

# Understanding spatial connectivity of individuals with non-uniform population density

BY PU WANG<sup>1,2</sup> AND MARTA C. GONZÁLEZ<sup>1,\*</sup>

<sup>1</sup>*Center for Complex Network Research, Department of Physics, Biology and Computer Science, Northeastern University, Boston, MA 02115, USA*

<sup>2</sup>*Center for Complex Network Research and Department of Physics, University of Notre Dame, Notre Dame, IN 46556, USA*

We construct a two-dimensional geometric graph connecting individuals placed in space within a given contact distance. The individuals are distributed using a measured country's density of population. We observe that while large clusters (group of individuals connected) emerge within some regions, they are trapped in detached urban areas owing to the low population density of the regions bordering them. To understand the emergence of a giant cluster that connects the entire population, we compare the empirical geometric graph with the one generated by placing the same number of individuals randomly in space. We find that, for small contact distances, the empirical distribution of population dominates the growth of connected components, but no critical percolation transition is observed in contrast to the graph generated by a random distribution of population. Our results show that contact distances from real-world situations as for WIFI and Bluetooth connections drop in a zone where a fully connected cluster is not observed, hinting that human mobility must play a crucial role in contact-based diseases and wireless viruses' large-scale spreading.

**Keywords:** hierarchical geometric graph; random geometric graph; population distribution; continuum percolation; percolation threshold; spatial networks

## 1. Introduction

For humans, population density is the number of people per unit area. Commonly, this may be calculated for a city, a country or the entire world. City population density is, however, heavily dependent on the definition of 'urban area' used: densities are often higher for the central municipality itself, compared with the more recently developed and administratively unincorporated suburban communities ([http://en.wikipedia.org/wiki/population\\_density](http://en.wikipedia.org/wiki/population_density)). Mobile phones provide the unique opportunity to estimate the distribution of population density with a spatial resolution never reached previously. By estimating the area of coverage of each mobile phone tower, we can measure the occupation of these areas, and estimate accurately the heterogeneous distribution of population

\*Author for correspondence ([marta.gonzalez.v@gmail.com](mailto:marta.gonzalez.v@gmail.com)).

One contribution of 14 to a Theme Issue 'Topics on non-equilibrium statistical mechanics and nonlinear physics'.

on a country scale. We placed individuals in space as measured from mobile phone usage and constructed a ‘spatial network’ to quantify the physical interconnectivity of individuals within a country. By these means, we investigated what the minimum contact distance is among individuals for which the entire population is fully connected, i.e. a giant cluster emerges. The understanding of the growth mechanism of a spatial network is crucial for diverse contact processes among humans; for example, disease transmissions such as SARS and influenza (Anderson *et al.* 2004; Hufnagel *et al.* 2004; Colizza *et al.* 2007; Riley 2007), wireless epidemics such as Bluetooth and WIFI router viruses (Hypponen 2006; Kleinberg 2007; Arenas *et al.* 2008; Hu *et al.* 2009) or information spreading.

A spatial network is a network embedded in Euclidean space: the interactions between the nodes depend on their spatial distance and usually take place between nearest neighbours (Medina *et al.* 2000; Yook *et al.* 2002; Kalapala *et al.* 2006). A random graph with a metric, often referred to as a random geometric graph (RGG), is constructed by assigning nodes uniformly in a  $d$ -dimensional space, linked whenever the distance is less than a certain value  $r$ ; this framework established the foundations of the continuum percolation theory (Gilbert 1961; Dall & Christensen 2002; Franceschetti *et al.* 2005), such that, for  $r > r_c$ , an unbounded connected component emerges. Although previous studies have provided mathematical methods to embed scale-free networks in a two-dimensional lattice (Rozenfeld *et al.* 2002; Nandi & Mana 2008) or in a continuous manifold (Herrmann *et al.* 2003), much less is known about the emergence of a connected component for geometric graphs with empirical distributions of population. The distribution of individuals in our measured sample is highly heterogeneous, with populated urban areas surrounded by less dense regions. The spatial random graph resulting from connecting individuals placed with such an empirical density of population is referred to here as a hierarchical geometric graph (HGG) in contrast to an RGG, which implies a uniform density of population.

