

# UC Irvine

## UC Irvine Previously Published Works

### Title

Concentration of random graphs and application to community detection

### Permalink

<https://escholarship.org/uc/item/3s56w2hf>

### Authors

Le, CM

Levina, E

Vershynin, R

### Publication Date

2018

### Copyright Information

This work is made available under the terms of a Creative Commons Attribution License, available at <https://creativecommons.org/licenses/by/4.0/>

Peer reviewed

# CONCENTRATION OF RANDOM GRAPHS AND APPLICATION TO COMMUNITY DETECTION

CAN M. LE, ELIZAVETA LEVINA AND ROMAN VERSHYNIN

**ABSTRACT.** Random matrix theory has played an important role in recent work on statistical network analysis. In this paper, we review recent results on regimes of concentration of random graphs around their expectation, showing that dense graphs concentrate and sparse graphs concentrate after regularization. We also review relevant network models that may be of interest to probabilists considering directions for new random matrix theory developments, and random matrix theory tools that may be of interest to statisticians looking to prove properties of network algorithms. Applications of concentration results to the problem of community detection in networks are discussed in detail.

## 1. INTRODUCTION

A lot of recent interest in concentration of random graphs has been generated by problems in network analysis, a very active interdisciplinary research area with contributions from probability, statistics, physics, computer science, and the social sciences all playing a role. Networks represent relationships (edges) between objects (nodes), and a network between  $n$  nodes is typically represented by its  $n \times n$  adjacency matrix  $A$ . We will focus on the case of simple undirected networks, where  $A_{ij} = 1$  when nodes  $i$  and  $j$  are connected by an edge, and 0 otherwise, which makes  $A$  a symmetric matrix with binary entries. It is customary to assume the graph contains no self-loops, that is,  $A_{ii} = 0$  for all  $i$ , but this is not crucial. In general, networks may be directed ( $A$  is not symmetric), weighted (the entries of  $A$  have a numerical value representing the strength of connection), and/or signed (the entries of  $A$  have a sign representing whether the relationship is positive or negative in some sense).

Viewing networks as random realizations from an underlying network model enables analysis and inference, with the added difficulty that we often only observe a single realization of a given network. Quantities of interest to be inferred from this realization may include various network properties such as the node degree distribution, the network radius, and community structure. Fundamental to these inferences is the question of how close a single realization of the matrix  $A$  is to the population mean, or the true model,  $\mathbb{E}A$ . If  $A$  is close to  $\mathbb{E}A$ , that is,  $A$  concentrates around its mean,

---

2010 *Mathematics Subject Classification.* 05C80 (primary) and 05C85, 60B20 (secondary).

then inferences drawn from  $A$  can be transferred to the population with high probability.

In this paper, we aim to answer the question “When does  $A$  concentrate around  $\mathbb{E}A$ ?” under a number of network models and asymptotic regimes. We also show that in some cases when the network does not concentrate, a simple regularization step can restore concentration. While the question of concentration is interesting in its own right, we especially focus on the implications for the problem of community detection, a problem that has attracted a lot of attention in the networks literature. When concentration holds, in many cases a simple spectral algorithm can recover communities, and thus concentration is of practical and not only theoretical interest.

## 2. RANDOM NETWORK MODELS

Our concentration results hold for quite general models, but, for the sake of clarity, we provide a brief review of network models, starting from the simplest model and building up in complexity.

**The Erdős-Rényi (ER) graph.** The simplest random network model is the Erdős-Rényi graph  $G(n, p)$  [20]. Under this model, edges are independently drawn between all pairs of nodes according to a Bernoulli distribution with success probability  $p$ . Although the ER model provides an important building block in network modeling and is attractive to analyze, it almost never fits network data observed in practice.

**The stochastic block model (SBM).** The SBM is perhaps the simplest network model with community structure, first proposed by [31]. Under this model, each node belongs to exactly one of  $K$  communities, and the node community membership  $c_i$  is drawn independently from a multinomial distribution on  $\{1, \dots, K\}$  with probabilities  $\pi_1, \dots, \pi_K$ . Conditional on the label vector  $c$ , edges are drawn independently between each pair of nodes  $i, j$ , with

$$\mathbb{P}(A_{ij} = 1) = B_{c_i c_j},$$

where  $B$  is a symmetric  $K \times K$  matrix controlling edge probabilities. Note that each community within SBM is an ER graph. The main question of interest in network analysis is estimating the label vector  $c$  from  $A$ , although model parameters  $\pi$  and  $P$  may also be of interest.

**The degree-corrected stochastic block model (DCSBM).** While the SBM does incorporate community structure, the assumption that each block is an ER graph is too restrictive for many real-world networks. In particular, ER graphs have a Poisson degree distribution, and real networks typically fit the power law or another heavy-tailed distribution better, since they often have “hubs”, influential nodes with many connections. An extension removing this limitation, the degree-corrected stochastic block model (DCSBM) was proposed by [34]. The DCSBM is like an SBM but with each node

assigned an additional parameter  $\theta_i > 0$  that controls its expected degree, and edges drawn independently with

$$\mathbb{P}(A_{ij} = 1) = \theta_i \theta_j B_{c_i c_j}.$$

Additional constraints need to be imposed on  $\theta_i$  for model identifiability; see [34, 61] for options.

**The latent space model (LSM).** Node labels under the SBM or the DCSBM can be thought of as latent (unobserved) node positions in a discrete space of  $K$  elements. More generally, latent positions can be modeled as coordinates in  $\mathbb{R}^d$ , or another set equipped with a distance measure. The LSM [30] assumes that each node  $i$  is associated with an unknown position  $x_i$  and edges are drawn independently between each pair of nodes  $i, j$  with probability inversely proportional to the distance between  $x_i$  and  $x_j$ . If latent positions  $x_i$  form clusters (for example, if they are drawn from a mixture of Gaussians), then a random network generated from this model exhibits community structure. Inferring the latent positions can in principle lead to insights into how the network was formed, beyond simple community assignments.

**Exchangeable random networks.** An analogue of de Finetti’s theorem for networks, due to Hoover and Aldous [32, 4], shows that any network whose distribution is invariant under node permutations can be represented by

$$A_{ij} = g(\alpha, \xi_i, \xi_j, \lambda_{ij}),$$

where  $\alpha$ ,  $\xi_i$  and  $\xi_j$  are independent and uniformly distributed on  $[0, 1]$ , and  $g(u, v, w, z) = g(u, w, v, z)$  for all  $u, v, w, z$ . This model covers all the previously discussed models as special cases, and the function  $g$ , called the graphon, can be estimated up to a permutation under additional assumptions; see [46, 25, 60].

