LW71] Francois Lorrain and Harrison White. "Structural Equivalence of Individuals in Social Networks." *Journal of Mathematical Sociology*, **1**:49–80, 1971.

[MAA16] Sophie Marbach, Karen Alim, Natalie Andrew, Anne Pringle, and Michael P Brenner. "Pruning to increase Taylor dispersion in Physarum polycephalum networks." *Physical review letters*, **117**(17):178103, 2016.

[MHK18] Fernando Munoz-Mendez, Ke Han, Konstantin Klemmer, and Stephen Jarvis. "Community Structures, Interactions and Dynamics in London's Bicycle Sharing Network." In *Proceedings of the 2018 ACM International Joint Conference and 2018 International Symposium on Pervasive and Ubiquitous Computing and Wearable Computers*, UbiComp '18, pp. 1015–1023, New York, NY, USA, 2018. ACM.

[MM17] Catherine Matias and Vincent Miele. "Statistical clustering of temporal networks through a dynamic stochastic block model." *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, **79**(4):1119–1141, 2017.

[MRM10] Peter J. Mucha, Thomas Richardson, Kevin Macon, Mason A. Porter, and Jukka-Pekka Onnela. "Community structure in time-dependent, multiscale, and multiplex networks." *Science*, **328**(5980):876–878, 2010.

[MRV18] Catherine Matias, Tabea Rebafka, and Fanny Villers. "A semiparametric extension of the stochastic block model for longitudinal networks: Semiparametric estimation in PPSBM." *Biometrika*, **105**(3):665–680, 06 2018.

[Mur26] Cecil D Murray. "The physiological principle of minimum work applied to the angle of branching of arteries." *The Journal of general physiology*, **9**(6):835, 1926.

[nac18] "Bike Share in the U.S.: 2017." Technical report, National Association of City Transportation Officials, May 2018. Available at `https://nacto.org/bike-share-statistics-2017/`.

[NBI19] Luis Natera, Federico Battiston, Gerardo Iñiguez, and Michael Szell. "Data-driven strategies for optimal bicycle network growth.", July 2019. arXiv:1907.07080.

[New06] Mark E. J. Newman. "Finding community structure in networks using the eigenvectors of matrices." *Physical Review E*, **74**(3):036104, 2006.

[New16] M. E. J. Newman. "Equivalence between modularity optimization and maximum likelihood methods for community detection." *Physical Review E*, **94**(5):052315, 2016.

[New18] Mark E. J. Newman. *Networks*. Oxford University Press, Inc., Oxford, UK, second edition, 2018.

[NG04] Mark E J Newman and Michelle Girvan. "Finding and evaluating community structure in networks." *Physical Review E*, **69**(2):026113, 2004.

[NKN04] Toshiyuki Nakagaki, Ryo Kobayashi, Yasumasa Nishiura, and Tetsuo Ueda. "Obtaining multiple separate food sources: behavioural intelligence in the

Physarum plasmodium." *Proceedings of the Royal Society of London. Series B: Biological Sciences*, **271**(1554):2305–2310, 2004.

[NS01] Krzysztof Nowicki and Tom A. B. Snijders. "Estimation and Prediction for Stochastic Blockstructures." *Journal of the American Statistical Association*, **96**(455):1077–1087, 2001.

[NYC17] NYCBS. "Citi Bike System Data.", 2017. Available at `https://s3.amazonaws.com/tripdata/201610-citibike-tripdata.zip`; last checked 2019-05-05.

[NYT01] Toshiyuki Nakagaki, Hiroyasu Yamada, and Agota Toth. "Path finding by tube morphogenesis in an amoeboid organism." *Biophysical chemistry*, **92**(1-2):47–52, 2001.

[NYU00] Toshiyuki Nakagaki, Hiroyasu Yamada, and Tetsuo Ueda. "Interaction between cell shape and contraction pattern in the Physarum plasmodium." *Biophysical chemistry*, **84**(3):195–204, 2000.

[PC16] Subhadeep Paul and Yuguo Chen. "Null Models and Modularity Based Community Detection in Multi-Layer Networks." *arXiv:1608.00623*, 2016.

