An Energy-Based Model for Spatial Social Networks

Alberto Antonioni[1], Mattia Egloff[2], Marco Tomassini[1]

[1]Information Systems Department, Faculty of Business and Economics,
[2]Section of Computer Science and Mathematical Methods, Faculty of Arts and Humanities,
University of Lausanne, Switzerland

Abstract

In the past decade, thanks to abundant data and adequate software tools, complex networks have been thoroughly investigated in many disciplines. Most of this work has dealt with networks in which distances do not have physical meaning and are just dimensionless quantities measured in terms of edge hops. However, in many cases the physical space in which networks are embedded and the actual distances between nodes are important, such as in geographical and transportation networks. The Random Geometric Graph (RGG) is a standard spatial network model that plays a role for spatial networks similar to the one played by the Erdös–Rényi random graph for relational ones. In this work we present an extension of the RGG construction to define a new model to build bi-dimensional spatial networks based on energy as realistic constraint to create the links. The constructed networks have several properties in common with those of actual social networks.

Introduction

Social networks arise in a wide range of contexts and really pervasive in our society. Examples range from corporate partnership connections, scientific collaborations, sexual contacts, film actors networks, to Facebook and other online social networks among others. In recent years much attention has been given to model these networks in order to gain a better understanding of their general structures and their functions like information flow, locating individuals, disease spread, etc. There is an increase in the number of network models in the literature (Toivonen et al., 2009) but, although general features common to all social networks are reproduced, such as the typically high clustering, none of them can represent all the typical characteristics of social networks in a realistic way. This is due to the fact that all these networks have formed and grow in ways that are similar but not identical. In other words, each actual network is an instance of a class of possible realizations and its particular structure depends on its particular history, frozen structures, dynamics, and many other factors.

It is generally believed that social networks possess the following main features:

- *positively skewed degree distribution*: the majority of agents have relatively small degrees, while a small number of agents may have large degrees.

- *high average clustering coefficient* $\overline{C}$: the conditional probability that two neighbors of an agent will be connected is much higher that what would be expected in a sparse random graph.

- *positive degree correlations*: the degrees of the neighbors are not independent and are similar on average.

- *small average shortest path length* $\overline{L}$: $\overline{L} \propto \log(N)$ i.e. it is rather small compared to the network size N.

- *existence of community structure*: clusters of agents which are highly connected within themselves but loosely connected to other subgroups.

In the last decade social networks as a topology-free relational graph structure have been much studied. A comprehensive literature would be too long to mention, but review articles and books contain a wealth of information on them e.g. (Boccaletti et al., 2006; Newman, 2010). Relational networks are those in which actual distances do not count and path lengths are computed by simply adding one for each link in the path. Social networks such as coauthorship networks are usually taken to be relational. However, Euclidean space is an important factor in many networks, for example transportation and communication networks among others are of this kind (see (Barthélemy, 2011) for a recent good review of the field). But spatial aspects can also be important for social networks. For example, while two Facebook friends might live in the US and Europe respectively and still be represented by a standard link in the net, it is also observed that many links will be among people that are geographically close. Thus, spatial considerations may play a role in social networks too. Therefore, while purely relational social network models are important and have been studied in depth (see e.g. Toivonen et al. (2006); Vázquez (2003); Kumpula et al. (2007); Catanzaro et al. (2004)), their spatial aspects are much less known (but see the following
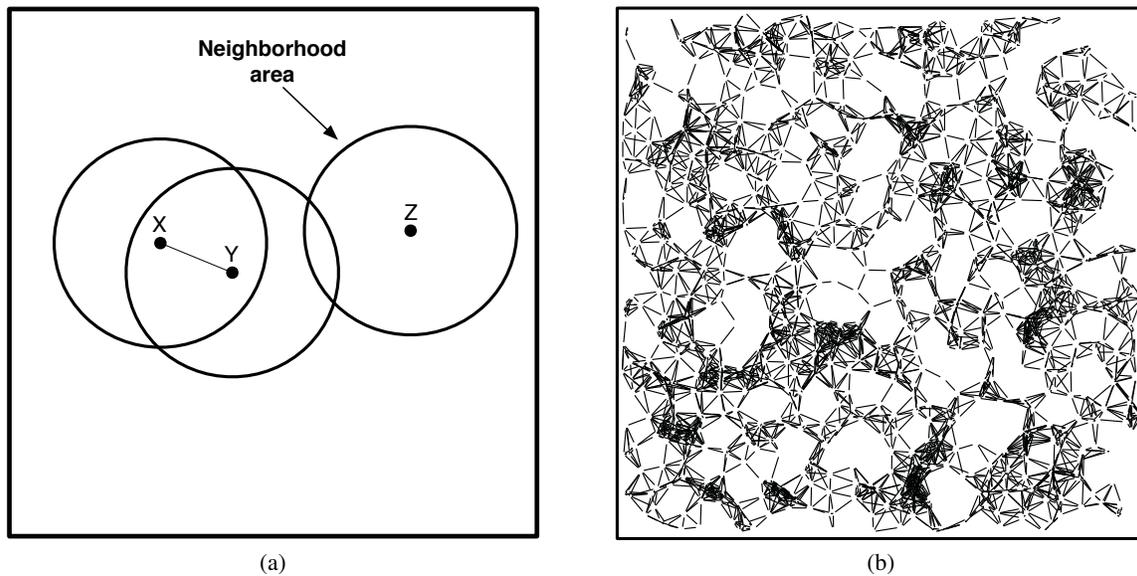
(a)   (b)