**Network models with overlapping communities.** In practice, it is often more reasonable to allow nodes to belong to more than one community. Multiple such models have been proposed, including the Mixed Membership Stochastic Block Model (MMSBM) [3], the Ball-Karrer-Newman Model (BKN) [7], and the OCCAM model [59]. MMSBM allows different memberships depending on which node the given node interacts with; the BKN models edges as a sum of multiple edges corresponding to different communities; and OCCAM relaxes the membership vector  $c$  under the SBM to have entries between 0 and 1 instead of exactly one “1”. All of these models are also covered by the theory we present, because, conditional on node memberships, all these networks are distributed according to an inhomogeneous Erdős-Rényi model, the most general model we consider, described next.

**The inhomogeneous Erdős-Rényi model.** All models described above share an important property: conditioned on node latent positions, edges are formed independently. The most general form of such a model is the inhomogeneous Erdős-Rényi model (IERM) [13], where each edge is independently drawn, with  $\mathbb{P}(A_{ij} = 1) = P_{ij}$ , where  $P = (P_{ij}) = \mathbb{E} A$ . Evidently, additional assumptions have to be made if latent positions of nodes (however they are defined) are to be recovered from a single realization of  $A$ . We will state concentration results under the IERM as generally as possible, and then discuss additional assumptions under which latent positions can also be estimated reliably.

**Scaling.** We have so far defined all the models for a fixed number of nodes  $n$ , but in order to talk about concentration, we need to determine how the expectation  $P_n = \mathbb{E} A_n$  changes with  $n$ . Most of the literature defines

$$P_n = \rho_n P$$

where  $P$  is a matrix with constant non-negative entries, and  $\rho_n$  controls the average expected degree of the network,  $d = d_n = n\rho_n$ . Different asymptotic regimes have been studied, especially under the SBM; see [1] for a review. Unless  $\rho_n \rightarrow 0$ , the average network degree  $d = \Omega(n)$ , and the network becomes dense as  $n$  grows. In the SBM literature, the regime  $d_n \gg \log n$  is sometimes referred to as semi-dense;  $d_n \rightarrow \infty$  but not faster than  $\log n$  is semi-sparse; and the constant degree regime  $d_n = O(1)$  is called sparse. We will elaborate on these regimes and their implications later on in the paper.

### 3. USEFUL RANDOM MATRIX RESULTS

We start from presenting a few powerful and general tools in random matrix theory which can help prove concentration bounds for random graphs.

**Theorem 3.1** (Bai-Yin law [6]; see [23] and for earlier result). *Let  $M = (M_{ij})_{i,j=1}^{\infty}$  be an infinite, symmetric, and diagonal-free random matrix whose entries above the diagonal are i.i.d. random variables with zero mean and variance  $\sigma^2$ . Suppose further that  $\mathbb{E} M_{ij}^4 < \infty$ . Let  $M_n = (M_{ij})_{i,j=1}^n$  denote the principal minors of  $M$ . Then, as  $n \rightarrow \infty$ ,*

$$\frac{1}{\sqrt{n}} \|M_n\| \rightarrow 2 \quad \text{almost surely.} \quad (3.1)$$

**Theorem 3.2** (Matrix Bernstein's inequality). *Let  $X_1, \dots, X_N$  be independent, mean zero,  $n \times n$  symmetric random matrices, such that  $\|X_i\| \leq K$  almost surely for all  $i$ . Then, for every  $t \geq 0$  we have*

$$\mathbb{P}\left\{\left\|\sum_{i=1}^N X_i\right\| \geq t\right\} \leq 2n \exp\left(-\frac{t^2/2}{\sigma^2 + Kt/3}\right).$$

Here  $\sigma^2 = \left\|\sum_{i=1}^N \mathbb{E} X_i^2\right\|$  is the norm of the “matrix variance” of the sum.

**Corollary 3.3** (Expected norm of sum of random matrices). *We have*

$$\mathbb{E} \left\| \sum_{i=1}^N X_i \right\| \lesssim \sigma \sqrt{\log n} + K \log n.$$

The following result gives sharper bounds on random matrices than matrix Bernstein's inequality, but requires independence of entries.

**Theorem 3.4** (Bandeira-van Handel [8] Corollary 3.6). *Let  $M$  be an  $n \times n$  symmetric random matrix with independent entries on and above the diagonal. Then*

$$\mathbb{E} \|M\| \lesssim \max_i \left( \sum_j \sigma_{ij}^2 \right)^{1/2} + \sqrt{\log n} \max_{i,j} K_{ij},$$

where  $\sigma_{ij}^2 = \mathbb{E} M_{ij}^2$  are the variances of entries and  $K_{ij} = \|M_{ij}\|_\infty$ .

**Theorem 3.5** (Seginer's theorem [54]). *Let  $M$  be a  $n \times n$  symmetric random matrix with i.i.d. mean zero entries above the diagonal and arbitrary entries on the diagonal. Then*

$$\mathbb{E} \|M\| \asymp \mathbb{E} \max_i \|M_i\|_2$$

where  $M_i$  denote the columns of  $M$ .

The lower bound in Seginer's theorem is trivial; it follows from the fact that the operator norm of a matrix is always bounded below by the Euclidean norm of any of its columns. The original paper of Seginer [54] proved the upper bound for non-symmetric matrices with independent entries. The present statement of Theorem 3.5 can be derived by a simple symmetrization argument, see [29, Section 4.1].

#### 4. DENSE NETWORKS CONCENTRATE

If  $A = A_n$  is the adjacency matrix of a  $G(n, p)$  random graph with a constant  $p$ , then the Bai-Yin law gives

$$\frac{1}{\sqrt{n}} \|A - \mathbb{E} A\| \rightarrow 2\sqrt{p(1-p)}.$$

In particular, we have

$$\|A - \mathbb{E} A\| \leq 2\sqrt{d} \tag{4.1}$$

with probability tending to one, where  $d = np$  is the expected node degree.