[Pei18] Tiago P Peixoto. "Bayesian stochastic blockmodeling." *arXiv:1705.10225*, 2018. Chapter in "Advances in Network Clustering and Blockmodeling", edited by P. Doreian, V. Batagelj, A. Ferligoj, (John Wiley & Sons, New York City, USA [forthcoming]).

[PHL19] A. Roxana Pamfil, Sam D. Howison, Renaud Lambiotte, and Mason A. Porter. "Relating modularity maximization and stochastic block models in multilayer networks." *SIAM Journal on Mathematics of Data Science (in press; arXiv:1804.01964)*, 2019.

[PLC17] Leto Peel, Daniel B Larremore, and Aaron Clauset. "The ground truth about metadata and community detection in networks." *Science Advances*, **3**(5):e1602548, 2017.

[POM09] Mason A. Porter, Jukka-Pekka Onnela, and Peter J. Mucha. "Communities in networks." *Notices of the American Mathematical Society*, **56**(9):1082–1097, 1164–1166, 2009.

[Por18] M. A. Porter. "WHAT IS... A Multilayer Network." *Notices of the American Mathematical Society*, **65**(11):1419–1423, 2018.

[PPD18] Lia Papadopoulos, Mason A Porter, Karen E Daniels, and Danielle S Bassett. "Network analysis of particles and grains." *Journal of Complex Networks*, **6**(4):485–565, 2018.

[Puk11] Patricia J Pukkila. "Coprinopsis cinerea." *Current Biology*, **21**(16):R616–R617, 2011.

[PWS14] J. Pfrommer, J. Warrington, G. Schildbach, and M. Morari. "Dynamic Vehicle Redistribution and Online Price Incentives in Shared Mobility Systems." *IEEE Transactions on Intelligent Transportation Systems*, **15**(4):1567–1578, Aug 2014.

[R C18] R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2018.

[RA15] R. A. Rossi and N. K. Ahmed. "Role Discovery in Networks." *IEEE Transactions on Knowledge and Data Engineering*, **27**:1112–1131, 2015.

[RAB09] M. Rosvall, D. Axelsson, and C. T. Bergstrom. "The map equation." *The European Physical Journal Special Topics*, **178**(1):13–23, Nov 2009.

[Rav85] E. J. Ravenstein. "The laws of migration." *Journal of the Statistical Society*, **46**:167–235, 1885.

[RC18] Giulio Rossetti and Rémy Cazabet. "Community Discovery in Dynamic Networks: A Survey." *ACM Comput. Surv.*, **51**(2):35:1–35:37, feb 2018.

[RD19] Marcus Roper and Emilie Dressaire. "Fungal Biology: Bidirectional Communication across Fungal Networks." *Current Biology*, **29**(4):R130–R132, 2019.

[RET11] Marcus Roper, Chris Ellison, John W Taylor, and N Louise Glass. "Nuclear and genome dynamics in multinucleate ascomycete fungi." *Current biology*, **21**(18):R786–R793, 2011.

[RKB11] Camille Roth, Soong Moon Kang, Michael Batty, and Marc Barthelemy. "Structure of urban movements: Polycentric activity and entangled hierarchical flows." *PloS One*, **6**(1):e15923, 2011.

[RMZ18] Gustavo Romanillos, Borja Moya-Gómez, Martin Zaltz-Austwick, and Patxi J Lamíquiz-Daudén. "The pulse of the cycling city: Visualising Madrid bike share system GPS routes and cycling flow." *Journal of Maps*, **14**(1):34–43, 2018.

[RS19] Marcus Roper and Agnese Seminara. "Mycofluidics: the fluid mechanics of fungal adaptation." *Annual Review of Fluid Mechanics*, **51**:511–538, 2019.

[RSH13] Marcus Roper, Anna Simonin, Patrick C Hickey, Abby Leeder, and N Louise Glass. "Nuclear dynamics in a fungal chimera." *Proceedings of the National Academy of Sciences*, **110**(32):12875–12880, 2013.

[RZ16] Gustavo Romanillos and Martin Zaltz Austwick. "Madrid cycle track: Visualizing the cyclable city." *Journal of Maps*, **12**(5):1218–1226, 2016.