Figure 1: (a): Neighborhood areas of nodes X, Y and Z. (b): An example of RGG with $N = 1000$ and $R = 0.056$, and average degree $\bar{k} = 10$ (for illustrative purposes the unit square is bounded).

works Boguñá et al. (2004), Wong et al. (2006) and Serrano et al. (2008) for some recent attempts). Our model is very simple and it is intended to be only a first step toward more realistic ones. It is based on the concept that each agent is initially given a constant amount of energy to be used to establish links with other agents. The spatial bias is given by the fact that links cost less if they are made with agents that are physically closer. The model gives rise to networks that have high clustering coefficient, positive degree assortativity, and modularity due to the appearance of communities. The degree distribution is rather peaked but the model could be easily modified to produce broader distributions.

The article has the following organization. In the next section we give a brief introduction to random geometric graphs, a spatial model that will be needed in the sequel. Next we describe our own model of a social spatial network. The following section presents and discusses the main numerical results, and we then give our conclusions.

Random Geometric Graph

The Random Geometric Graph (RGG) is obtained when the points located in the plane are connected according to a given geometric rule. The simplest rule is a proximity rule which states that nodes only within a certain distance are connected. There is an extensive mathematical literature on geometric graph and the random case was studied by physicists in the context of continuum percolation (Penrose, 2003; Dall and Christensen, 2002; Barthélemy, 2011).

In this work we refer to the following construction process for a RGG with N nodes and radius R:

- the N nodes are placed on the unit space $\Omega \in \mathbb{R}^2$ with

uniform distribution,

- an edge is created for every pair of nodes whose distance is $r < R$. The distance is given by the standard Euclidean metric on $\mathbb{R}^2$.

Furthermore, we shall assume that the unit space Ω is the square $[0, 1]^2$ with cyclic boundary conditions (torus). In Fig. 1a three nodes X, Y, Z and their neighborhood areas are depicted. Nodes X and Y are connected through an edge since they are within the neighborhood area of each other respectively, while Z is not connected to Y even if it is sharing a common area with it. Figure 1b shows a RGG realization with $N = 1000$ and $R = 0.056$; in this case, for illustrative purposes, Ω is a bounded unit square.

It is also possible to adopt different shapes of neighborhood area generated according to other metrics. For example, the Manhattan distance is sometimes used to model mobility networks (Glauche et al., 2003; Di Crescenzo et al., 2012). The general properties of these networks are very close to those using the more common Euclidean distance, which are the ones we describe here.

The average degree $\bar{k}$ of a RGG can be easily estimated by the formula $\bar{k} = \rho V$, where ρ is the node density, representing the number of nodes within a unit space, and V is the neighborhood area. In this case $\rho = N$, since Ω is an unit space, and $V = \pi R^2$. In conclusion, $\bar{k} = \pi N R^2$.

