UNIVERSIDADE ESTADUAL DE CAMPINAS
INSTITUTO DE MATEMÁTICA, ESTATÍSTICA E COMPUTAÇÃO CIENTÍFICA
DEPARTAMENTO DE MATEMÁTICA APLICADA

Alan Albert Piovesana

An introduction to spatial networks: network generating models

Campinas
07/07/2021

Alan Albert Piovesana

An introduction to spatial networks: network generating models*

Monografia apresentada ao Instituto de Matemática, Estatística e Computação Científica da Universidade Estadual de Campinas como parte dos requisitos para obtenção de créditos na disciplina Projeto Supervisionado, sob a orientação do(a) Prof. Dr. José Antônio Brum.

*This work was part of a larger undergraduate research project sponsored by CNPQ-PIBIC, and started in Sep/2020.

Resumo

Este texto é uma breve revisão bibliográfica baseada em Barthélemy [2011] no tópico de modelos de geração de redes espaciais. O estudo de teoria de redes complexas é introduzido, assim como os principais tipos de modelos de geração de redes não-espaciais. Tais modelos são então adaptados para o caso espacial por diversos caminhos. Este estudo servirá como parte de um projeto maior de iniciação científica em redes.

Abstract

This text is a brief literature review based on Barthélemy [2011] on the topic of spatial network generating models. The field of network science is introduced, as well as the main types of models for non-spatial networks. These are then adapted to spatial networks through several paths. This study will serve as part of a larger undergraduate research project on networks.

Contents

1 Introduction

Network science has been a very active field in the last years. It lies in the intersection between several other fields, and has many real-world applications. It was born as a branch of graph theory (Bollobás [1998]) and it focuses specially on large graphs with special topological properties.

The field has become very popular for its wide-range applications, since basically any interacting system can be represented as a graph, and the mathematical toolkit can be applied to all such systems. Examples of systems studied in network science include biological systems, such as networks of cell metabolism (Holme [2011]), large-scale networks of internet connection (Yook et al. [2002]), functional neuronal networks (Zheng et al. [2020]) networks of language semantics (Steyvers and Tenenbaum [2005]), among many others.

A network is composed of two types of objects: nodes (or vertices) and links (or edges). They can represent all types of agents and interactions, not being restricted to physical connections. Examples include friendship networks (Marathe et al. [2013]) and airline route networks (Guo et al. [2019]).

This field gained momentum with the works of Erdős and Rényi (Erdős and Rényi [1959]), who pioneered on creating models that underlie graph formation. They focused on modeling how a certain type of random network can be formed, instead of simply studying *a priori* graphs. This type of modeling had another milestone in the works of Watts and Strogatz (Watts and Strogatz [1998]), who introduced the concept of "small-world" networks, and later with Albert and Barabási (Barabási and Albert [1999]), with "scale-free" networks. All these models will be discussed in the next sections.

Spatial networks, on the other hand, evolved as a branch of network science. Differently from its parent research field, in which the position of nodes are immaterial, so networks can be treated as topological objects that define isomorphism classes, in spatial networks, there are "wiring costs" associated with linking the nodes, such that the probability of linking nodes which are far away from each other is smaller than closer nodes, as we will see in the next sections.

These networks are usually better models for several of real-world networks,

so that friendship networks will be impacted by the geographic location of the nodes, transportation networks depend on distance, and communication networks have a short range, depending on the physical equipment used to communicate the signal. Apart from physical distances, spatial networks can represent any kind of distance, such as the social distance as measured by wage, social-professional categories, etc (Barthélemy [2011]). These observations motivate us to study spatial networks more thoroughly.

Similarly to the network formation models cited (Erdős and Rényi [1959], Watts and Strogatz [1998], Barabási and Albert [1999]), there are also spatial network generating models, some of which are adaptations of those three models. These models can be good approximations to spatial networks seen in the real world, so that, in order to start studying spatial networks, it is wise to begin with those models, and this is the goal of this text: to present the main network-generating models for non-spatial networks and then some examples of mappings to spatial networks, focusing on understanding what remains the same and what needs to be changed in this mapping.

This text was inspired by the review article (Barthélemy [2011]) and most of the models here presented are ones described in that work.

2 Some definitions on networks

2.1 Non-spatial networks

The definitions used in this section are based on Bollobás [1998].