This work is organized as follows. In §2, we characterize the data and present the simulation methods. In §3, we show the size of the largest cluster as a function of the contact distance of individuals and we present the distribution of cluster sizes for fixed values of contact distances. In §4, we compare the growth mechanism of the largest cluster on an HGG with an RGG. Conclusions are provided in §5.

## 2. Dataset and methods

The studied dataset was collected by a mobile phone carrier for billing and operational purposes during a one-month period. Privacy was ensured by identifying users with a security key (hash code). This dataset consisted of the full mobile communication pattern of 6.2 million users, recording the location of each user with a mobile tower-level resolution each time the user made or received a phone call or SMS, resulting in 1.1 billion location data points during our observational period recorded from over 10 000 mobile phone towers’ coordinates.

A Voronoi diagram (figure 1) around each tower offers a reasonable estimate of the tower’s service area, and its area is obtained by Delaunay triangulations (Fu *et al.* 2006). The Delaunay triangulation for a set  $P$  of points in the plane

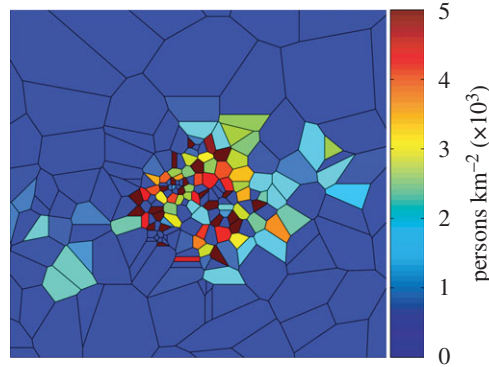


Figure 1. Diagram of the spatial area assigned to each tower coloured according to the density of users detected within it. The figure shows a subset of towers: each polygon is a Voronoi cell calculated over the entire set of tower's coordinates.

is a triangulation  $DT(P)$ , such that no point in  $P$  is inside the circumcircle of any triangle in  $DT(P)$ . The circumcircle centre of each Delaunay triangulation connecting a tower contributes a vertex of the tower's Voronoi cell. We calculated the Voronoi diagram for the set of mobile tower coordinates and found that the distribution of the tower areas followed a power law (Barabási & Albert 1999; Arenas *et al.* 2008) with an exponent  $k = -0.81$  in the region where the tower's area is smaller than  $100 \text{ km}^2$  (figure 2*a*), which implies a heterogeneous spatial distribution among mobile phone tower areas.

The mobile phone tower's ID is recorded each time a user has a phone communication through a mobile phone tower. For each user, we calculated the most frequently used tower during the entire month and assigned the user to this unique tower's area. By counting the number of users in each tower area, we obtained the population distribution at a country scale (figure 2*b*). We observed that it followed a power law with an exponent  $k = -1.16$  in the region where the population density was smaller than  $10\,000 \text{ persons km}^{-2}$ , revealing a hierarchical spatial distribution of the population, which is in contrast with the Poisson distribution observed in an RGG. Having the population for each tower, we constructed our HGG as follows: first, we placed each user randomly within the area of the corresponding tower; second, we connected two users if their distance was smaller than a contact distance  $r$ . By doing these two steps, we generated HGGs with different values of  $r$  (figure 2*c*). For relevant applications, the contact distance  $r$  can be the disease infection distance for contact-based diseases and the Bluetooth or connection range for mobile phones or wireless routers (Wang *et al.* 2009).