The degree distribution of RGGs with a sufficiently large number of nodes can be estimated by the Poisson distribution with parameter $\lambda = \bar{k}$ (Dall and Christensen, 2002).

The average clustering coefficient is given averaging on all node's individual clustering coefficients (Newman,

2010). This property on RGGs was extensively studied in the work of Dall and Christensen (2002), in which they have found the law for the average clustering coefficient as a function of the dimension of the space. Here the dimension is equal to two, and it is possible to demonstrate that the average clustering coefficient tends to $1 - \frac{3\sqrt{3}}{4\pi} \sim 0.5865$, for large values of N and for all 2-dimensional RGGs in the Euclidean space. This important result depends on the particular construction of RGGs. The average clustering coefficient tends to the ratio of the average shared neighborhood area of two connected nodes and the whole neighborhood area. It is clear that changing the radius R this fraction maintains the same value.

Due to its construction process, in a RGG there is positive degree-degree correlation. This property is commonly detected studying the *assortativity coefficient*, which is the Pearson correlation coefficient of degree between pairs of connected nodes (Boccaletti et al., 2006). In the recent work of Antonioni and Tomassini (2012), it has been demonstrated that the assortativity coefficient tends to the average clustering coefficient value for any d-dimensional RGG. Many other properties of RGGs have been studied in Penrose (2003).

Energetic Spatial Network

In order to construct more realistic social networks with spatial structure, we now consider the following two realistic assumptions for a spatial social network (see Fig. 2):

- *limited neighborhood*: a given node may create links only within the set of nodes in its neighborhood area given by the radius R.

- *distance cost*: creating a long link is more costly in terms of energy expenditure than creating one closer to the focal node.
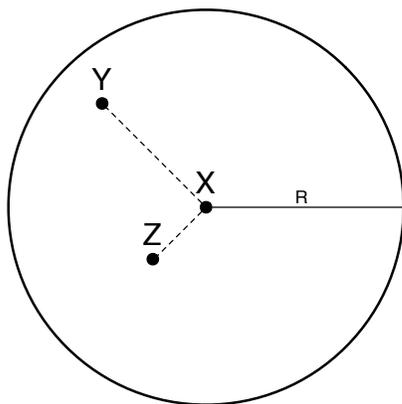

Figure 2: Node X can be linked only to nodes within its neighborhood area given by radius R. In this case, creating a link with Y is more costly than creating one with Z.

In this model we assume that *energy*, which is constant and the same for each node, is a resource provided to nodes in order to create and maintain their links. We shall call networks constructed according to this model as Energetic Spatial Networks (ESNs). Our hypothesis is thus that each node has limited energy available to create its acquaintances, which are increasingly costly with increasing physical distance. This feature can be also assumed in actual social networks, in which real distance may play an important role in order to establish a connection. It is indeed reasonable to think that in order to minimize the efforts and to maintain a social tie most individuals tend to be connected with their spatial neighbors, at least in social networks that are not fully mediated by communication devices. The present model creates a static network; dynamical aspects might be included in the future by requiring that maintaining links through time also costs a certain amount of energy.

The construction process to build ESNs with N nodes, radius R, and initial energy E, can be summarized as follows:

1. The N nodes are randomly placed with uniform distribution on the unit space $\Omega \in \mathbb{R}^2$. All nodes have the same initial energy equal to E.

2. A node X is picked uniformly at random in the set of all nodes, and Y is chosen uniformly at random in the set of nodes whose distance from X is $r < R$.

3. An edge between X and Y is created if d_{XY}, which is the Euclidean distance between X and Y, is less than E_X (residual energy of X) and E_Y. If the edge is created, then the residual energies of X and Y are both decremented by d_{XY}.

4. Steps 2–3 are repeated until no more edges can be created according to the linking rule.

The unit space Ω can be seen, similarly to the RGGs construction, as the square $[0,1]^2$ with cyclic boundary conditions (torus). It is rather clear that this construction process produces RGGs for $E \to \infty$, while, for $\{R, E\} \to \infty$ complete graphs are obtained.

