

# UC Berkeley

## International Conference on GIScience Short Paper Proceedings

### Title

Exploring the Structure of Contact Networks at Multiple Scales

### Permalink

<https://escholarship.org/uc/item/26g8f7cd>

### Journal

International Conference on GIScience Short Paper Proceedings, 1(1)

### Authors

Gao, Peng  
Bian, Ling

### Publication Date

2016

### DOI

10.21433/B31126g8f7cd

Peer reviewed

# Exploring the Structure of Contact Networks at Multiple Scales

P. Gao and L. Bian

Department of Geography, University at Buffalo, The State University of New York, Buffalo, NY 14261, United States

Email: {pgao3, lbian}@buffalo.edu

## Abstract

This study examines the scale effects on the network structure and spatial structure of contact networks. Regular grids with different cell sizes are used to divide one observed and two reference random networks into multiple scales. Four metrics are used to represent the two structures. Results show that the network structure of the observed network is sensitive to scale changes at fine scales. In comparison, the clustered spatial structure is scale independent.

## 1. Introduction

Contact networks play a critical role in disease dispersion, as repeatedly stressed in reports on the most dangerous communicable diseases, such as SARS, H1N1, and Ebola (Ferguson *et al.* 2005). Understanding the properties of contact networks helps gain insights into how to prevent and control the dispersion of these diseases (Keeling and Eames 2005).

Disease dispersion is inherently a spatial process, while scale is involved in all spatial phenomena (Wu and Wu 2013). There are rarely studies addressing the scale effects on network properties. Contact networks have both network structure and spatial structure (Barthélemy 2011), while the spatial structure has received attention only in recent years (Bian 2013).

This study aims to evaluate (1) the scale effects on the network structure and spatial structure of contact networks, and (2) the ranges of scale at which contact networks are scale dependent. To achieve these goals, three networks, one observed and two randomly structured, are partitioned into multiple scales using regular grids. The properties of the resultant networks represented by four metrics are compared across scales.

## 2. Contact Network

This study uses a contact network previously constructed for a metropolitan community in the Northeastern US (Bian *et al.* 2012). Each individual belongs to a family and most belong to a workplace. There are two types of contact relationships between individuals, those between family members and those between co-workers. The network represents individuals as nodes and the contact relationships as edges, resulting in a total of 64,726 nodes and 194,683 edges. The two types of relationships are treated as family and co-worker edges, respectively.

To examine the scale effects, the contact network is projected into space. Nodes are projected according to their home locations. A family edge is within a home location with a zero physical distance. The co-worker edges are between two different home locations and have various physical distances.

### 3. Methods

#### 3.1 Four Metrics

Three metrics are used to represent the network structure, including the relative size of the largest component  $S$ , clustering coefficient  $cc$ , and relative average path length  $l$ . A component is a cluster of nodes within a network. All nodes are directly or indirectly (through a chain of edges) connected to all other nodes within the cluster but disconnected with nodes in other clusters.  $S$  is the ratio of the number of nodes in the largest component to the total number of nodes in the network (Newman 2010). A greater  $S$  indicates a more cohesive network.

The clustering coefficient of a given node is the number of edges between its neighboring nodes divided by the number of all possible edges between the neighboring nodes (Newman 2010).  $cc$  of the network is the average clustering coefficient over all individual nodes. A higher  $cc$  means a stronger locally clustered structure.

The path length is the number of consecutive edges between a pair of nodes.  $l$  is the average length of all shortest paths divided by the shortest path length between the pair of nodes that are most apart in the network (Newman 2010). A shorter  $l$  implies a more efficiently connected network.

The statistical distribution of edge distance ( $Dist$ ) measures the spatial structure of the network (Barthélemy 2011). A negatively skewed distribution indicates the dominance of short edges, thus a spatially clustered structure, while a normal distribution indicates a spatially scattered structure.