A *graph* G is an ordered pair of disjoint sets (V, E) with E being a subset of the set of unordered pairs of V. E is called the *edges* set while V is the *vertices* set.

If two vertices x and y are connected by an edge of G, which means that $\{x, y\} \in E(G)$, then the vertices x and y are called *neighboring* or *adjacent*.

The *degree* of a node x is the number of all (different) nodes that connect to it, which is equivalent to the cardinality of $\{x : \{x, y\} \in E(G)\}$.

Two graphs G and G' are said to be *isomorphic* iff one can find a correspondence between their vertex sets preserving adjacency. This means that $G = (V, E)$ is isomorphic to $G' = (V', E')$ iff there is a bijection $\phi : V \to V'$ such that $\{x, y\} \in E$ iff $\{\phi(x), \phi(y)\} \in E'$. This definition is what allows one to, intuitively, "deform" a graph, since the graph after the transformation is an isomorphism of the initial one.

Planar networks (or planar graphs) are graphs that can be embedded in the plane, without having edges crossing themselves (Bollobás [1998]).

Spatial networks, on the other hand, are those graphs whose elements (edges and nodes) are geometrical elements embedded in a metric space (Barthélemy [2011]). These can be either planar or non-planar graphs.

Finally, some remarks on the expressions used in this text. The words "network" and "graph" are taken as synonyms, so ”vertex” and ”node” are equivalent, the same applying to ”edge” and ”link”.

Also, the word "distance" is used to describe physical distances in spatial networks, while "length" was taken as a measure of the number of hops needed to go from one node to another one in non-spatial networks. This follows the convention of Zitin et al. [2014].

3 Results and discussion: network generating models

3.1 Non-spatial networks

3.1.1 Random networks

This class of theoretical graphs was created by Paul Erdős and Alfréd Rényi (Erdős and Rényi [1959]), and represents the paradigm for many null models of "random graph". It is generated by running through all pairs of nodes and connecting them with a uniform probability p. This generates an average number of links on a network given by:

$$\langle E \rangle = p\frac{N(N-1)}{2} \tag{1}$$

And this equation makes up for an average (individual) degree of:

$$\langle k \rangle = 2\frac{\langle E \rangle}{N} = p(N-1) \tag{2}$$

Since the attribution of links for each pair of nodes is independent and identical for each of them, the degree distribution - probability that a randomly picked node has k network connections (Zitin et al. [2014]) - is binomial:

$$P(k) = \binom{N-1}{k} p^k (1-p)^{N-1-k} \tag{3}$$

Where N is the number of nodes in the network and k is the degree. This generates a network that, on the limit of a large number of nodes N, converges to a Gaussian degree distribution, by the Central Limit Theorem (Bárány and Vu [2007]).

This model was a major breakthrough in graph theory as it was one of the first to actually trying to understand how one generates graphs with specific properties (Barabási [2014]), not assuming the topology as given.

However, this model fails to predict many properties found in real-world networks, since it is usually not the case that nodes have uniform linkage probability, and some of the nodes in most networks have large degrees while others are more peripheral, something that would be almost forbidden by the Gaussian distribution.

3.1.2 Small-world networks

Small-world networks are those that have *short average path length* and *high clustering coefficient*.

The former is related to the number of hops it takes to go from one node to any other in the network, by taking the shortest possible path. This number is averaged out over all nodes, and if this value grows "slowly" (logarithmically) with the network size (number of nodes), the graph is regarded as having short average path length.

The latter, on the other hand, is defined as 3x the number of triangles in the network divided by the number of node pairs that link to a common node. This means that in networks with large clustering coefficients, if two nodes connect to a third one, they are likely to connect to each other as well, thus closing the triangle (Zitin et al. [2014]).

The property of being small-world is found in many real-world networks, and it differs from networks generated by the Erdős-Rényi process in the sense that, although the latter usually present small node-to-node distances, they also show small clustering coefficients.

Steven Strogatz and Duncan Watts worked out a theoretical model for describing those networks by using a network in which nodes were uniformly placed on a ring lattice and randomly attributed shortcuts between non-first neighbors, which allowed for hops between different sides of the circumference. This resulted on a network whose average path length was small and also shown a large clustering coefficient. This gave birth to the Watts-Strogatz network model (Watts and Strogatz [1998]).