[SBC13] Christian M Schneider, Vitaly Belik, Thomas Couronné, Zbigniew Smoreda, and Marta C González. "Unravelling daily human mobility motifs." *Journal of The Royal Society Interface*, **10**:20130246, 2013.

[Sch18] Charles Schmidt. "Active Travel for All? The Surge in Public Bike-Sharing Programs." *Environmental Health Perspectives*, **128**, Aug 2018. available at `https://doi.org/10.1289/EHP3754`.

[SCL13a] Jia Shu, Mabel Chou, Qizhang Liu, Chung Teo, and I-Lin Wang. "Models for Effective Deployment and Redistribution of Bicycles Within Public Bicycle-Sharing Systems." *Operations Research*, **61**:1346–1359, 11 2013.

[SCL13b] Jia Shu, Mabel Chou, Qizhang Liu, Chung Teo, and I-Lin Wang. "Models for Effective Deployment and Redistribution of Bicycles Within Public Bicycle-Sharing Systems." *Operations Research*, **61**:1346–1359, 11 2013.

[SGS17] Emanuele Strano, Andrea Giometto, Saray Shai, Enrico Bertuzzo, Peter J Mucha, and Andrea Rinaldo. "The scaling structure of the global road network." *Royal Society Open Science*, **4**:170590, 2017.

[She81] Thomas F Sherman. "On connecting large vessels to small. The meaning of Murray's law." *The Journal of general physiology*, **78**(4):431–453, 1981.

[SLC16] Marta Sarzynska, Elizabeth A. Leicht, Gerardo Chowell, and Mason A. Porter. "Null models for community detection in spatially embedded, temporal networks." *Journal of Complex Networks*, **4**(3):363–406, 2016.

[SM50] Jack Sherman and Winifred J Morrison. "Adjustment of an inverse matrix corresponding to a change in one element of a given matrix." *The Annals of Mathematical Statistics*, **21**(1):124–127, 1950.

[SN97] Tom A. B. Snijders and Krzysztof Nowicki. "Estimation and Prediction for Stochastic Block-Structures for Graphs with Latent Block Structure." *Journal of Classification*, **14**:75–100, 1997.

[SSF15] Divya Singhvi, Somya Singhvi, Peter I. Frazier, Shane G. Henderson, Eoin O' Mahony, David B. Shmoys, and Dawn B. Woodard. "Predicting bike usage for

New York City's bike sharing system." In *Association for the Advancement of Artificial Intelligence Proceedings*, 2015.

[SST16] Natalie Stanley, Saray Shai, Dane Taylor, and Peter J Mucha. "Clustering network layers with the strata multilayer stochastic block model." *IEEE Transactions on Network Science and Engineering*, **3**(2):95–105, 2016.

[Sta18] Stan Development Team. "RStan: The R interface to Stan.", 2018. R package version 2.18.2.

[Tai14] Aqil Taiyeb. "TP19: Spatial Networks and Human Mobility: An Application of the Intervening Opportunities Model to the London Cycle Hire Scheme." Undergraduate Thesis, Department of Physics, University of Oxford (available at `https://www.math.ucla.edu/~mason/research/aqil-TP19-final.pdf`), 2014.

[THA00a] NAKAGAKI Toshiyuki, YAMADA Hiroyasu, and TÓTH Ágota. "Maze-solving by an amoeboid organism." *Nature*, **407**:470, 2000.

[THA00b] NAKAGAKI Toshiyuki, YAMADA Hiroyasu, and TÓTH Ágota. "Maze-solving by an amoeboid organism." *Nature*, **407**:470, 2000.

[TT93] TT Tanyimboh and AB Templeman. "Calculating maximum entropy flows in networks." *Journal of the Operational Research Society*, **44**(4):383–396, 1993.

[TTS10] Atsushi Tero, Seiji Takagi, Tetsu Saigusa, Kentaro Ito, Dan P Bebber, Mark D Fricker, Kenji Yumiki, Ryo Kobayashi, and Toshiyuki Nakagaki. "Rules for biologically inspired adaptive network design." *Science*, **327**(5964):439–442, 2010.