Results

Figure 3 shows empirical values of the normalized average degree, average clustering coefficient, and assortativity coefficient of a realization of an ESN as a function of the initial energy E (Fig. 3a), and of the radius R (Fig. 3b). The normalized average degree k_{norm} shown in Figs. 3 is obtained by dividing the average degree $\bar{k}$ of an ESN by the average degree of a RGG with the same radius R. This means that for $\bar{k}_{norm} \to 1$ the ESN can be approximated by a RGG. In Fig. 3a, the parameters of an ESN are $N = 10000$ and $R = 0.04$. We can observe that for high values of the initial energy E (> 1.5), ESN features become closer to those of a
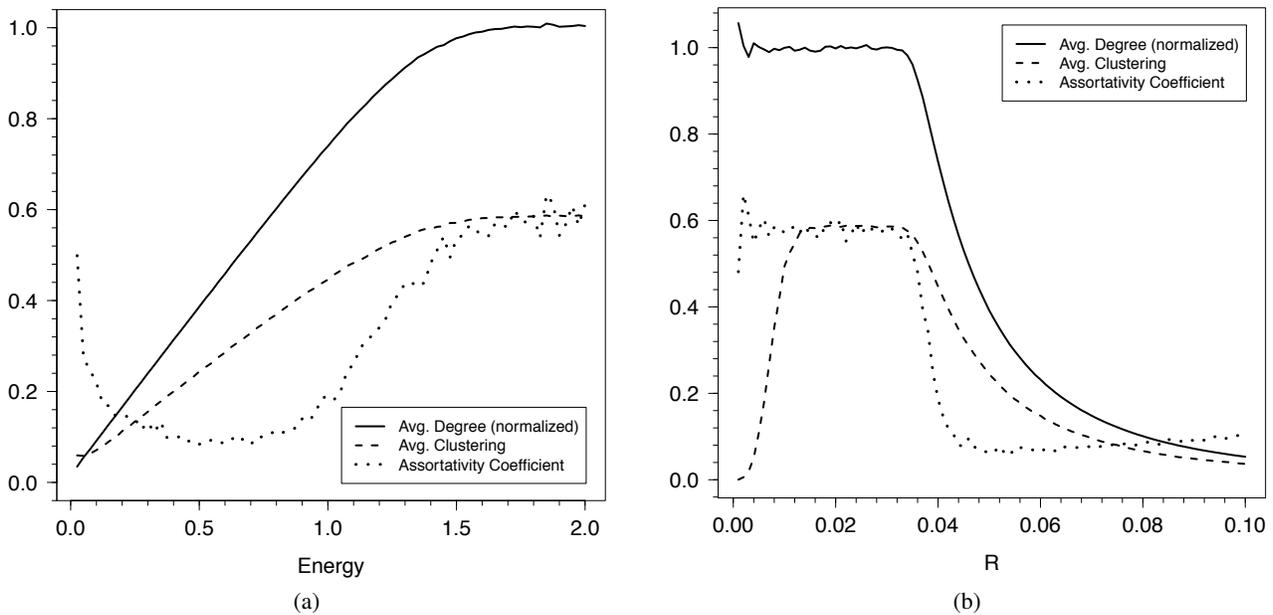
(a)

(b)

Figure 3: Numerical values of normalized average degree, average clustering coefficient, and assortativity coefficient of an ESN as a function of : (a) the initial energy E and (b) the radius R. (a): The other ESN parameters are $N = 10000$ and $R = 0.04$. The average degree is normalized dividing by 50, which represents the average degree value of a RGG with parameters $N = 10000$ and $R = 0.04$. (b): The other ESN parameters are $N = 10000$ and $E = 1.0$. The average degree value is normalized dividing by $10000\pi R^2$, which represents the average degree of a RGG with $N = 10000$.
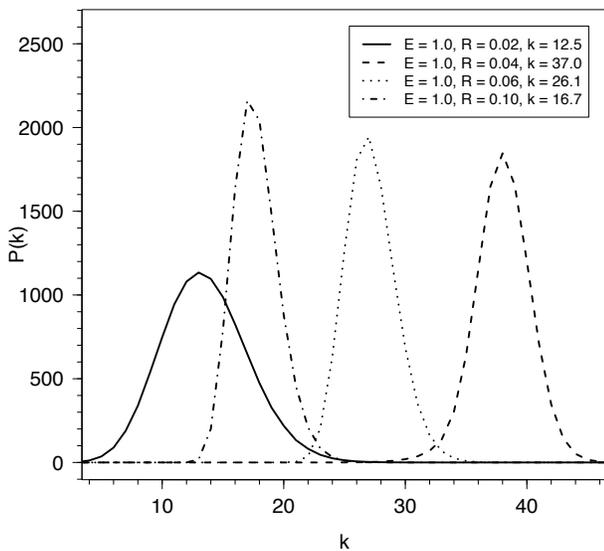