#### 3.2 Three Networks

In addition to the observed network, a random-node network and a random-edge network are simulated to serve as references to the observed network. The two random networks keep the identical number of nodes and edges and identical degree distributions as the observed network.

The random-node network keeps the network structure of the observed network, but alters the spatial structure by randomizing node locations. The random-edge network keeps the spatial structure of the observed network, but alters the network structure by randomly shuffling edges between nodes. Both random networks are generated 1,000 times. The average of the metrics is used to represent the properties of the random networks for the subsequent analysis.

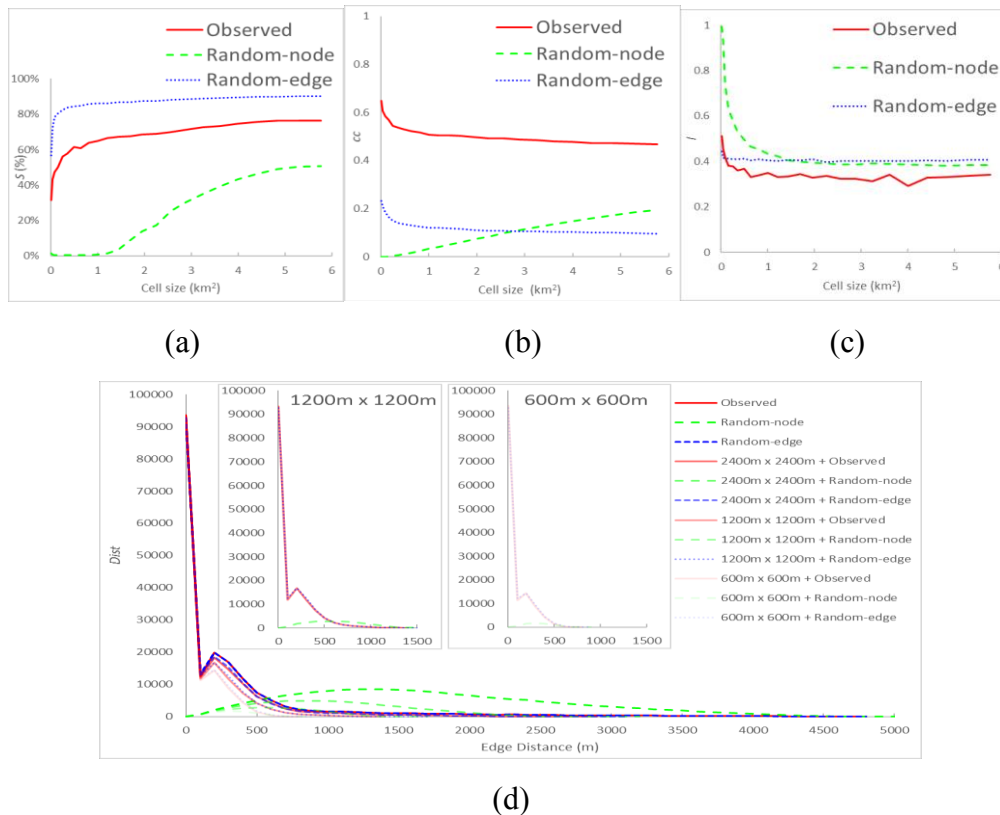
#### 3.3 Division of Networks

The study area is divided into 24 levels of regular grids ranging from 100m x 100m to 2400m x 2400m using 100m increments. During the division, those edges across boundaries of grid cells are eliminated, while those within the cells are kept. In this way, each cell contains its own network, called a 'unit network'. Properties of the unit networks are calculated using the four metrics and compared across scales and between the three networks.

