Barthelemy, Marc
Spatial networks. A complete introduction: from graph theory and statistical physics to real-world applications.
Springer [2022] 437 pp.

Draft review by David Aldous

The field of spatial networks is fascinating to me, partly because such networks are easy to visualize and indeed are familiar from maps, and partly because problems and models have arisen independently from many diverse sources, and methodologies have ranged across the spectrum from theorem-proof mathematics to heuristics to data analysis. Seeking to cover such a spectrum within any particular field, at an introductory level, is very challenging, and so such books are quite rare. The first half of the book (Chapters 1-10) focusses on descriptive statistics relevant to the type of questions one can ask about network structure, and the second half (Chapters 11-20) describes and studies a range of network models. Being written by a statistical physicist, the book has a particular style that may be unfamiliar to mathematicians. The style is that of a survey – outlining numerous results (around 500 papers are cited) via descriptions of statistics and models and using what I would call back-of-envelope calculations and simulation results together with real world data examples. Very different from a typical mathematical textbook or monograph which goes deeply into a few topics via a systematic development of theory. Fortunately the book is mostly technically undemanding, meaning it should mostly be accessible to an undergraduate mathematics student.

A *graph*, in discrete mathematics, is an abstract set of vertices and edges, and the word network is best considered to mean a graph with some context-dependent extra structure. In contrast, a spatial network is embedded in the plane. That is, the vertices have specific positions in the plane. Familiar real-world examples are *communication networks* such as road, railway and fiber-optic networks, and *distribution networks* such as electricity, gas and water supply. In those cases the edges are also explicitly situated in the plane; in other examples such as airline routes or wireless communication the edges are implicit. Realistic modeling of any one real-world network would involve many intricate details of population and geographical features. The field under consideration tends to stand back from such details and instead studies relations between statistical features of networks, and how these features affect economic or other activities involving the network. Talking about *statistical features* in the first half of the book naturally suggests the consideration of probability models in the second half.

The field of spatial networks touches upon, though is conceptually different from, surprisingly many well-developed areas of theorem-proof mathematics, noted in this review as (**a, b, c, . . .**). There is (**a**) the classical theory of random graphs [5, 9, 10, 16] centered upon the simplest Erdös-Rényi model, and since 2000 a vast literature (**b**) on preferential attachment, "small worlds" and related models has appeared. But this is mostly not spatial. Non-random (**c**) planar graphs (Chapters 2-3) are classical [14], and also there exists a technically sophisticated recent literature studying the random case, under the phrase (**d**) *random planar maps* [4, 12]. However this is not truly *spatial*, in that it identifies isomorphic graphs, so there are only a finite number of n-vertex planar graphs. The resulting model of uniform random maps is very interesting as mathematics, and is related to the (rather speculative) physics notion of *quantum gravity*, but has no apparent relevance to the everyday networks within direct human experience.

Chapter 4 starts by observing that several of the standard descriptive statistics for general networks are (in practice) irrelevant for spatial networks. In particular, the much-studied context of power-law tails for distribution of vertex degrees (number of edges at a vertex) is implausible – the spatial setting *imposes severe constraints on the degree of a node and its distribution which is generally peaked around its average*" (section 4.1.1). As one would guess, for urban road networks the average is around 4. Other uninformative (in the spatial setting) statistics include clustering coefficients and assortativity measures. A more surprising empirical observation is that edge lengths (relative to their average) in areas of different population density (urban or rural) have roughly the same distribution (section 4.1.2), an observation lacking any plausible theoretical explanation. Real world spatial networks invariably have some notion of traffic flow and some notion of routes between vertices, and the study of routes is perhaps the most fundamental issue for spatial networks. In simple models, there are two notions of the *length* of a route. *Hop-length* counts the number of edges in the route, in other words imagines each edge having length 1. Or the model may include real-valued edge-lengths (interpretable as transit time) and the sum of such edge-lengths is what I will call the *geometric-length* of the route. The latter seems more realistic and also more natural mathematically, because (for generic edge-lengths) the shortest route will be unique for geometric-length, whereas hop-lengths are small integers and so there will typically be multiple shortest hop-length routes.

One vague general and much-studied question: which vertices or edges are *important* in the sense of having heavy traffic, that is being in many routes? This in turn should relate to which edges are *critical* in the sense that their failure would have substantial effect on the functioning of the network. This general issue, called *betweenness centrality*, is the topic of Chapter 5. The simplest formulation is to consider all start-end pairs (v_1, v_2) and their shortest routes, and define (in the geometric-length case) BC(v) to be the proportion of such routes that pass through v. In the hop-length case one needs the more awkward definition of BC(v) as the average, over pairs (v_1, v_2), of the proportion of shortest-hop-length routes that pass through v. Chapter 5 outlines calculations of BC for simple examples of networks (grids and trees) and also treats two more elaborate cases: first a "grid-tree" model consisting of a square grid with a tree attached to each side, and second some heuristics for an "increasing density of vertices" type of continuum limit. There is also discussion of the real world distribution over v of BC(v) for urban street networks.