Figure 4: Degree distribution functions for ESNs with $N = 10000$ and $E = 1.0$. The other parameter, the radius R, assumes the following values: $0.02, 0.04, 0.06$ and 0.1.

RGG since in this case the energy is more than enough to allow the creation of all possible links within radius R. In fact, average clustering and assortativity coefficients both tend to the characteristic RGG value 0.5865. Therefore, with this model, it is possible to select a value for E that will produce a desired high clustering coefficient for the ESN. This is not possible with the standard RGG model, which converges by construction to a fixed value.

In Fig. 3b, the parameters of the ESN are $N = 10000$ and $E = 1$. For small values of R (< 0.04), ESNs can be considered as RGGs, because the radius is rather small compared to the initial energy and the nodes can build all the possible links in their neighborhood areas. For larger values of R the network becomes more interesting. The clustering coefficient and the assortativity coefficient both decrease and thus suitable values of R for a social network are around $0.04 - 0.06$. The normalized average degree is approximately 1 for $R < 0.04$, but then tends to zero for larger values of R. This means that ESNs become sparser than a RGG with the same radius R and thus R should not go beyond 0.06 for a realistically connected network.

For $E = 1$, Figure 4 depicts the degree distribution functions of four realizations of ESNs with different values of R. The thick curve with $R = 0.02$ gives rise to a standard RGG as we have seen above (see also Fig. 3b). The other curves correspond to three ESNs and are rather peaked. Although relational social networks usually have rather broad
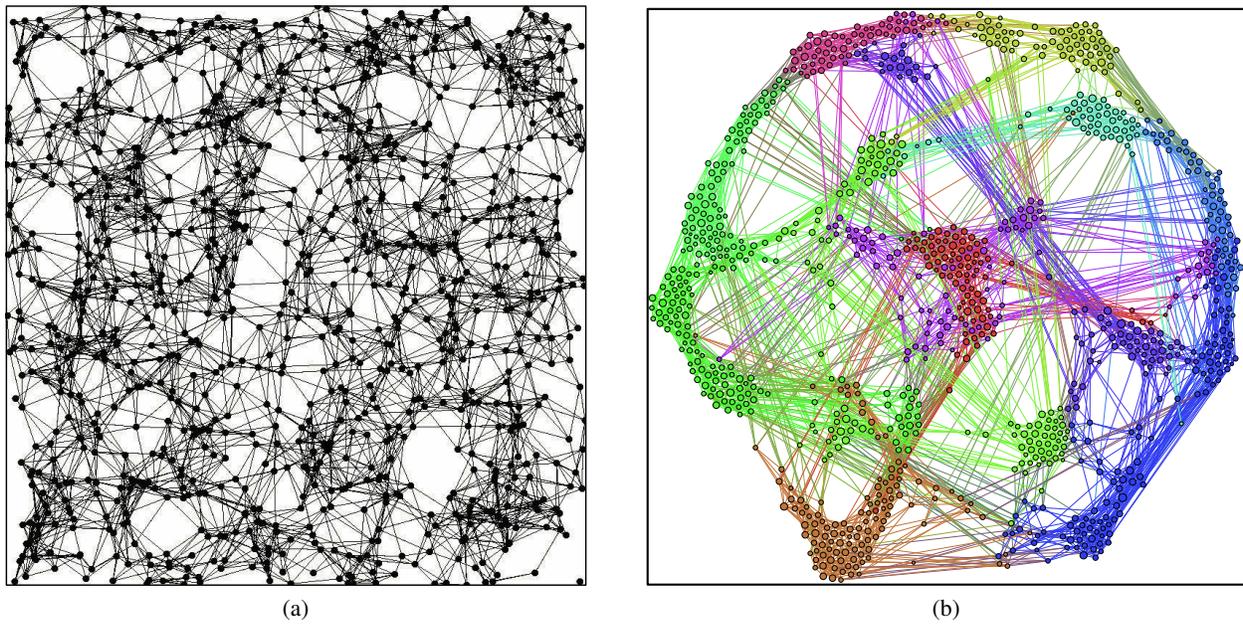
Figure 5: Realization of an ESN with $N = 1000$, $E = 0.64$, and $R = 0.1$. The average degree is about $\bar{k} = 10$, while the average clustering coefficient is equal to 0.205 and the assortativity coefficient is 0.135. The modularity is $M = 0.758$ (Blondel et al., 2008). (a) Spatial representation, and (b) graphical representation given by the OpenOrd algorithm (Martin et al., 2011).