## 4. Results and Discussion

### 4.1 Observed Network

The three network structure metrics ( $S$ ,  $cc$ , and  $l$ ) are plotted against the 24 cell sizes in Figures 1a-c. The spatial structure metric  $Dist$  for cell sizes of 600m x 600m, 1200m x 1200m, and 2400m x 2400m, are shown in Figure 1d. The three network metrics of the observed network are scale dependent, as their values vary with scale (Figures 1a-c). All three metrics show a characteristic scale of 0.6 km<sup>2</sup>. At scales finer than 0.6 km<sup>2</sup>, the value of the relative size of the largest component is low and that of the clustering coefficient and relative average path length are high. This indicates the unit networks are fragmented, locally clustered, inefficient, and consequently robust against disease dispersion. Beyond the characteristic scale, the values of the three metrics level off, behaving independently of scale change. The network structure at these coarser scales is more cohesive (higher  $S$ ), remains clustered (similar  $cc$ ), and is more efficient (lower  $l$ ), thus are more vulnerable to disease dispersion.



**Figure 1. The four metrics.** (a) The relative size of the largest component, (b) the clustering coefficient, (c) the relative average path length, and (d) the statistical distribution of edge distance (including that of the three original networks).

Unlike the network structure metrics, the spatial structure metric,  $Dist$ , appears to be scale independent. Although absolute quantities of edge distance change, the response of  $Dist$  to the scale change is invariant. It peaks at 0m that reflects the large number of 0-distance family

edges that are not affected by the division. The second peak between 0-800m is caused by a large supply of short co-worker edges. The 800m diminishing point is equivalent to the characteristic scale of  $0.6\text{km}^2$ . Such a short-edge dominant pattern indicates that the unit networks are highly clustered in space. Such networks facilitate short distance disease dispersion and lead to an epidemic surge in small areas.

## 4.2 Discussion and Conclusions

The two random networks help reveal the unique properties of the observed contact network. In terms of network structure, the observed network is sensitive to scale change at fine scales. In terms of spatial structure, the observed network is scale independent at all scales. Compared to the random-edge network that has a random network structure, the observed network tends to be 'dense', thus more cohesive, clustered, and efficient. Compared to the random-node network that has a random spatial structure, the observed network is spatially clustered.

The network structure and spatial structures are closely related. Spatially, the co-worker edges are mostly short (<800m). This can be attributed to urban dwellers' preference of being in close proximity to schools and workplaces (Ben-Akiva and Lerman 1985). Those who live close to each other tend to go to the same schools or workplaces. In other words, the closer ones are more connected. Findings of this research help deploy intervention strategies to spatially targeted areas.

## Acknowledgements

Research reported in this publication was supported in part by the National Institute of General Medical Sciences of the National Institutes of Health under Award Number R01GM108731. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

## References

- Barthélemy, M. (2011). Spatial networks. *Physics Reports*, 499, 1-101
- Ben-Akiva, M.E., & Lerman, S.R. (1985). *Discrete choice analysis: theory and application to travel demand*. Cambridge, MA: MIT press
- Bian, L. (2013). Spatial approaches to modeling dispersion of communicable diseases—A review. *Transactions in GIS*, 17, 1-17
- Bian, L., Huang, Y., Mao, L., Lim, E., Lee, G., Yang, Y., Cohen, M., & Wilson, D. (2012). Modeling individual vulnerability to communicable diseases: A framework and design. *Annals of the Association of American Geographers*, 102, 1016-1025
- Ferguson, N.M., Cummings, D.A., Cauchemez, S., Fraser, C., Riley, S., Meeyai, A., Iamsirithaworn, S., & Burke, D.S. (2005). Strategies for containing an emerging influenza pandemic in Southeast Asia. *Nature*, 437, 209-214
- Keeling, M.J., & Eames, K.T. (2005). Networks and epidemic models. *Journal of the Royal Society Interface*, 2, 295-307
- Newman, M.E.J. (2010). *Networks: An introduction*. Oxford: Oxford University Press
- Wu, J., & Wu, T. (2013). Ecological resilience as a foundation for urban design and sustainability. *Resilience in Ecology and Urban Design* (pp. 211-229). New York: Springer