The algorithm that generates this class of networks is the following (Watts and Strogatz [1998]):

Given N nodes, a desired mean degree K and a parameter β with $0 \leq \beta \leq 1$, in the limit of large networks, which, in this case, corresponds to: $N >> K >> \ln K >> 1$:

- Build a regular ring lattice with N nodes, each connected to K neighbors, equally split between each side
- For each node, take every edge connecting it to all its $K/2$ rightmost neighbors and rewire them with probability β, by disconnecting the node to its neighbor and

reconnecting it to any other node, without self-loops.

This generates a network that is small-world and which converges to a Erdős-Rényi one in the limit of large probability of shortcut: $\beta \to 1$.

Also note that, with a uniform distribution of shortcuts, there are no "wiring costs" associated with it, so large range connections are as likely as shorter ones. This is an important point, since it is modified in spatial adaptations, as we will see later on.

3.1.3 Scale-free networks

This is another property shared by many real-world networks and that also gave rise to a theoretical network model. It refers to networks showing degree distributions not fitting into the random network model created by Erdős and Rényi (Erdős and Rényi [1959]). In this model, the degree distribution is a Gaussian curve: most nodes' degrees have values around a mean and very few diverge too much from this mean. Scale-free networks, on the other hand, show power-law degree distributions of the form:

$$P(k) \sim k^{-\gamma}, \tag{4}$$

in which $2 < \gamma < 3$ gives rise to distributions that are well defined in their first moments but are ill-defined in their second-moments.

These differences are shown in the graphs below, from the website (Barabási et al. [2021]), comparing histograms of degree distributions of networks generated by the Erdős-Rényi (ER) model and scale-free networks, also called Barabási-Albert (BA):

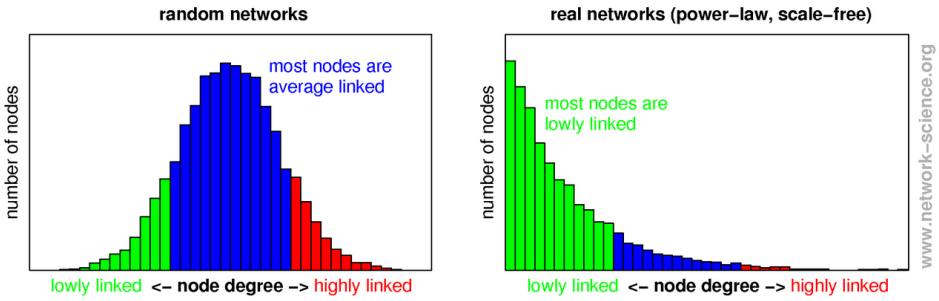

Figure 1: Comparison between ER and BA degree distributions.

These networks are formed by the so-called *preferential attachment model*, in which the probability of a new node n linking to an existing one i is proportional to the

degree of i. This is a basic rich-get-richer process, which describes graphs such as citation networks (Wang et al. [2010]), in which the most cited authors are most likely to be cited than others. In a plain BA model, since the network is generated by placing nodes one by one over time, older nodes tend to become hubs.

As a final remark, notice that, although this class of networks is a good approximation for many real-world networks, it is a simplified model that has generated many others depicting more complex phenomena. One that can be cited is the network fitness model, which considers that often what determines whether or not a node will become a hub is not its age, but a metric called *fitness*, which, for instance, accounts for the success of new players in tech industries, as explained in the book Barabási [2014].

3.2 Spatial networks

3.2.1 Planar random networks

To adapt the classical random network model for spatial networks, one possible idea is to place the nodes in fixed points in space and then connect them randomly. This generates arbitrary networks, most likely non-planar ones, from which one can only keep the links that preserve planarity.