Chapter 6 talks briefly about (**e**) first-passage percolation [3], that is study of the shortest geometric-length-route between distance vertices. This has been extensively studied as theorem-proof mathematics in the model of random edge-lengths on the square (and other) lattices. For more general models in the continuous plane, such as the proximity networks described later, there is the fascinating longstanding KPZ conjecture [7] that $\chi = 2\xi - 1$ for the scaling exponents χ and ξ defined by: for shortest routes between vertices at Euclidean distance d
variance of route length $\asymp d^{2\chi}$
maximum deviation of route from linear $\asymp d^{\xi}$.
Such universality results are a central goal of theoretical statistical physics, and represent extremely challenging open problems for rigorous proofs. On the other hand, for real-world networks a much more prominent issue is the actual route length – how does this compare to Euclidean distance?

Chapter 7 considers the notion of the *simplest* route, defined here as minimizing the number of turns required, where *turn* is rather arbitrarily defined as a turn of more than 30 degrees at an intersection. (Some automobile navigation systems offer this as an alternative to the fastest route). One can then define the "simplicity profile" $S(d) \geq 1$ of a road network to be the average ratio simplest/shortest lengths as a function of Euclidean distance d. This profile and its maximum $S^* = \max_d S(d)$ is intended as a rough indication of some form of "complexity" of the network. One can compare the profiles for real networks (roads or railways) and simple models of those and biological vein networks. A tentative conclusion is that there is a systematic difference between communication and distribution networks. Section 7.2 describes another intriguing (though not so convincing to me) topic, complexity in human perception. This is illustrated by considering a route through a subway network that needs to be memorized, thereby bringing "entropy of information" into play in the route description.

For general networks, a major subject in both theory and applications has been *community detection*, that is finding regions where most vertices and most of their neighbors have a given attribute. In a spatial network, possession of an attribute will often be somewhat correlated with spatial position. So one might modify general algorithms to incorporate spatial knowledge, or ask whether ignoring spatial knowledge may lead to erroneous results. This is discussed in Chapter 8 via examples and a simple spatial analog of the widely-used "planted bisection" benchmark model for comparing algorithms on general networks.

Chapter 9 considers urban street networks, which partition the city into blocks (*cells*, in geometry). Each block has some area and can be assigned some measure of shape (e.g. area/diameter^2) and so a scatter diagram of shape versus area provides information about the local structure of the network, enabling comparison of different cities. Also, an exact bijection between planar graphs and trees is described: this is one point of contact with the “non-spatial” theory of random mappings mentioned earlier, the development of that theory being heavily based on such bijections.

Chapter 10 describes data on the growth of real road networks in two locations (Paris; and a 125 km^2 region north of Milan) over the last several centuries. As one would expect from elementary scaling arguments, the number of intersections (vertices) increased linearly with population increase, and the total length of the network increased as population$^{1/2}$. Studying the network statistics introduced in previous chapters, one can see differences between the “organic” growth in the Milan region and the extensive planned renovation of Paris by Haussmann in the mid-1800s.

Part 2 of the book focusses on probability models of networks and their analysis. The *statistical physics* phrase in the title might be discouraging for readers who have never taken such a course, but here it is more a viewpoint, emphasizing scaling arguments, power law behavior and “universality”, meaning behavior that is similar for a range of models rather than dependent on the details of a particular model.

Chapters 11 and 12 give brief discussions of spatial variants of the much-studied Erdös-Rényi and “small worlds” networks. In particular there is an outline of the following celebrated result of Kleinberg. Start with the 2-dimensional grid. Add extra edges at random, with the probability of an edge of length d scaling as $d^{-\alpha}$ for real $\alpha > 0$. Then for the “critical value” $\alpha = 2$, but no other value, one can navigate through the network to a given destination at Euclidean distance D using a small number of steps, in fact $O(\log^2 D)$ steps rather than D^β for $\beta(\alpha) > 0$.

A large literature on non-spatial networks studies variants of a basic *preferential attachment* model for growing random networks one vertex at a time. In the basic model, a new vertex v^* is linked to a fixed number of existing vertices v with relative probabilities depending on the degree of v (increasing linearly, in the simplest case). Chapter 13 describes natural spatial analogs, in which attachment probabilities depend on both the degree of v and the distance of v from the new vertex v^*. Properties of statistics for route lengths and degree distributions are derived heuristically. In a more elaborate model, envisaging roads in a city with non-uniform population density, the attachment probabilities can depend also on the local density and the level of traffic through the existing vertex.