Can we expect a similar concentration for sparser Erdős-Rényi graphs, where  $p$  is allowed to decrease with  $n$ ? The method of [24] adapted by Feige and Ofek [21] gives

$$\|A - \mathbb{E} A\| = O(\sqrt{d}) \tag{4.2}$$

under the weaker condition  $d \gtrsim \log n$ , which is optimal, as we will see shortly. This argument actually yields (4.2) for inhomogeneous random graphs  $G(n, (p_{ij}))$  as well, and for  $d = \max_{ij} np_{ij}$ , see e.g. [38, 17].

Under a weaker assumption  $d = np \gg \log^4 n$ , Vu [58] proved a sharper bound for  $G(n, p)$ , namely

$$\|A - \mathbb{E} A\| = (2 + o(1))\sqrt{d}, \quad (4.3)$$

which essentially extends (4.1) to sparse random graphs. Very recently, Benaych-Georges, Bordenave and Knowles [10] were able to derive (4.3) under the optimal condition  $d \gg \log n$ . More precisely, they showed that if  $4 \leq d \leq n^{2/13}$ , then

$$\mathbb{E} \|A - \mathbb{E} A\| \leq 2\sqrt{d} + C \sqrt{\frac{\log n}{1 + \log(\log(n)/d)}}.$$

The argument of [10] applies more generally to inhomogeneous random graphs  $G(n, (p_{ij}))$  under a regularity condition on the connection probabilities  $(p_{ij})$ . It even holds for more general random matrices that may not necessarily have binary entries.

To apply Corollary 3.3 to the adjacency matrix  $A$  of an ER random graph  $G(n, p)$ , decompose  $A$  into a sum of independent random matrices  $A = \sum_{i < j} X_{ij}$ , where each matrix  $X_{ij}$  contains a pair of symmetric entries of  $A$ , i.e.  $X_{ij} = A_{ij}(e_i e_i^\top + e_j e_j^\top)$  where  $(e_i)$  denotes the canonical basis in  $\mathbb{R}^n$ . Then apply Corollary 3.3 to the sum of mean zero matrices  $X_{ij} - p$ . It is quick to check that  $\sigma^2 \leq pn$  and obviously  $K \leq 2$ , and so we conclude that

$$\mathbb{E} \|A - \mathbb{E} A\| \lesssim \sqrt{d \log n} + \log n, \quad (4.4)$$

where  $d = np$  is the expected degree. The same argument applies more generally to inhomogeneous random graphs  $G(n, (p_{ij}))$ , and it still gives (4.4) when

$$d = \max_i \sum_j p_{ij}$$

is the maximal expected degree.

The logarithmic factors in bound (4.4) are not optimal, and can be improved by applying the result of Bandeira and van Handel (Theorem 3.4) to the centered adjacency matrix  $A - \mathbb{E} A$  of an inhomogeneous random graph  $G(n, (p_{ij}))$ . In this case,  $\sigma_{ij}^2 = p_{ij}$  and  $K_{ij} \leq 1$ , so we obtain the following sharpening of (4.4).

**Proposition 4.1** (Concentration of inhomogeneous random graphs). *Let  $A$  be the adjacency matrix of an inhomogeneous random graph  $G(n, (p_{ij}))$ . Then*

$$\mathbb{E} \|A - \mathbb{E} A\| \lesssim \sqrt{d} + \sqrt{\log n}, \quad (4.5)$$

where  $d = \max_i \sum_j p_{ij}$  is the expected maximal degree.

In particular, if the graph is not too sparse, namely  $d \gtrsim \log n$ , then the optimal concentration (4.3) holds, i.e.

$$\mathbb{E} \|A - \mathbb{E} A\| \lesssim \sqrt{d}.$$

This recovers a result of Feige-Ofek [21].

A similar bound can be alternatively proved using the general result of Seginer (Theorem 3.5). If  $A$  is the adjacency matrix of  $G(n, p)$ , it is easy to check that  $\mathbb{E} \max_i \|A_i\|_2 \lesssim \sqrt{d} + \sqrt{\log n}$ . Thus, Seginer’s theorem implies the optimal concentration bound (4.5) as well. Using simple convexity arguments, one can extend this to inhomogeneous random graphs  $G(n, (p_{ij}))$ , and get the bound (4.5) for  $d = \max_{ij} np_{ij}$ , see [29, Section 4.1].

One may wonder if Seginer’s theorem holds for matrices with independent but not identically distributed entries. Unfortunately, this is not the case in general; a simple counterexample was found by Seginer [54], see [8, Remark 4.8]. Nevertheless, it is an open conjecture of Latala that Seginer’s theorem does hold if  $M$  has independent *Gaussian* entries, see the papers [52, 56] and the survey [57].

## 5. SPARSE NETWORKS CONCENTRATE AFTER REGULARIZATION

**5.1. Sparse networks do not concentrate.** In the sparse regime  $d = np \ll \log n$ , the Bai-Yin’s law for  $G(n, p)$  fails. This is because in this case, degrees of some vertices are much higher than the expected degree  $d$ . This causes some rows of the adjacency matrix  $A$  to have Euclidean norms much larger than  $\sqrt{d}$ , which in turn gives

$$\|A - \mathbb{E} A\| \gg \sqrt{d}.$$

In other words, concentration fails for very sparse graphs; there exist outlying eigenvalues that escape the interval  $[-2, 2]$  where the spectrum of denser graphs lies according to (3.1). For precise description of this phenomenon, see the original paper [35], a discussion in [8, Section 4] and the very recent work [9].

**5.2. Sparse networks concentrate after regularization.** One way to regularize a random network in the sparse regime is to remove high degree vertices altogether from the network. Indeed, [21] showed that for  $G(n, p)$ , if we drop all vertices with degrees, say, larger than  $2d$ , then the remaining part of the network satisfies  $\|A - \mathbb{E} A\| = O(\sqrt{d})$  with high probability. The argument in [21] is based on the method developed by [24] and it is extended to the IERM in [38, 17].

Although removal of high degree vertices restores concentration, in practice this is a bad idea, since the loss of edges associated with “hub” nodes in an already sparse network leads to a considerable loss of information, and in particular community detection tends to break down. A more gentle regularization proposed in [36] does not remove high degree vertices, but reduces the weights of their edges just enough to keep the degrees bounded by  $O(d)$ .