### 3. Analysis of cluster sizes

In this section, we calculate the size of the largest connected clusters (Cohen *et al.* 2000; Chen *et al.* 2007) of the HGG formed by all the users as a function of the contact distance  $r$ . A cluster is defined as a group of connected nodes (Stauffer & Aharony 1991). As we increased  $r$  from 10 to 50 m, large clusters gradually

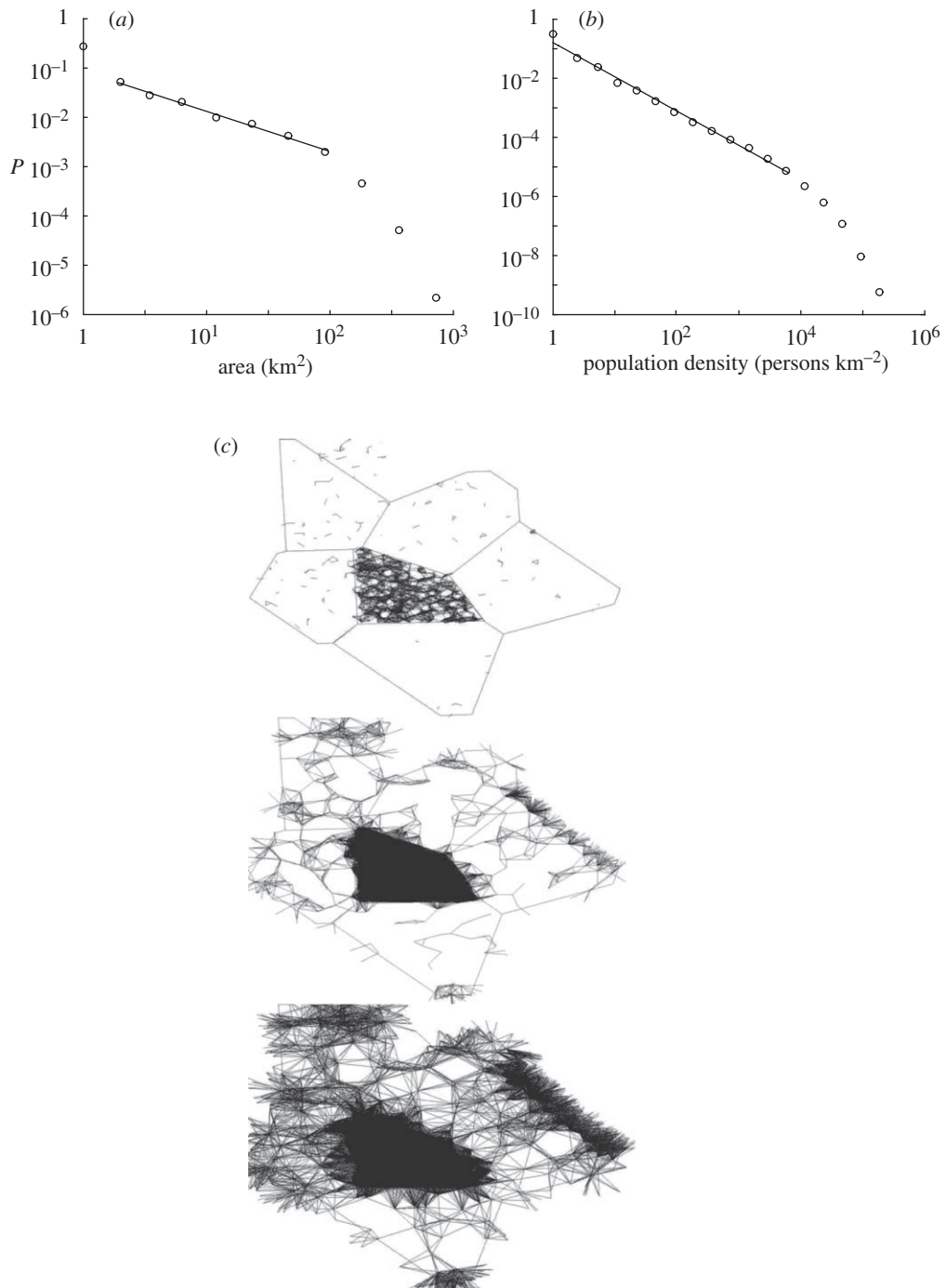


Figure 2. (a) Voronoi diagram area distribution:  $k \sim -0.81$  of the towers' service areas. The power law behaviour has been checked by statistical testing with an  $R^2 > 0.99$ . (b) Distribution of Voronoi diagrams' population density:  $k \sim -1.16$  in each tower's service area. The power law behaviour has been checked by statistical testing with an  $R^2 > 0.99$ . (c) The HGG with different contact distance  $r$ :  $r = 10, 30, 50$ .