By doing this on a network of $N = 1000$ nodes, the review articles provides the network depicted in Figure 2:

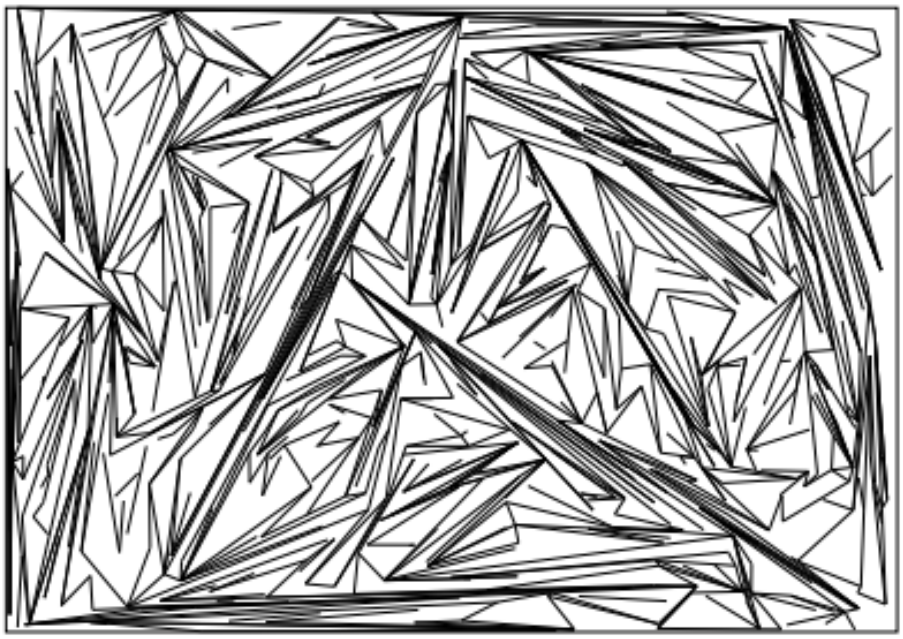

Figure 2: Spatial Erdős-Rényi network with $N = 1000$.

Apart from this straightforward generalization, we can have other models that adapt ER networks. One example is the Waxman model (Waxman [1988]). In this model, the nodes are uniformly distributed in the plane just like the previous model, but instead of using a uniform distribution to connect the edges, it connects nodes i and j with probability given by:

$$P(i,j) = \beta\, e^{-d_E(i,j)/d_0} \tag{5}$$

Where d_0 represents the typical length of the links in this network, while β is a parameter responsible for controlling the density of nodes in the network. This model

thus represents the exponential difficulty of connecting nodes which are far-apart in the plane. In the limit of large d_0 the probability converges to the constant β, so one recovers the previously cited ER spatial model.

This model, although very simple, served as basis for more sophisticated ones used to describe the topology of the Internet (Zegura et al. [1996], Barthélemy [2011]).

Note that both of the models shown in this section are planar.

3.2.2 Small-world networks

Several real-world networks that show this property, described in previous sections, are actually embedded in space, so it is important to have a spatial analogue of this type of networks.

One spatial variant of the Watts-Strogatz model places the nodes on a regular d-dimensional lattice, using periodic boundary conditions (Barthélemy [2011]). The reason for replacing the ring lattice is to represent the wiring costs of the random shortcuts. After adding the nodes, links are added with a probability that depends on the link's length (that is, the physical distance d between the end nodes):

$$q(d) = d^{-\alpha} \tag{6}$$

With this definition, we can see that, for large values of α, the probability of linking nodes which are far-apart is small, so the network's average shortest path $\langle l \rangle$ scales as:

$$\langle l \rangle \sim N^{1/d} \tag{7}$$

Since this is a super-logarithmic behavior, large values of α tune the network to converge to a "large-world" one.

On the other hand, if α is small, the average shortest path scales as:

$$\langle l \rangle \sim \log N \tag{8}$$

,

and we recover the small-world behavior.

There is, therefore, a critical value α_c which separates the two types of behavior. This threshold is shown to be $\alpha = d + 1$ (Petermann and Rios [2005]).

Other interesting models that adapt the small-world one to spatial networks are models that grow the network from zero, placing the nodes one by one. The first of them is called *sphere network model* and works as follows (Zitin et al. [2014]):

Place the nodes in the surface of a d-dimensional hypersphere. All nodes act as identical point charges, and new nodes being added to the sphere displace the already existing ones so as to obtain "electrostatic energy equilibrium" in the whole surface. The force used to mutually repel the charges is a generalized electrostatic force proportional to $r^{-\alpha}$, r being the distance between the charges, and $\alpha = d - 1$, not necessarily equating the value of 2 in the Coulomb exponent.