**Theorem 5.1** (Concentration of regularized adjacency matrices). *Consider a random graph from the inhomogeneous Erdős-Rényi model  $G(n, (p_{ij}))$ , and let  $d = \max_{ij} np_{ij}$ .*



Consider any subset of at most  $10n/d$  vertices, and reduce the weights of the edges incident to those vertices in an arbitrary way, but so that all degrees of the new (weighted) network become bounded by  $2d$ . For any  $r \geq 1$ , with probability at least  $1 - n^{-r}$  the adjacency matrix  $A'$  of the new weighted graph satisfies

$$\|A' - \mathbb{E} A\| \leq Cr^{3/2}\sqrt{d}.$$

Proving concentration for this kind of general regularization requires different tools. One key result we state next is the Grothendieck-Pietsch factorization, a general and well-known result in functional analysis [48, 49, 55, 50] which has already been used in a similar probabilistic context [37, Proposition 15.11]. It compares two matrix norms, the spectral norm  $\ell_2 \rightarrow \ell_2$  and the  $\ell_\infty \rightarrow \ell_2$  norm.

**Theorem 5.2** (Grothendieck-Pietsch factorization). *Let  $B$  be a  $k \times m$  real matrix. Then there exist positive weights  $\mu_j$  with  $\sum_{j=1}^m \mu_j = 1$  such that*

$$\|B\|_{\infty \rightarrow 2} \leq \|BD_\mu^{-1/2}\| \leq 2\|B\|_{\infty \rightarrow 2},$$

where  $D_\mu = \text{diag}(\mu_j)$  denotes the  $m \times m$  diagonal matrix with weights  $\mu_j$  on the diagonal.

*Idea of the proof of Theorem 5.1 by network decomposition.* The argument in [21] becomes very complicated for handling the general regularization in Theorem 5.1. A simpler alternative approach was developed by [36] for proving Theorem 5.1. The main idea is to decompose the set of entries  $[n] \times [n]$  into different subsets with desirable properties. There exists a partition (see Figure 1c for illustration)

$$[n] \times [n] = \mathcal{N} \cup \mathcal{R} \cup \mathcal{C}$$

such that  $A$  concentrates on  $\mathcal{N}$  even without regularization, while restrictions of  $A$  onto  $\mathcal{R}$  and  $\mathcal{C}$  have small row and column sums, respectively. It is easy to see that the degree regularization does not destroy the properties of  $\mathcal{N}$ ,  $\mathcal{R}$  and  $\mathcal{C}$ . Moreover, it creates a new property, allowing for controlling the columns of  $\mathcal{R}$  and rows of  $\mathcal{C}$ . Together with the triangle inequality, this implies the concentration of the entire network.

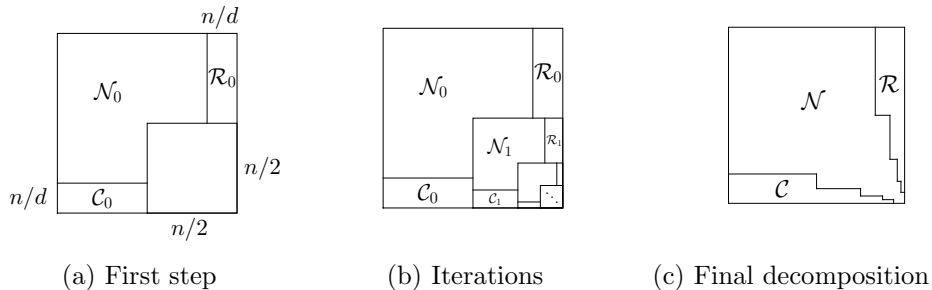


Figure 1. Constructing network decomposition iteratively.

The network decomposition is constructed by an iterative procedure. We first establish concentration of  $A$  in  $\ell_\infty \rightarrow \ell_2$  norm using standard probability techniques. Next, we upgrade this to concentration in the spectral norm  $\|(A - \mathbb{E}A)_{\mathcal{N}_0}\| = O(\sqrt{d})$  on an appropriate (large) subset  $\mathcal{N}_0 \subseteq [n] \times [n]$  using the Grothendieck-Pietsch factorization (Theorem 5.2). It remains to control  $A$  on the complement of  $\mathcal{N}_0$ . That set is small; it can be described as a union of a block  $\mathcal{C}_0$  with a small number of rows, a block  $\mathcal{R}_0$  with a small number of columns and an exceptional (small) block (see Figure 1a). Now we repeat the process for the exceptional block, decomposing it into  $\mathcal{N}_1$ ,  $\mathcal{R}_1$ , and  $\mathcal{C}_1$ , and so on, as shown in Figure 1b. At the end, we set  $\mathcal{N} = \cup_i \mathcal{N}_i$ ,  $\mathcal{R} = \cup_i \mathcal{R}_i$  and  $\mathcal{C} = \cup_i \mathcal{C}_i$ . The cumulative error from this iterative procedure can be controlled appropriately; see [36] for details.

**5.3. Concentration of the graph Laplacian.** So far, we have looked at random graphs through the lens of their adjacency matrices. Another matrix that captures the structure of a random graph is the Laplacian. There are several ways to define the Laplacian; we focus on the symmetric, normalized Laplacian,

$$\mathcal{L}(A) = D^{-1/2}AD^{-1/2}.$$

Here  $D = \text{diag}(d_i)$  is the diagonal matrix with degrees  $d_i = \sum_{j=1}^n A_{ij}$  on the diagonal. The reader is referred to [18] for an introduction to graph Laplacians and their role in spectral graph theory. Here we mention just two basic facts: the spectrum of  $\mathcal{L}(A)$  is a subset of  $[-1, 1]$ , and the largest eigenvalue is always one.

In the networks literature in particular, community detection has been mainly done through spectral clustering on the Laplacian, not on the adjacency matrix. We will discuss this in more detail in Section 6, but the primary reason for this is degree normalization: as discussed in Section 2, real networks rarely have the Poisson or mixture of Poissons degree distribution that characterizes the stochastic block model; instead, “hubs”, or high degree vertices, are common, and they tend to break down spectral clustering on the adjacency matrix itself.

Concentration of Laplacians of random graphs has been studied by [47, 16, 51, 33, 26]. Just like the adjacency matrix, the Laplacian is known to concentrate in the dense regime  $d = \Omega(\log n)$ , and it fails to concentrate in the sparse regime. However, the reasons it fails to concentrate are different. For the adjacency matrix, as we discussed, concentration fails in the sparse case because of high degree vertices. For the Laplacian, it is the low degree vertices that destroy concentration. In fact, it is easy to check that when  $d = o(\log n)$ , the probability of isolated vertices is non-vanishing; and each isolated vertex contributes an eigenvalue of 0 to the spectrum of  $\mathcal{L}(A)$ , which is easily seen to destroy concentration.