[VFP15] Eugenio Valdano, Luca Ferreri, Chiara Poletto, and Vittoria Colizza. "Analytical computation of the epidemic threshold on temporal networks." *Physical Review X*, **5**(2):021005, 2015.

[VMG16] Toni Valles-Catala, Francesco A Massucci, Roger Guimera, and Marta Sales-Pardo. "Multilayer stochastic block models reveal the multilayer structure of complex networks." *Physical Review X*, **6**(1):011036, 2016.

[WB18] Jon Wergin and Ralph Buehler. "Where do Bikeshare Bikes Actually Go? An Analysis of Capital Bikeshare Trips Using GPS Data." *Transportation Research Record*, **2662**, 01 2018.

[WBE97] Geoffrey B West, James H Brown, and Brian J Enquist. "A general model for the origin of allometric scaling laws in biology." *Science*, **276**(5309):122–126, 1997.

[WK15] Dan Weikel and Soumya Karlamangla. "San Francisco residents relying less on private automobiles." *Los Angeles Times*, February 23 2015.

[XFS10] Eric P. Xing, Wenjie Fu, and Le Song. "A state-space mixed membership blockmodel for dynamic network tomography." *The Annals of Applied Statistics*, **4**(2):535–566, 06 2010.

[XH14] Kevin S. Xu and Alfred O. Hero. "Dynamic Stochastic Blockmodels for Time-Evolving Social Networks." *IEEE Journal of Selected Topics in Signal Processing*, **8**(4):552–562, 2014.

[XW18] Xiao-Feng Xie and Zunjing Jenipher Wang. "Examining travel patterns and characteristics in a bikesharing network and implications for data-driven decision supports: Case study in the Washington DC area." *Journal of Transport Geography*, **71**:84–102, 2018.

[Yan16] Xiaoran Yan. "Bayesian model selection of stochastic block models." In *Proceedings of the 2016 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining*, pp. 323–328. IEEE Press, 2016.

[YCZ11] Tianbao Yang, Yun Chi, Shenghuo Zhu, Yihong Gong, and Rong Jin. "Detecting communities and their evolutions in dynamic social networks—A Bayesian approach." *Machine Learning*, **82**(2):157–189, Feb 2011.

[YSJ14] Xiaoran Yan, Cosma Shalizi, Jacob E Jensen, Florent Krzakala, Cristopher Moore, Lenka Zdeborová, Pan Zhang, and Yaojia Zhu. "Model selection for degree-corrected block models." *Journal of Statistical Mechanics: Theory and Experiment*, **2014**(5):P05007, 2014.

[ZAH14] Chen Zhong, Stefan Müller Arisona, Xianfeng Huang, Michael Batty, and Gerhard Schmitt. "Detecting the dynamics of urban structure through spatial network analysis." *International Journal of Geographical Information Science*, **28**(11):2178–2199, 2014.

[Zam77] M Zamir. "Shear forces and blood vessel radii in the cardiovascular system." *The Journal of general physiology*, **69**(4):449–461, 1977.

[ZGW18] Zhong-Yuan Zhang, Yujie Gai, Yu-Fei Wang, Hui-Min Cheng, and Xin Liu. "On equivalence of likelihood maximization of stochastic block model and constrained nonnegative matrix factorization." *Physica A*, **503**:687–697, 2018.

[ZKK18] Emnwei Zhu, Maham Khan, Philipp Kats, Shreya Santosh Bamne, and Stanislav Sobolevsky. "Digital Urban Sensing: A Multi-layered Approach." *arXiv e-prints*, 2018. 1809.01280.

[ZMN17] Xiao Zhang, Cristopher Moore, and Mark EJ Newman. "Random graph models for dynamic networks." *The European Physical Journal B*, **90**(10):200, 2017.

[ZWL83] M Zamir, SM Wrigley, and BL Langille. "Arterial bifurcations in the cardiovascular system of a rat." *The Journal of general physiology*, **81**(3):325–335, 1983.

[ZYM13] Yaojia Zhu, Xiaoran Yan, and Cristopher Moore. "Oriented and degree-generated block models: Generating and inferring communities with inhomogeneous degree distributions." *Journal of Complex Networks*, **2**(1):1–18, 2013.