The model then works by adding the nodes one by one at random with uniform probability density per unit area and then adding links that connect each node to their m closest neighbors. After these connections, we relax the position of all nodes so as to minimize the potential energy, by means of a gradient descent method (Zitin et al. [2014]):

$$\frac{d\vec{x}_i}{dt} = P\left[\vec{F}_i\right] \tag{9}$$

Here, $\vec{x}_i$ is the position vector of each node in the unitary hypersphere, while $P\left[\vec{F}_i\right]$ is the projection of the force $\vec{F}_i$ to the sphere's surface. This force is given by (generalized) Coulomb's law:

$$\vec{F}_i = \sum_{j \neq i} \frac{\vec{r}_{ij}}{r_{ij}^d} \tag{10}$$

Where $\vec{r}$ is the displacement vector between nodes i and j: $\vec{r}_{ij} = \vec{x}_i - \vec{x}_j$.

The authors add that this model works best for large numbers of nodes N, for which the node distribution can get approximately uniform.

There are other adaptations of this model, for instance, the so-called *Plum Pudding Network Model*, in which the relaxation step allows the nodes to move to the interior of the sphere, inside a ball of unitary radius. This model lends its name from the atomic model of J. J. Thomson, from 1904 (Hentschel [2009]).

Note that the sphere network model generates networks whose average shortest

path length metric scales as $\langle l \rangle \sim \log N$. This happens because the older nodes are repelled as the network grows, thus creating links between far away nodes in the final network. The clustering coefficient is also nonzero for large N, a feat that would not be present in ER networks, thus indicating a small-world behavior, and showing that this type of spatial network growth can generate another class of networks similar to WS in its topological properties.

Note finally that the models presented in this section are not planar, like the ones discussed in the spatial ER subsection.

3.2.3 Scale-free networks

To have scale-free networks, as discussed in the section of non-spatial networks, the process to be considered - preferential attachment - is based on network growth, like the spherical network model, discussed in the previous subsection. In scale-free, or Barabási-Albert (BA) networks, new nodes are more likely to attach to well-connected ones, called hubs, which are the oldest ones in the original BA model (Barabási and Albert [1999]) and the fittest ones in more recent adaptations (Bianconi and Barabási [2001]).

In this context, the generation of scale-free spatial networks is also done using generating processes, but this time the wiring costs are taken into consideration, and they restrict long-range connections to appear only if they are sufficiently relevant to the overall connectivity of the network (that is, hubs are being connected).

In the original BA network, a new node n connects to each existing node i with probability proportional to the degree of i, that is, the probability of link between these nodes is:

$$P(n, i) \sim k_i \tag{11}$$

In most adapted spatial versions, the new node location is drawn from a uniform distribution in space. The attachment probability is then modified as:

$$P(n, i) \sim k_i \, F(d_E(n, i)) \tag{12}$$

Where $F(d_E(n, i))$ is a function of the Euclidean distance between nodes n

and i. Usually, this is a decreasing function of distance, so new nodes tend to connect to hubs, unless these are too far away, in which case they link to less connected nearby nodes (Barthélemy [2011]). This can be seen in airplane networks, in which small local airports will connect to larger ones up until a certain distance, above which it is better to connect to other local nearby airports.

Next, we briefly comment on two possible forms for the function F.

Finite range

This choice is when F is a decreasing exponential with distance, given by:

$$F(d) = e^{-d/r_c} \tag{13}$$

What happens in this case is the introduction of a new scale in the system, called the *interaction range*, r_c. This controls how much the distance influences the system. If r_c is large (or at least, of the order of the system size) the exponential term gets very small, and F approaches unity, and the network returns to its scale-free behavior, independent of distance.

Note that, this case can be considered a generalization of the Waxman model for non-planar spatial networks, in an adaptation that takes into account the degree of each node in the probability of linkage to other nodes.

Power law

This other option is when F is a decaying power law of the distance:

$$F(d) = d^{\alpha} \tag{14}$$

In this case, the degree distribution has two regimes: $\alpha > -1$, for which it has a broad shape and $\alpha < -1$ for which it quickly decays. For $\alpha = 0$ one gets back the BA model.

4 Conclusion

In this text we introduced the field of network science with a particular focus on spatial networks. Three milestone methods of network generation were presented, and these were adapted to include wiring costs in their spatial versions.