degree distribution functions (Newman, 2001a,b; Barabási et al., 2002; Tomassini and Luthi, 2007), spatial constraints do not allow agents to have too many links to other nodes. The effect is particularly evident in technological and transportation networks where hard constraints such as rail crossing or economic factors in cable length in power grids, for instance, set well defined limits to the network's connectivity (see Barthélemy (2011) and references therein). The above factors arising from spatial physical constraints are less important for social acquaintances but, to some extent, they also influence the spatial structure of social networks. Indeed, social networks can actually be seen as a mix of relational and spatial factors. However, part of the observed effect on the degree distribution function is certainly attributable to the fact that we give the same constant amount of initial energy to all nodes. A simple improvement of the model would consist in attributing energy according to a more complex distribution such as a power-law or another suitable function.

The presence of communities, which are clusters of densely connected nodes, is a common feature of all social networks (Newman, 2010). Communities may arise in many ways, for instance people having common interests, people sharing the same culture or religion, going to the same school, living in the same area, as is the case in our model networks, and many others. We have studied the cluster structure of ESNs using one of the several heuristic community detection algorithms (Blondel et al., 2008)

and we have found that ESNs present several communities. Figure 5 shows an ESN with $N = 1000$, $E = 0.64$, and $R = 0.1$. Figure 5a is a true geographical representation of the network and, although it might superficially look very similar to the RGG image of Fig. 1b, it possesses features such as the presence of longer links and the lack of linking to all the neighbors falling into the disk-shaped neighborhood area for a given node. Instead, Fig. 5b shows with different colors the communities found in the network by the community detection algorithm. To improve the rendering, Fig. 5b does not take into account the Euclidean metrics in the node and link disposition. Modularity M (Newman, 2006) for the graph is rather high at a value of $M = 0.758$ which means that the communities are rather well defined. Although the M measure is not without drawbacks, it still provides a rather clear-cut indication of the significance of the network's community structure.

Conclusion

In this article we have proposed an original model for the construction of social networks having a spatial dimension. We started from the random geometric graph model and we added a few ingredients in order to generate networks that possess most of the statistical features shown by actual spatial social networks. The main idea is to attribute a limited amount of energy to nodes, the same for all of them. Nodes can spend this resource to link to other nodes as a function of their Euclidean distance, longer links being more expensive

than shorter ones. In this way we obtain networks that do indeed look similar in many ways to actual ones from the point of view of their statistical features. In particular, the modeled networks have high clustering, positive degree correlation, and the presence of community structure. Moreover, properties such as clustering and assortativity can be tuned to some extent by changing the parameters E and R. The degree distribution function is peaked due, in part, to spatial constraints but especially because of the homogeneous distribution of energy and radii among the nodes. This consideration suggest ideas for further research with the purpose of getting more realistic social networks. For example, one could assume that nodes are not be placed at random on the space, in order to form even more clustered networks and have more community structures. Furthermore, it is reasonable to consider non-constant distributions of the energy among the nodes which would probably produce some more connected individuals and thus a broader degree distribution. The linking process is bilateral in the present version, i.e. both partners must pay the same amount of energy to create the connection. One-way links could also be considered and the model could be extended to make it dynamical allowing for link suppression as well as link formation.