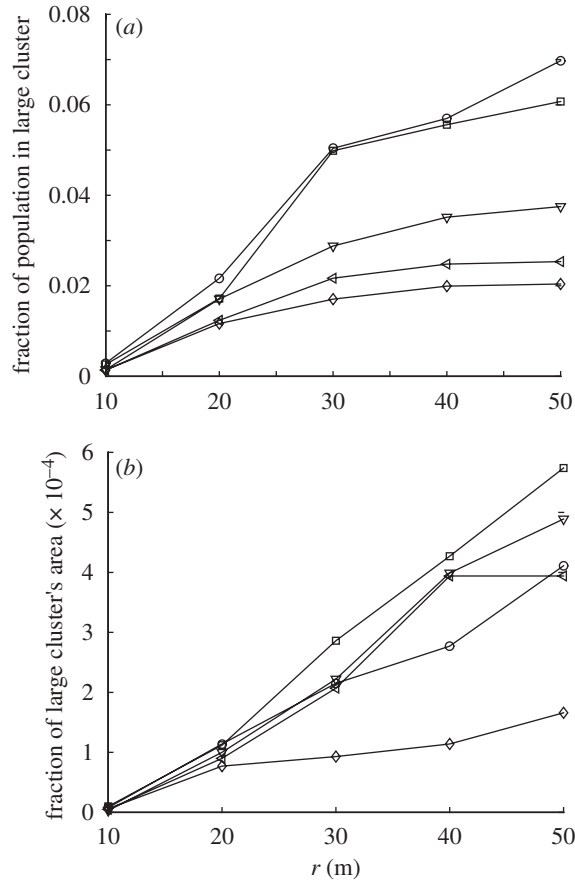


Figure 3. (a) The size of the top five large clusters as a function of the contact distance  $r$ . (b) The area covered by each of the top five large clusters as a function of the contact distance  $r$ . Circles, first; squares, second; inverse triangles, third; left-facing triangles, fourth; diamonds, fifth.

emerged. In figure 3a, we measured the size of the top five large clusters. Note that the largest cluster contains more than 5 per cent of the whole population at a contact distance as small as  $r = 30$  m. Next, we measured the areas covered by the top five large clusters by defining the area of a cluster as the total area of towers that contains at least one user within it. We observed that in HGG, while a large cluster can have a relatively large population, its area of coverage is very small. For example, the largest cluster at  $r = 50$  m has 7 per cent of the entire user population (figure 3a) and less than 0.06 per cent of the country's area (figure 3b). These results offer us a comprehensive picture of how the large clusters distribute in real space.

To get a better understanding of the growth mechanism of the HGG, we measured the distribution of the clusters' sizes (figure 4). We found that as we increased  $r$  from 10 to 50 m, the largest cluster's size gradually increased from ten thousand to half-million users, resulting from the connection of small isolated clusters.

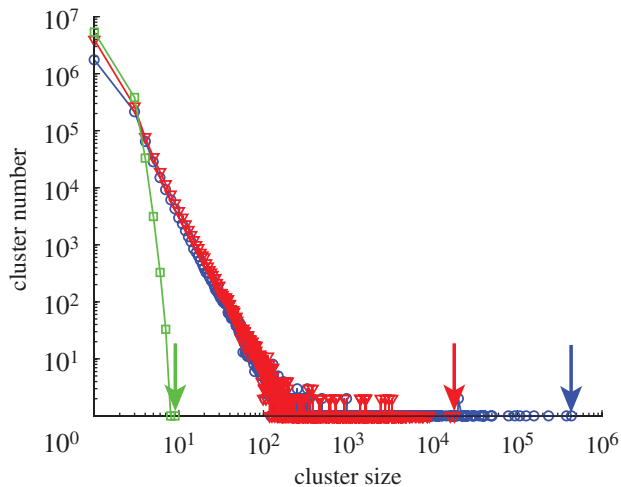


Figure 4. The cluster size distribution of HGG and RGG. The arrows point out the largest cluster's size for each condition. Circles, HGG,  $r = 50$  m; reverse triangles, HGG,  $r = 10$  m; squares, RGG,  $r = 50$  m.