Multiple ways to regularize the Laplacian in order to deal with the low degree vertices have been proposed. Perhaps the two most common ones are adding a small constant to all the degrees on the diagonal of  $D$  [16],

and adding a small constant to all the entries of  $A$  before computing the Laplacian. Here we focus on the latter regularization, proposed by [5] and analyzed by [33, 26]. Choose  $\tau > 0$  and add the same number  $\tau/n$  to all entries of the adjacency matrix  $A$ , thereby replacing it with

$$A_\tau := A + \frac{\tau}{n} \mathbf{1}\mathbf{1}^\top \quad (5.1)$$

Then compute the Laplacian as usual using this new adjacency matrix. This regularization raises all degrees  $d_i$  to  $d_i + \tau$ , and eliminates isolated vertices, making the entire graph connected. The original paper [5] suggested the choice  $\tau = \rho \bar{d}$ , where  $\bar{d}$  is the average node degree and  $\rho \in (0, 1)$  is a constant. They showed the estimator is not particularly sensitive to  $\rho$  over a fairly wide range of values away from 0 (too little regularization) and 1 (too much noise). The choice of  $\rho = 0.25$  was recommended by [5] but this parameter can also be successfully chosen by cross-validation on the network [39].

The following consequence of Theorem 5.1 shows that regularization (5.1) indeed forces the Laplacian to concentrate.

**Theorem 5.3** (Concentration of the regularized Laplacian). *Consider a random graph drawn from the inhomogeneous Erdős-Rényi model  $G(n, (p_{ij}))$ , and let  $d = \max_{i,j} np_{ij}$ . Choose a number  $\tau > 0$ . Then, for any  $r \geq 1$ , with probability at least  $1 - e^{-r}$  we have*

$$\|\mathcal{L}(A_\tau) - \mathcal{L}(\mathbb{E} A_\tau)\| \leq \frac{Cr^2}{\sqrt{\tau}} \left(1 + \frac{d}{\tau}\right)^{5/2}.$$

In the next section, we discuss why concentration of the adjacency matrix and/or its Laplacian is important in the context of community detection, the primary application of concentration in network analysis.

## 6. APPLICATION TO COMMUNITY DETECTION

Concentration of random graphs has been of such interest in networks analysis primarily because it relates to the problem of community detection; see [22, 27, 1] for reviews of community detection algorithms and results. We should specify that, perhaps in a slight misnomer, “community detection” refers to the task of assigning each node to a community (typically one and only one), not to the question of whether there are communities present, which might be a more natural use of the term “detection”.

Most of the theoretical work linking concentration of random graphs to community detection has focused on the stochastic block model (SBM), defined in Section 2, which is one of the many special cases of the general IERM we consider. For the purpose of this paper, we focus on the simplest version of the SBM for which the largest number of results has been obtained so far, also known as the balanced planted partition model  $G(n, \frac{a}{n}, \frac{b}{n})$ . In this model, there are  $K = 2$  equal-sized communities with  $n/2$  nodes each. Edges between vertices within the same community are drawn independently with

probability  $a/n$ , and edges between vertices in different communities with probability  $b/n$ . The task is to recover the community labels of vertices from a single realization of the adjacency matrix  $A$  drawn from this model. The large literature on both the recovery algorithms and the theory establishing when a recovery is possible is very nicely summarized in the recent excellent review [1], where we refer the reader for details and analogues for a general  $K$  (now available for most results) and the asymmetric SBM (very few are available). In the following subsections we give a brief summary for the symmetric  $K = 2$  case which does not aim to be exhaustive.

**6.1. Community detection phase transition.** *Weak recovery*, sometimes also called *detection*, means performing better than randomly guessing the labels of vertices. The phase transition threshold for weak recovery was first conjectured in the physics literature by [19], and proved rigorously by [44, 45, 43], with follow-up and related work by [2, 40, 14]. The phase transition result says that there exists a polynomial time algorithm which can classify more than 50% of the vertices correctly as  $n \rightarrow \infty$  with high probability if and only if

$$(a - b)^2 > 2(a + b).$$

Performing better than random guessing is the weakest possible guarantee of performance, which is of interest in the very sparse regime of  $d = (a+b)/2 = O(1)$ ; when the degree grows, weak recovery becomes trivial. This regime has been mostly studied by physicists and probabilists; in the statistics literature, *consistency* has been of more interest.

**6.2. Consistency of community detection.** Two types of consistency have been discussed in the literature. Strong consistency, also known as exact recovery, means labeling *all* vertices correctly with high probability, which is, as the name suggests, a very strong requirement. Weak consistency, or “almost exact” recovery, is the weaker and arguably more practically reasonable requirement that the fraction of misclassified vertices goes to 0 as  $n \rightarrow \infty$  with high probability.

*Strong consistency* was studied first, in a seminal paper [12], as well as by [43, 41, 29, 15]. Strong consistency is achievable, and achievable in polynomial time, if

$$\left| \sqrt{\frac{a}{\log n}} - \sqrt{\frac{b}{\log n}} \right| > \sqrt{2}$$

and not possible if  $\left| \sqrt{a/n} - \sqrt{b/n} \right| < \sqrt{2}$ . In particular, strong consistency is normally only considered in the semi-dense regime of  $d/\log n \rightarrow \infty$ .

*Weak consistency*, as one would expect, requires a stronger condition than weak recovery but a weaker one than strong consistency. Weak consistency

is achievable if and only if

$$\frac{(a-b)^2}{a+b} = \omega(1)$$

see for example [43]. In particular, weak consistency is achievable in the semi-sparse regime of  $d \rightarrow \infty$ .

*Partial recovery*, finally, refers to the situation where the fraction of misclassified vertices does not go to 0, but remains bounded by a constant below 0.5. More specifically, partial recovery means that for a fixed  $\varepsilon > 0$  one can recover communities up to  $\varepsilon n$  mislabeled vertices. For the balanced symmetric case, this is true as long as

$$\frac{(a-b)^2}{a+b} = O(1)$$

which is primarily relevant when  $d = O(1)$ . Several types of algorithms are known to succeed at partial recovery in this very sparse regime, including non-backtracking walks [44, 40, 14], spectral methods [17] and methods based on semidefinite programming [28, 42].