In this adaptation, we saw planar models generating ER networks by randomly assigning nodes to a plane and connecting them, keeping only the links preserving planarity. The other planar ER model presented is the Waxman model, that assigns a decaying exponential to the probability of linkage between nodes.

Going forward as far as network-generating models are concerned, we have presented models that generate networks with small-world behavior. One of them works by placing nodes on a regular lattice and assigning links whose probabilities decay with the distance. Another model shown in this section was the sphere network model, which is based on growing a network from zero, placing nodes one by one on the surface of a sphere, and rearranging the other nodes as to minimize a potential energy based on a generalized-Coulomb force.

Lastly, models that mimic the scale-free behavior were presented. These are based on the modification of the step in the BA network formation that attributes the probability of linking new nodes to previous existing ones, also by using growth models. The probability of linkage is changed to accommodate not only the proportionality to the node's degree but also some kind of decaying function of spatial distance between the nodes, so that the new node will only link to far away nodes if these are hubs, otherwise it will prefer to be linked to local neighbors.

The approach followed by this text is almost entirely based on the review paper (Barthélemy [2011]), and served as an introduction to the topic of spatial networks, a field which we will be exploring in more depth in the next months in order to apply it to the scope of our other studies in networks.

References

Albert-László Barabási and Réka Albert. Emergence of scaling in random networks. *Science*, 286(5439):509–512, 1999.

Albert-László Barabási. *Linked: How Everything Is Connected to Everything Else and What It Means for Business, Science, and Everyday Life*. Basic Books, 2014.

Albert-László Barabási et al. Comparison between degree distributions, 2021. URL http://www.network-science.org/powerlaw_scalefree_node_degree_distribution.html.

Marc Barthélemy. Spatial networks. *Physics Reports*, 499(1-3):1–101, 2011.

G Bianconi and A.-L Barabási. Competition and multiscaling in evolving networks. *Europhysics Letters (EPL)*, 54(4):436–442, 2001.

Béla Bollobás. *Modern Graph Theory*. Springer Science & Business Media, 1998.

Imre Bárány and Van Vu. Central limit theorems for Gaussian polytopes. *The Annals of Probability*, 35(4):1593 – 1621, 2007.

P. Erdős and A. Rényi. On random graphs. *Publicationes Mathematicae*, pages 290–297, 1959.

Weisi Guo et al. Global air transport complex network: multi-scale analysis. *SN Applied Sciences*, 1:680, 06 2019.

Klaus Hentschel. Atomic models, j.j. thomson's "plum pudding" model. 01 2009.

Petter Holme. Metabolic robustness and network modularity: A model study. *PLOS ONE*, 6(2):1–7, 02 2011.

Achla Marathe et al. Analysis of friendship network and its role in explaining obesity. *ACM transactions on intelligent systems and technology*, 4(3):1–21, 06 2013.

T. Petermann and P. Rios. Spatial small-world networks: A wiring-cost perspective. *arXiv: Disordered Systems and Neural Networks*, 2005.

Mark Steyvers and Joshua B. Tenenbaum. The large-scale structure of semantic networks: Statistical analyses and a model of semantic growth. *Cognitive Science*, 29(1):41–78, 2005.

Ming-Yang Wang, Guang Yu, and Da-Ren Yu. The scale-free model for citation network. In *2010 IEEE International Conference on Intelligent Computing and Intelligent Systems*, volume 1, pages 773–776, 2010.

Duncan J. Watts and Steven H. Strogatz. Collective dynamics of 'small-world' networks. *Nature*, 393:440–442, 1998.

B. M. Waxman. Routing of multipoint connections. *IEEE J. Sel. Areas Commun.*, 6: 1617–1622, 1988.

Soon-Hyung Yook et al. Modeling the internet's large-scale topology. *Proceedings of the National Academy of Sciences of the United States of America (PNAS)*, 99(21): 13382–13386, 2002.

E. Zegura et al. How to model an internetwork. *Proceedings of IEEE INFOCOM '96. Conference on Computer Communications*, 2:594–602, 1996.

Muhua Zheng et al. Geometric renormalization unravels self-similarity of the multiscale human connectome. *Proceedings of the National Academy of Sciences*, 117(33): 20244–20253, 2020.

Ari Zitin et al. Spatially embedded growing small-world networks. *Scientific Reports*, 4: 7047, 2014.