For comparison, we constructed an RGG with the same population and average population density: we first calculated the average population density  $\rho_{\text{ave}}$  from the population data and multiplied this value by each tower area,  $A_{\text{tower}}$ , to get the number of users,  $N_{\text{tower}}$ , for each tower. Next, we randomly assigned  $N_{\text{tower}}$  users to each tower area and generated the RGG by connecting two users within contact distance. In the RGG, however, the largest component's size is only 9 at  $r = 50$  m, which is much smaller than the largest component for the  $r = 10$  m empirical case (figure 4). For contact distances smaller than 100 m, which have relevance to technological applications, we observed heterogeneous distribution of cluster size with the presence of large clusters for HGG when compared with RGG owing to the presence of small areas with high density of population.

In the next section, we study a large extent of contact distances. We analyse how a giant emerges for our HGG in contrast to the critical percolation transition observed in an RGG.

#### 4. The emergence of the giant cluster in a hierarchical geometric graph and random geometric graph

The simulation conditions to study the giant cluster emergence involve larger contact distances in which the simulations for millions of nodes are computationally untreatable (Dall & Christensen 2002). In order to compare the giant cluster growth's mechanisms for an HGG with those for an RGG, we rescaled the problem to a total population of 100 000 users. To generate the HGG, we rescaled each area keeping the same empirical density of population per tower area. We constructed an RGG with the same population and total area, placing the users at random positions within the space, and, as before, two users were connected within a given contact distance  $r$ .

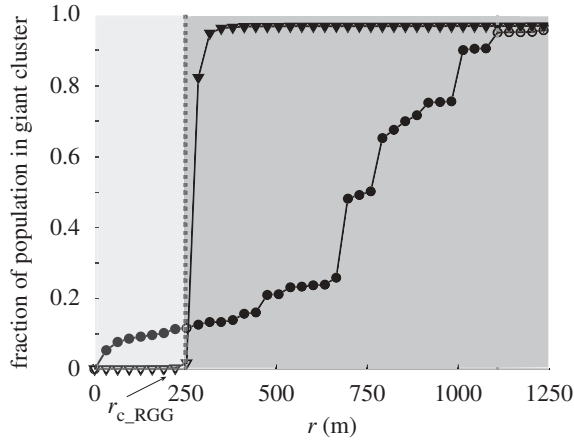


Figure 5. Giant cluster size as a function of contact distance  $r$  for HGG and RGG after shrinking the 6.2 million users to 100 000 users. Light-shaded region, zone 1; dark-shaded region, zone 2. Circles, HGG; triangles, RGG.

For each geometric graph, we calculated the size of the largest cluster versus  $r$ . As shown in figure 5, there is a clear difference in the emergence of the giant cluster for each graph. For the RGG, we measured the expected percolation transition beyond a certain threshold  $r_{c\_RGG}$ . In contrast, the growth of the giant cluster for HGG does not present a percolation transition, and it has a stair-like growth pattern caused by the connection of highly dense regions. In figure 5, we clearly see two zones separated by the RGG's percolation threshold: in zone 1, there is a relatively large cluster (about 0.1) for HGG at small  $r$ , while RGG remains totally disconnected. In zone 2, beyond  $r_{c\_RGG} \sim 250$  m RGG percolates, while the fully connected cluster for HGG is measured at much larger distances  $r \sim 1100$  m (figure 5). The reason is the population's hierarchical spatial distribution (figures 1 and 2b), which helps the emergence of large clusters within small highly populated areas, but also blocks the emergence of a giant cluster. As shown in figures 4 and 5, it is expected that for empirical population distributions and realistic contact distances (as for WIFI routers or Bluetooth ranges) scenarios drop in zone 1, in which clusters are much bigger and emerge easier for HGG than for RGG.