**6.3. Concentration implies recovery.** As an example application of the new concentration results, we demonstrate how to show that *regularized spectral clustering* [5, 33], one of the simplest and most popular algorithms for community detection, can recover communities in the sparse regime of constant degrees. In general, spectral clustering works by computing the leading eigenvectors of either the adjacency matrix or the Laplacian, or their regularized versions, and running the  $k$ -means clustering algorithm on the rows of the  $n \times k$  matrix of leading eigenvectors to recover the node labels. In the simplest case of the balanced  $K = 2$  model  $G(n, \frac{a}{n}, \frac{b}{n})$ , one can simply assign nodes to two communities according to the sign of the entries of the eigenvector  $v_2(A')$  corresponding to the second smallest eigenvalue of the (regularized) adjacency matrix  $A'$ .

Let us briefly explain how concentration results validate recovery from the regularized adjacency matrix or regularized Laplacian. If concentration holds and the regularized matrix  $A'$  is shown to be close to  $\mathbb{E}A$ , then standard perturbation theory (i.e., the Davis-Kahan theorem, see e.g. [11]) implies that  $v_2(A')$  is close to  $v_2(\mathbb{E}A)$ , and in particular, the signs of these two eigenvectors must agree on most vertices. An easy calculation shows that the signs of  $v_2(\mathbb{E}A)$  recover the communities exactly: the eigenvector corresponding to the second smallest eigenvalue of  $\mathbb{E}A$  (or the second largest of  $\mathcal{L}(A)$ ) is a positive constant on one community and a negative constant on the other. Therefore, the signs of  $v_2(A')$  recover communities up to a small fraction of misclassified vertices and, as always, up to a permutation of community labels. This argument remains valid if we replace the regularized adjacency matrix  $A'$  with regularized Laplacian  $\mathcal{L}(A_\tau)$ .

**Corollary 6.1** (Partial recovery from a regularized adjacency matrix for sparse graphs). *Let  $\varepsilon > 0$  and  $r \geq 1$ . Let  $A$  be the adjacency matrix drawn from the stochastic block model  $G(n, \frac{a}{n}, \frac{b}{n})$ . Assume that*

$$(a - b)^2 > C(a + b)$$

*where  $C$  is a constant depending only on  $\varepsilon$  and  $r$ . For all nodes with degrees larger than  $2a$ , reduce the weights of the edges incident to them in an arbitrary way, but so that all degrees of the new (weighted) network become bounded by  $2a$ , resulting in a new matrix  $A'$ . Then with probability at least  $1 - e^{-r}$ , the signs of the entries of the eigenvector corresponding to the second smallest eigenvalue of  $A'$  correctly estimate the partition into two communities, up to at most  $\varepsilon n$  misclassified vertices.*

**Corollary 6.2** (Partial recovery from a regularized Laplacian for sparse graphs). *Let  $\varepsilon > 0$  and  $r \geq 1$ . Let  $A$  be the adjacency matrix drawn from the stochastic block model  $G(n, \frac{a}{n}, \frac{b}{n})$ . Assume that*

$$(a - b)^2 > C(a + b) \tag{6.1}$$

*where  $C$  is a constant depending only on  $\varepsilon$  and  $r$ . Choose  $\tau$  to be the average degree of the graph, i.e.  $\tau = (d_1 + \dots + d_n)/n$ . Then with probability at least  $1 - e^{-r}$ , the signs of the entries of the eigenvector corresponding to the second largest eigenvalue of  $\mathcal{L}(A_\tau)$  correctly estimate the partition into the two communities, up to at most  $\varepsilon n$  misclassified vertices.*

As we have discussed, the Laplacian is typically preferred over the adjacency matrix in practice, because the variation in node degrees is reduced by the normalization factor  $D^{-1/2}$  [53]. Figure 2 shows the effect of regularization for the Laplacian of a random network generated from  $G(n, \frac{a}{n}, \frac{b}{n})$  with  $n = 50$ ,  $a = 5$  and  $b = 0.1$ . For plotting purposes, we order the nodes so that the first  $n/2$  nodes belong to one community and the rest belong to the other community. Without regularization, the two leading eigenvectors of the Laplacian localize around a few low degree nodes, and therefore do not contain any information about the global community structure. In contrast, the second leading eigenvector of the regularized Laplacian (with  $\tau = 0.1\bar{d}$ ) clearly reflects the communities, and the signs of this eigenvector alone recover community labels correctly for all but three nodes.

## 7. DISCUSSION

Great progress has been made in recent years, and yet many problems remain open. Open questions on community detection under the SBM, in terms of exact and partial recovery and efficient (polynomial time) algorithms are discussed in [1], and likely by the time this paper comes out in print, some of them will have been solved. Yet the focus on the SBM is unsatisfactory for many practitioners, since not many real networks fit this model well. Some of the more general models we discussed in Section 2 fix some of the problems of the SBM, allowing for heterogeneous degree

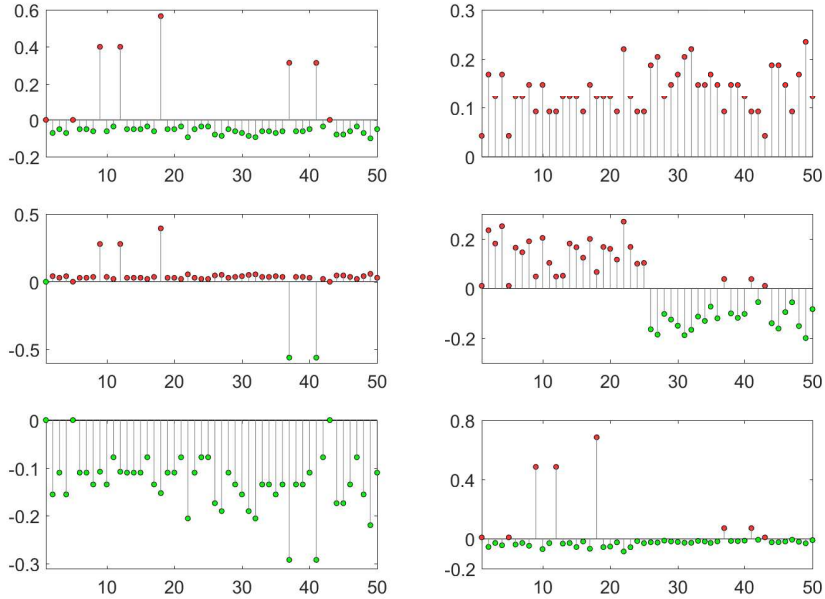


Figure 2. Three leading eigenvectors (from top to bottom) of the Laplacian (left) and the regularized Laplacian (right). The network is generated from  $G(n, \frac{a}{n}, \frac{b}{n})$  with  $n = 50$ ,  $a = 5$  and  $b = 0.1$ . Nodes are labeled so that the first 25 nodes belong to one community and the rest to the other community. Regularized Laplacian is computed from  $A + 0.1\bar{d}/n\mathbf{1}\mathbf{1}^\top$ .