Chapters 14–16 give brief accounts of some models that have been well-studied in the theorem-proof literature. Chapter 14 considers the random Voronoi tessellation, constructed over the Poisson point process in the plane. This was studied in (**f**) classical *stochastic geometry* [8] as a tractable way of randomly partitioning the plane into cells. Also mentioned are dynamical processes by which cells are split by random lines. Chapter 15 continues with another model which has been well-studied more recently, the (**g**) *random ge-*

ometric graph [15] in which edges are placed between each pair of vertices at distance less than some assigned threshold value. Chapter 16 describes what are usually called (**h**) *proximity graphs* [6], defined by putting an edge between two vertices if some specified region based on those points is empty of vertices. This slightly counter-intuitive definition implies that a vertex gets linked to a few nearby vertices in such a way that the network is always connected.

Many of the world's largest cities have extensive subway networks, which (for intuitively rather obvious reasons) typically have a characteristic shape, consisting of a well-connected central core with branches radiating outwards. Chapter 17 considers a very simple model consisting of a "star" of lines radiating out from a center, intersecting with one circular "ring" line. Adding extra structure to this model – a non-uniform density of users, and costs from congestion effects one can do calculations of betweenness centrality and traffic flows. For instance one sees a trade-off (as parameters vary) regarding whether routes go via the center or via the ring.

Chapters 18 and 19 deal with optimal networks. Chapter 18 briefly reviews two areas of well-studied theory. The *minimum spanning tree* is the shortest network connecting given points. Some of its quantitative behavior is described for two probability models: **(i)** the randomly edge-weighted complete graph [2], and the Poisson point process on the plane. More relevant to spatial networks is the concept of a **(j)** t-spanner [13], a network in which the route length between any two vertices is at most t times the Euclidean distance. Chapter 19 takes a more detailed look at the case of subway networks. One general problem formulation is to assume a Gaussian population density and seek the network of given total length that is "optimal" according to some specific criterion. Of course one hopes that the optimal network in such a model will show the "core and branches" structure observed in real-world subway networks. This turns out to be a surprisingly challenging problem that has not been successfully answered, because it would require consideration of all possible topologies. Chapter 19 gives extensive calculations for specific topologies.

Chapter 20 returns to models for growing random networks one vertex at a time. Here one considers a local optimization (benefit minus cost) rule to decide how to link a new vertex to the existing network. The "cost" is proportional to the length added. The "benefit" is assessed using a conventional "gravity model" of traffic, that the level of traffic between vertices at distance d scales as $d^{-\alpha}$. Analysis of such models is extremely difficult, and here is restricted to tree-like networks.

Commentary: This book succeeds admirably in its stated "complete introduction" goal, covering briefly a huge range of questions and models that have been studied, with numerous references to the original papers which may be consulted for more extensive analysis. Of course I have been unable to mention all the topics in this review. The focus on road networks as examples reflects the author's own research, reasonably enough, though more data from other examples might have helped readers appreciate the scope of the field.

To readers from theorem-proof mathematics, the parts not related to existing theory may be most intriguing, to see the different approach. One often starts with some elementary scaling argument, illustrated by the following well-known example (section 14.2). You want to decide where to site a chain of retail stores, keeping in mind that population density $\lambda(z)$ varies greatly across positions z, and seeking to minimize the mean distance from a customer to the closest store. The answer is to make the store density be proportional to $\lambda^{2/3}(z)$. As another example, to design a new high speed railway network to link the N largest cities in a country of area A, the total length must generically scale as $\sqrt{NA}$. A longer network (larger c in the length $c\sqrt{NA}$) will be more efficient in providing shorter routes, but there is an obvious "law of diminishing returns" – how much extra are we willing to spend in order to reduce average route length from 1.3 to 1.2 times Euclidean distance? From a Poisson model and the type of networks described in section 4.2.5, one can numerically obtain $c \approx 2$ as a good trade-off.

To my own taste, it is the study of optimal networks that offers the greatest challenge for rigorous theory. The example above, and the "subway"example from Chapter 19, illustrate a general difficulty faced in rigorous theory. Orders of magnitude are usually rather obvious, while actual numbers – the constants c in first order asymptotics – seem beyond the reach of theory and rely on numerical estimates from simulations. Indeed a key constant in the entire field of probability on the plane – the critical parameter in continuum percolation – seems beyond reach. Sophisticated statistical physics has studied universality of scaling exponents at critical values of systems, but (notwithstanding the KPZ conjecture from Chapter 6) it is not clear if there are any analogs for spatial networks that relate to natural real-world questions.