## 5. Conclusions

This work offers a spatial picture of a geometric graph with a non-uniform distribution of population, here named a hierarchical geometric graph (HGG): large clusters emerge in urban areas surrounded by many small clusters in less dense areas. This situation blocks the critical emergence of a giant cluster for HGG, in contrast to the results for RGGs. As we observed in §4, for distances smaller than 250 m, which is the critical contact distance for the RGG percolation transition, HGG has larger clusters with up to 7 per cent of the entire population, while RGG remains completely disconnected. The growth mechanism of the giant



cluster for HGG is stair like and governed by the connection of dense small regions; a fully connected cluster is measured beyond 1100 m. Our results offer evidence that, under heterogeneous density of population and realistic contact distances of the order of a few metres, human mobility (Brockmann *et al.* 2006; Gonzalez *et al.* 2008) must play a crucial role for large-scale epidemics in the world (Anderson *et al.* 2004; Hufnagel *et al.* 2004; Colizza *et al.* 2007; Riley 2007). We provide here a numerical analysis of the spatial network formation under an empirical distribution of population, an analytical formulation of the results presented here remains open to further studies.

We thank J. Park and C. Song for discussions and comments on the manuscript. This work was supported by the James S. McDonnell Foundation 21st Century Initiative in Studying Complex Systems, the National Science Foundation within the DDDAS (CNS-0540348), ITR (DMR-0426737) and IIS-0513650 programmes, the Defense Threat Reduction Agency Award HDTRA1-08-1-0027 and the US Office of Naval Research Award N00014-07-C.

## References

- Anderson, R. M., Fraser, C., Ghani, A. C., Donnelly, C. A., Riley, S., Ferguson, N. M., Leung, G. M., Lam, T. H. & Hedley, A. J. 2004 Epidemiology, transmission dynamics and control of SARS: the 2002–2003 epidemic. *Phil. Trans. R. Soc. Lond. B* **359**, 1091–1105. (doi:10.1098/rstb.2004.1490)
- Arenas, A., Díaz-Guilera, A., Kurths, J., Moreno, Y. & Zhou, C. 2008 Synchronization in complex networks. *Phys. Rep.* **469**, 93–153. (doi:10.1016/j.physrep.2008.09.002)
- Barabási, A.-L. & Albert, R. 1999 Emergence of scaling in random networks. *Science* **286**, 509–512. (doi:10.1126/science.286.5439.509)
- Brockmann, D., Hufnagel, L. & Geisel, T. 2006 The scaling laws of human travel. *Nature* **439**, 462–465. (doi:10.1038/nature04292)
- Chen, Y., Gerald, P., Reuven, C., Reuven, S., Borgatti, S. P., Liljeros, F. & Stanley, H. E. 2007 Percolation theory applied to measures of fragmentation in social networks. *Phys. Rev. E* **75**, 046107. (doi:10.1103/PhysRevE.75.046107)
- Cohen, R., Erez, K., ben-Avraham, D. & Havlin, S. 2000 Resilience of the Internet to random breakdowns. *Phys. Rev. Lett.* **85**, 4626–4628. (doi:10.1103/PhysRevLett.85.4626)
- Colizza, V., Barrat, A., Barthelemy, M., Valleron, A.-J. & Vespignani, A. 2007 Modeling the worldwide spread of pandemic influenza: baseline case and containment interventions. *PLoS Med.* **4**, 95–110. (doi:10.1371/journal.pmed.0040013)
- Dall, J. & Christensen, M. 2002 Random geometric graphs. *Phys. Rev. E* **66**, 016121. (doi:10.1103/PhysRevE.66.016121)
- Franceschetti, M., Booth, L., Cook, M., Meester, R. & Bruck, J. 2005 Continuum percolation with unreliable and spread out connections. *J. Stat. Phys.* **118**, 721–734. (doi:10.1007/s10955-004-8826-0)
- Fu, T., Yin, X. & Zhang, Y. 2006 Voronoi algorithm model and the realization of its program. *Comput. Simul.* **23**, 89–91.
- Gilbert, E. N. 1961 Random plane networks. *J. SIAM* **9**, 533–543.
- González, M. C., Hidalgo, C. A. & Barabási, A.-L. 2008 Understanding individual human mobility patterns. *Nature* **435**, 779–782. (doi:10.1038/nature06958)
- Herrmann, C., Barthelemy, M. & Provero, P. 2003 Connectivity distribution of spatial networks. *Phys. Rev. E* **68**, 026128. (doi:10.1103/PhysRevE.68.026128)
- Hu, H., Myers, S., Colizza, V. & Vespignani, A. 2009 WiFi networks and malware epidemiology. *Proc. Natl Acad. Sci. USA* **106**, 1318–1323. (doi:10.1073/pnas.0811973106)
- Hufnagel, L., Brockmann, D. & Geisel, T. 2004 Forecast and control of epidemics in a globalized world. *Proc. Natl Acad. Sci. USA* **101**, 15124–15129. (doi:10.1073/pnas.0308344101)
- Hyyonen, M. 2006 Malware goes mobile. *Sci. Am.* **295**, 70–77.