distributions and overlapping communities, for instance. A bigger problem lies in the fixed  $K$  regime; it is not realistic to assume that as the size of the network grows, the number of communities remains fixed. A more realistic model is the “small world” scenario, where the size of communities remains bounded or grows very slowly with the number of nodes, the number of communities grows, and connections between many smaller communities happen primarily through hub nodes. Some consistency results have been obtained for a growing  $K$ , but we are not aware of any results in the sparse constant degree regime so far. An even bigger problem is presented by the so far nearly universal assumption of independent edges; this assumption violates commonly observed transitivity of friendships (if A is friends with B and B is friends with C, A is more likely to be friends with C). There are other types of network models that do not rely on this assumption, but hardly any random matrix results apply there. Ultimately, network analysis involves a lot more than community detection: link prediction, network denoising, predicting outcomes on networks, dynamic network modeling over

time, and so on. We are a long way away from establishing rigorous theoretical guarantees for any of these problems to the extent that we have for community detection, but given how rapid progress in the latter area has been, we are hopeful that continued interest from the random matrix community will help shed light on other problems in network analysis.

## REFERENCES

- [1] E. Abbe. Community detection and the stochastic block model. *arXiv:1703.10146*, 2017.
- [2] E. Abbe, A. S. Bandeira, and G. Hall. Exact recovery in the stochastic block model. *arXiv:1405.3267*, 2014.
- [3] E. M. Airoldi, D. M. Blei, S. E. Fienberg, and E. P. Xing. Mixed membership stochastic blockmodels. *J. Machine Learning Research*, 9:1981–2014, 2008.
- [4] D. J. Aldous. Representations for partially exchangeable arrays of random variables. *J. Multivariate Analysis*, 11:581–598, 1981.
- [5] A. A. Amini, A. Chen, P. J. Bickel, and E. Levina. Pseudo-likelihood methods for community detection in large sparse networks. *The Annals of Statistics*, 41(4):2097–2122, 2013.
- [6] Z. D. Bai and Y. Q. Yin. Necessary and sufficient conditions for almost sure convergence of the largest eigenvalue of a wigner matrix. *The Annals of Probability*, 16(4):1729–1741, 1988.
- [7] B. Ball, B. Karrer, and M. E. J. Newman. An efficient and principled method for detecting communities in networks. *Physical Review E*, 34:036103, 2011.
- [8] A. Bandeira and R. V. Handel. Sharp nonasymptotic bounds on the norm of random matrices with independent entries. *The Annals of Probability*, 44(4):2479–2506, 2016.
- [9] F. Benaych-Georges, C. Bordenave, and A. Knowles. Largest eigenvalues of sparse inhomogeneous Erdős-Rnyi graphs. *arXiv:1704.02953*, 2017.
- [10] F. Benaych-Georges, C. Bordenave, and A. Knowles. Spectral radii of sparse random matrices. *arXiv:1704.02945*, 2017.
- [11] R. Bhatia. *Matrix Analysis*. Springer-Verlag New York, 1996.
- [12] P. J. Bickel and A. Chen. A nonparametric view of network models and Newman-Girvan and other modularities. *Proc. Natl. Acad. Sci. USA*, 106:21068–21073, 2009.
- [13] B. Bollobas, S. Janson, and O. Riordan. The phase transition in inhomogeneous random graphs. *Random Structures and Algorithms*, 31:3–122, 2007.
- [14] C. Bordenave, M. Lelarge, and L. Massoulié. Non-backtracking spectrum of random graphs: community detection and non-regular Ramanujan graphs. *arxiv:1501.06087*, 2015.
- [15] T. T. Cai and X. Li. Robust and computationally feasible community detection in the presence of arbitrary outlier nodes. *The Annals of Statistics*, 43(3):1027–1059, 2015.
- [16] K. Chaudhuri, F. Chung, and A. Tsiatas. Spectral clustering of graphs with general degrees in the extended planted partition model. *Proceedings of Machine Learning Research*, 23:35.1 – 35.23, 2012.
- [17] P. Chin, A. Rao, and V. Vu. Stochastic block model and community detection in the sparse graphs : A spectral algorithm with optimal rate of recovery. *Proceedings of Machine Learning Research*, 40:391–423, 2015.
- [18] F. R. K. Chung. *Spectral Graph Theory*. CBMS Regional Conference Series in Mathematics, 1997.
- [19] A. Decelle, F. Krzakala, C. Moore, and L. Zdeborová. Asymptotic analysis of the stochastic block model for modular networks and its algorithmic applications. *Physical Review E*, 84:066106, 2011.
- [20] P. Erdős and A. Rényi. On random graphs. I. *Publ. Math. Debrecen*, 6:290–297, 1959.