On the applied side, it seems to me that the main conceptual issue in devising models that might illuminate the structure of real-world spatial networks (which reflect human activity) is the extreme variation of population density; how does one choose a function $\rho(z)$ to model population density throughout a generic country?

The models considered in the book have simple descriptions, and one can readily think of variant models in an attempt to be more realistic. Simulation studies of such models offer opportunities for undergraduate research, and I have done a number of such projects with undergraduates.

From the viewpoint of technically sophisticated theory, one challenge already mentioned is to make rigorous the KPZ conjecture. Here is a different idea, not mentioned in this book. Random spanning trees on the 2-dimensional lattice have been studied, with the goal of establishing a scaling limit as a random spanning tree linking almost all points in the continuum plane [11]. This approach to defining continuum networks has apparently not been studied, but one can try an alternate approach, as follows. Instead of edges, take routes as primitives. So a realization of the random network is a collection of routes, with a route specified between each pair of points, satisfying natural consistency conditions. In particular, this approach allows one to construct networks which are exactly (Euclidean) scale-invariant, a property that is only possible in the continuum. Constructions and analysis of this class of models remains an open research topic [1]. It is intriguing that scale-invariance implies there are should be straight edges of all length scales, echoing a comment from this book's section on simple routes: *almost always beneficial to take long straight lines when they exist* (section 7.1.2).

References

[1] David Aldous and Karthik Ganesan. True scale-invariant random spatial networks. *Proc. Natl. Acad. Sci. USA*, 110(22):8782–8785, 2013.

[2] David J. Aldous and Shankar Bhamidi. Edge flows in the complete random-lengths network. *Random Structures Algorithms*, 37(3):271–311, 2010.

[3] Antonio Auffinger, Michael Damron, and Jack Hanson. *50 years of first-passage percolation*, volume 68 of *University Lecture Series*. American Mathematical Society, Providence, RI, 2017.

[4] Jean Bertoin, Nicolas Curien, and Igor Kortchemski. Random planar maps and growth-fragmentations. *Ann. Probab.*, 46(1):207–260, 2018.

[5] Béla Bollobás. *Random graphs*, volume 73 of *Cambridge Studies in Advanced Mathematics*. Cambridge University Press, Cambridge, second edition, 2001.

[6] Prosenjit Bose, Vida Dujmović, Ferran Hurtado, John Iacono, Stefan Langerman, Henk Meijer, Vera Sacristán, Maria Saumell, and David R. Wood. Proximity graphs: E, δ, Δ, χ and ω. *Internat. J. Comput. Geom. Appl.*, 22(5):439–469, 2012.

[7] Sourav Chatterjee. The universal relation between scaling exponents in first-passage percolation. *Ann. of Math. (2)*, 177(2):663–697, 2013.

[8] Sung Nok Chiu, Dietrich Stoyan, Wilfrid S. Kendall, and Joseph Mecke. *Stochastic geometry and its applications*. Wiley Series in Probability and Statistics. John Wiley & Sons, Ltd., Chichester, third edition, 2013.

[9] Rick Durrett. *Random graph dynamics*, volume 20 of *Cambridge Series in Statistical and Probabilistic Mathematics*. Cambridge University Press, Cambridge, 2007.

[10] Alan Frieze and MichałKaroński. *Introduction to random graphs*. Cambridge University Press, Cambridge, 2016.

[11] Christophe Garban, Gábor Pete, and Oded Schramm. The scaling limits of the minimal spanning tree and invasion percolation in the plane. *Ann. Probab.*, 46(6):3501–3557, 2018.

[12] J.-F. Le Gall. The Brownian map: a universal limit for random planar maps. In *XVIIth International Congress on Mathematical Physics*, pages 420–428. World Sci. Publ., Hackensack, NJ, 2014.

[13] Giri Narasimhan and Michiel Smid. *Geometric spanner networks*. Cambridge University Press, Cambridge, 2007.

[14] T. Nishizeki and N. Chiba. *Planar graphs: theory and algorithms*, volume 140 of *North-Holland Mathematics Studies*. North-Holland Publishing Co., Amsterdam, 1988. Annals of Discrete Mathematics, 32.

[15] Mathew Penrose. *Random geometric graphs*, volume 5 of *Oxford Studies in Probability*. Oxford University Press, Oxford, 2003.

[16] Remco van der Hofstad. *Random graphs and complex networks. Vol. 1*. Cambridge Series in Statistical and Probabilistic Mathematics, [43]. Cambridge University Press, Cambridge, 2017.