Acknowledgments

References

Antonioni, A. and Tomassini, M. (2012). Degree correlations in random geometric graphs. *Physical Review E*, 86:037101.

Barabási, A.-L., Jeong, H., Néda, Z., Ravasz, E., Schubert, A., and Vicsek, T. (2002). Evolution of the social network of scientific collaborations. *Physica A*, 311:590–614.

Barthélemy (2011). Spatial networks. *Physics Reports*, 499:1–101.

Blondel, V. D., Guillaume, J.-L., Lambiotte, R., and Lefebvre, E. (2008). Fast unfolding of communities in large networks. *Journal of Statistical Mechanics: Theory and Experiment*, 2008:P10008.

Boccaletti, S., Latora, V., Moreno, Y., Chavez, M., and Hwang, D.-U. (2006). Complex networks: Structure and dynamics. *Physics Reports*, 424:175–308.

Boguñá, M., Pastor-Satorras, R., Díaz-Guilera, A., and Arenas, A. (2004). Models of social networks based on social distance attachment. *Physical Review E*, 70:056122.

Catanzaro, M., Caldarelli, G., and Pietronero, L. (2004). Assortative model for social networks. *Physical Review E*, 70(3):037101.

Dall, J. and Christensen, M. (2002). Random geometric graphs. *Physical Review E*, 66:016121.

Di Crescenzo, G., Kondareddy, Y., and Zhang, T. (2012). Concrete synthetic modeling of vehicular networks as random geometric graphs. In *Communications (ICC), 2012 IEEE International Conference on*, pages 1–5. IEEE.

Glauche, I., Krause, W., Sollacher, R., and Greiner, M. (2003). Continuum percolation of wireless ad hoc communication networks. *Physica A: Statistical Mechanics and its Applications*, 325:577–600.

Kumpula, J. M., Onnela, J.-P., Saramäki, J., Kaski, K., and Kertész, J. (2007). Emergence of communities in weighted networks. *Physical Review Letters*, 99:228701.

Martin, S., Brown, W. M., Klavans, R., and Boyack, K. W. (2011). Openord: an open-source toolbox for large graph layout. In *IS&T/SPIE Electronic Imaging*, pages 786806–786806. International Society for Optics and Photonics.

Newman, M. E. (2006). Modularity and community structure in networks. *Proceedings of the National Academy of Sciences*, 103(23):8577–8582.

Newman, M. E. J. (2001a). Scientific collaboration networks. I. network construction and fundamental results. *Physical Review E*, 64:016131.

Newman, M. E. J. (2001b). Scientific collaboration networks. II. shortest paths, weighted networks, and centrality. *Physical Review E*, 64:016132.

Newman, M. E. J. (2010). *Networks: An Introduction*. Oxford University Press, Oxford, UK.

Penrose, M. (2003). *Random Geometric Graphs*. Oxford University Press, Oxford, UK.

Serrano, M. A., Krioukov, D., and Boguná, M. (2008). Self-similarity of complex networks and hidden metric spaces. *Physical Review Letters*, 100:078701.

Toivonen, R., Kovanen, L., Kivelä, M., Onnela, J.-P., Saramäki, J., and Kaski, K. (2009). A comparative study of social network models: Network evolution models and nodal attribute models. *Social Networks*, 31:240–254.

Toivonen, R., Onnela, J.-P., Saramäki, J., Hyvönen, J., and Kaski, K. (2006). A model for social networks. *Physica A: Statistical Mechanics and its Applications*, 371:851–860.

Tomassini, M. and Luthi, L. (2007). Empirical analysis of the evolution of a scientific collaboration network. *Physica A*, 385:750–764.

Vázquez, A. (2003). Growing network with local rules: Preferential attachment, clustering hierarchy, and degree correlations. *Physical Review E*, 67:056104.

Wong, L. H., Pattison, P., and Robins, G. (2006). A spatial model for social networks. *Physica A: Statistical Mechanics and its Applications*, 360(1):99–120.