- Kalapala, V., Sanwalani, V., Clauset, A. & Moore, C. 2006 Scale invariance in road networks. *Phys. Rev. E* **73**, 026130. (doi:10.1103/PhysRevE.73.026130)
- Kleinberg, J. 2007 The wireless epidemic. *Nature* **449**, 287–288. (doi:10.1038/449287a)
- Medina, A., Matta, I. & Byers, J. 2000 On the origin of power laws in internet topologies. *ACM SIGCOMM Comput. Commun. Rev.* **30**, 18–28. (doi:10.1145/505680.505683)
- Nandi, A. K. & Manna, S. S. 2008 Construction of spatial scale-free networks on a lattice by using a democratic rule. *J. Stat. Mech.* 07009. (doi:10.1088/1742-5468/2008/07/P07009)
- Riley, S. 2007 Large-scale spatial-transmissions models of infectious disease. *Science* **316**, 1298–1301. (doi:10.1126/science.1134695)
- Rozenfeld, A. F., Cohen, R., Ben-Avraham, D. & Havlin, S. 2002 Scale-free networks on lattices. *Phys. Rev. Lett.* **89**, 218701. (doi:10.1103/PhysRevLett.89.218701)
- Stauffer, D. & Aharony, A. 1991 *An introduction to percolation theory*. Boca Raton, FL: CRC Press.
- Wang, P., González, M. C., Hidalgo, C. A. & Barabási, A.-L. 2009 Understanding the spreading patterns of mobile phone viruses. *Science* **324**, 1071–1076.
- Yook, S. H., Jeong, H. & Barabási, A.-L. 2002 Modeling the internet's large-scale topology. *Proc. Natl Acad. Sci. USA* **99**, 13 382–13 386. (doi:10.1073/pnas.172501399)

# Understanding spatial connectivity of individuals with non-uniform population density

Pu Wang and Marta C. González

*Phil. Trans. R. Soc. A* 2009 **367**, 3321-3329  
doi: 10.1098/rsta.2009.0089

---

## References

**This article cites 23 articles, 7 of which can be accessed free**  
<http://rsta.royalsocietypublishing.org/content/367/1901/3321.full.html#ref-list-1>

## Rapid response

**Respond to this article**  
<http://rsta.royalsocietypublishing.org/letters/submit/roypta;367/1901/3321>

## Subject collections

Articles on similar topics can be found in the following collections

[mathematical physics](#) (79 articles)  
[statistical physics](#) (16 articles)

## Email alerting service

Receive free email alerts when new articles cite this article - sign up in the box at the top right-hand corner of the article or click [here](#)

---

To subscribe to *Phil. Trans. R. Soc. A* go to:  
<http://rsta.royalsocietypublishing.org/subscriptions>

---