- [21] U. Feige and E. Ofek. Spectral techniques applied to sparse random graphs. *Wiley InterScience*, 2005.
- [22] S. Fortunato. Community detection in graphs. *Physics Reports*, 486(3-5):75 – 174, 2010.
- [23] Z. Fredi and J. Komls. The eigenvalues of random symmetric matrices. *Combinatorica*, 1:3:233–241, 1981.
- [24] J. Friedman, J. Kahn, and E. Szemerédi. On the second eigenvalue in random regular graphs. *Proceedings of the 21st Annual ACM Symposium in Theory of Computing*, pages 587–598, 1989.
- [25] C. Gao, Y. Lu, and H. H. Zhou. Rate-optimal graphon estimation. *Ann. Statist.*, 43(6):2624–2652, 2015.
- [26] C. Gao, Z. Ma, A. Y. Zhang, and H. H. Zhou. Achieving optimal misclassification proportion in stochastic block model. *arXiv:1505.03772*, 2015.
- [27] A. Goldenberg, A. X. Zheng, S. E. Fienberg, and E. M. Airoldi. A survey of statistical network models. *Foundations and Trends in Machine Learning*, 2:129–233, 2010.
- [28] O. Guédon and R. Vershynin. Community detection in sparse networks via Grothendieck’s inequality. *arXiv:1411.4686*, 2014.
- [29] B. Hajek, Y. Wu, and J. Xu. Achieving exact cluster recovery threshold via semidefinite programming. *IEEE Transactions on Information Theory*, 62(5):2788–2797, 2016.
- [30] P. D. Hoff, A. E. Raftery, and M. S. Handcock. Latent space approaches to social network analysis. *Journal of the American Statistical Association*, 97:1090–1098, 2002.
- [31] P. W. Holland, K. B. Laskey, and S. Leinhardt. Stochastic blockmodels: first steps. *Social Networks*, 5(2):109–137, 1983.
- [32] D. Hoover. Relations on probability spaces and arrays of random variables. Technical report, Institute for Advanced Study, Princeton, NJ, 1979.
- [33] A. Joseph and B. Yu. Impact of regularization on spectral clustering. *arXiv:1312.1733*, 2013.
- [34] B. Karrer and M. E. J. Newman. Stochastic blockmodels and community structure in networks. *Physical Review E*, 83:016107, 2011.
- [35] M. Krivelevich and B. Sudakov. The largest eigenvalue of sparse random graphs. *Combin Probab Comput*, 12:61–72, 2003.
- [36] C. M. Le, E. Levina, and R. Vershynin. Concentration and regularization of random graphs. *Random Structures and Algorithms*, 51(3):538–561, 2017.
- [37] M. Ledoux and M. Talagrand. *Probability in Banach spaces: Isoperimetry and processes*. Springer-Verlag, Berlin, 1991.
- [38] J. Lei and A. Rinaldo. Consistency of spectral clustering in stochastic block models. *Ann. Statist.*, 43(1):215–237, 2015.
- [39] T. Li, E. Levina, and J. Zhu. Network cross-validation by edge sampling. *arXiv:1612.04717*, 2016.
- [40] L. Massoulié. Community detection thresholds and the weak Ramanujan property. In *Proceedings of the 46th Annual ACM Symposium on Theory of Computing*, STOC ’14, pages 694–703, 2014.
- [41] F. McSherry. Spectral partitioning of random graphs. *Proc. 42nd FOCS*, pages 529–537, 2001.
- [42] A. Montanari and S. Sen. Semidefinite programs on sparse random graphs and their application to community detection. *Proceedings of the 48th Annual ACM symposium on Theory of Computing (STOC ’16)*, pages 814–827, 2016.
- [43] E. Mossel, J. Neeman, and A. Sly. Consistency thresholds for binary symmetric block models. *arXiv:1407.1591*, 2014.
- [44] E. Mossel, J. Neeman, and A. Sly. A proof of the block model threshold conjecture. *arXiv:1311.4115*, 2014.

- [45] E. Mossel, J. Neeman, and A. Sly. Reconstruction and estimation in the planted partition model. *Probability Theory and Related Fields*, 2014.
- [46] S. C. Olhede and P. J. Wolfe. Network histograms and universality of blockmodel approximation. *PNAS*, 111(41):14722–14727, 2013.
- [47] R. Oliveira. Concentration of the adjacency matrix and of the Laplacian in random graphs with independent edges. *arXiv:0911.0600*, 2010.
- [48] A. Pietsch. *Operator Ideals*. North-Holland Amsterdam, 1978.
- [49] G. Pisier. *Factorization of linear operators and geometry of Banach spaces*. CBMS Regional Conference Series in Mathematics, 60. AMS, Providence, 1986.
- [50] G. Pisier. Grothendiecks theorem, past and present. *Bulletin (New Series) of the American Mathematical Society*, 49(2):237–323, 2012.
- [51] T. Qin and K. Rohe. Regularized spectral clustering under the degree-corrected stochastic blockmodel. In *Advances in Neural Information Processing Systems*, pages 3120–3128, 2013.
- [52] S. Riemer and C. Schütt. On the expectation of the norm of random matrices with non-identically distributed entries. *Electron. J. Probab.*, 18(29):1–13, 2013.
- [53] P. Sarker and P. Bickel. Role of normalization in spectral clustering for stochastic blockmodels. *Ann. Statist.*, 43(3):962–990, 2015.
- [54] Y. Seginer. The expected norm of random matrices. *Combinatorics, Probability and Computing*, 9(2):149–166, 2000.
- [55] N. Tomczak-Jaegermann. *Banach-Mazur distances and finite-dimensional operator ideals*. John Wiley & Sons, Inc., New York, 1989.
- [56] R. van Handel. On the spectral norm of Gaussian random matrices. *Trans. Amer. Math. Soc.*, 369(11):8161–8178, 2017.
- [57] R. van Handel. Structured random matrices. In *Carlen E., Madiman M., Werner E. (eds) Convexity and Concentration. IMA Vol. 161, Springer*, 2017.
- [58] V. Vu. Spectral norm of random matrices. *Combinatorica*, 27(6):721–736, 2007.
- [59] Y. Zhang, E. Levina, and J. Zhu. Detecting overlapping communities in networks using spectral methods. *arXiv:1412.3432*, 2014.
- [60] Y. Zhang, E. Levina, and J. Zhu. Estimating network edge probabilities by neighbourhood smoothing. *Biometrika*, 104(4):771–783, 2017.
- [61] Y. Zhao, E. Levina, and J. Zhu. Consistency of community detection in networks under degree-corrected stochastic block models. *Annals of Statistics*, 40(4):2266–2292, 2012.

DEPARTMENT OF STATISTICS, UNIVERSITY OF CALIFORNIA, DAVIS, ONE SHIELDS AVE,  
DAVIS, CA 95616, U.S.A.

*E-mail address:* [canle@ucdavis.edu](mailto:canle@ucdavis.edu)

DEPARTMENT OF STATISTICS, UNIVERSITY OF MICHIGAN, 1085 S. UNIVERSITY AVE,  
ANN ARBOR, MI 48109, U.S.A.

*E-mail address:* [elevina@umich.edu](mailto:elevina@umich.edu)

DEPARTMENT OF MATHEMATICS, UNIVERSITY OF CALIFORNIA, IRVINE, 340 ROWLAND  
HALL, IRVINE, CA 92697, U.S.A.

*E-mail address:* [rvershyn@uci.edu](mailto:rvershyn@uci.edu)