Applied Mathematics and Network Science

Daryl DeFord

1 Networks and Applied Mathematics

At the beginning of this note I would like to express a little bit of my philosophy of applied mathematics, noting clearly that this is simply a reflection of my preferences and thoughts and not intended as a statement of how anyone else should "do" or think about mathematics. My main philosophy of applied problems grows out of the modeling process, in that I think about applied mathematics as the natural result of embedding real–world data and problems in rigid mathematical objects where we can apply the theorems and techniques of "pure" mathematics to great effect.

There are (at least) two distinct types of problems that arise under this type of modeling. The first are the purely mathematical problems of determining the properties of the portion of the mathematical space that the embedding reaches. Frequently these are different than those that are usually considered as the result of the "regular" progress of mathematics. The entire field of mathematical approaches for complex networks is a perfect example of this type, where questions such as the power law degree distribution of generative models or the importance of the clustering coefficient to network structure were unlikely to have developed in the combinatorial study of graphs.

Secondly, there are questions that arise in the application of known results to the embedded data. A relevant example to my research is graph Laplacian, most of whose properties have been developed in analogy with the heat diffusion operator on smooth manifolds. This is a little closer to the setting of what most people think of as traditional applied mathematics, as the mathematical techniques as the goal tends to be developing a better understanding of the structure of the real world objects themselves rather than the properties of the mathematical space they inhabit. I tend to view these problem types as mostly orthogonal in the sense that they are independently interesting and fully justifiable as legitimate avenues of mathematical inquiry (YMMV).

Finally, notice that neither of these problem types necessarily evaluates its application against a notion of correctness with respect to real world validation. While this is a process that can lead to interesting iterations, going back and forth tweaking modeling hypotheses and techniques, and is of course of prime importance when considering the real–world usefulness of these methods, from my perspective this has little bearing on the mathematical interestingness of the problems.

In addition to the question of problem selection, one of the problems facing the field of complex networks right now is the level of rigor that is requires to establish a result. In my view this is related to the concerns addressed above in the following way: mathematical abstractions are never completely accurate representations of the real–world objects that are the actual interest of study. Thus, the best we can hope for (note that this is not the best that happens – the "unreasonable effectiveness of mathematics in the sciences" [7] is a real phenomena) is that the mathematical results we can prove provide intuition and guidelines for how we should evaluate the results of our real–world experiments and data gathering.

This plays out time and again in the field, and indeed one of the most important skills that a researcher can possess is the ability to construct simple mathematical models of real world data about which we can prove results to help evaluate our intuition from the real data we encounter. In this sense, my philosophy is a Bayesian one – mathematical examples can offer signposts to update our intuition about the objects we encounter in data analysis. For many networks problems it is often difficult to even state a general result that would describe a useful way to transfer facts learned about a particular example to a "generic" model but the intuition can still be valuable for practitioners.

In this regard, my applied work is representative of the standard types of argument in the field of complex networks, with computational results buttressed by theoretical proofs wherever possible but unafraid to attack problems for which there do not yet exist the frameworks to describe them theoretically.

2 Why are networks not graphs?

Network science is a highly interdisciplinary and rapidly advancing field. Frequently, results and intuition have developed around ideas from a computational or experimental perspective that took many years to fully justify (e.g. the power law degree distribution of the preferential attachment model [1, 2]) or have even been overturned by more careful theoretical examinations (such as the rush to show that various families of empirical networks obey power law degree distributions, even though this is actually quite rare [4, 3]). Partially, this tendency is due to the fact that many of the problems do not have unique optimal solutions or are mathematically intractable and hence approximate or heuristic solutions are necessary. This is particularly true for problems that are motivated by solving real world problems or are attached to actual data, where there is usually a pressing concern that can be answered for a particular case, even if there does not appear to be a natural generalization to an entire class of objects.

Many of the graphs encountered in mathematical (combinatorial) settings are quite distinct from the ones that we use to represent real world data and problems. In addition to the structural differences, explored in the remainder of this section, the questions that we are interested in are usually quite different. Queries such as: “What is the chromatic number of the internet?”, “How many perfect matchings are possible on the FaceBook graph?”, or even something as simple as “How many nodes or edges are in the world wide web?” are at best ill–formed questions, as these networks are constantly changing and evolving. Additionally, it is important to note that the same is true in the other direction: “Which is the most important node in the Peterson graph?” or “Which node controls the information flow through the Cayley graph of $SL_2(\mathbb{Z}/p\mathbb{Z})$?” or “What is the community structure in the knight’s tour graph?” are similarly ill formed questions.

A reasonable counterargument might be that although the relevant questions are different the fundamental objects of study are still graphs, in the sense that they are formed of vertices and edges, and hence networks should be considered a subfield of graph theory. In fact, this was exactly the argument that I made as a graph theorist, with interests in enumerative problems, encountering networks for the first time. However, discussions with knowledgeable experts and increasing familiarity with the networks literature have added a great deal to my perspective and demonstrated why my initial argument misses the point.

Most importantly, over–identifying a real world system with a particular instance of a model and ignoring the context in which the data arises can lead to misleading conclusions about real world processes. This is always an issue with mathematical models but tends to be exacerbated in the setting of networks, where the abstract objects are so simplified. Additionally, many of the systems under consideration are dynamic, with properties that change with respect to time and measurement process. Even though generalizations such as weighted, directed, or multiplex networks, among others, have been introduced to capture a higher degree of resolution of the data, these are still abstractions and can even introduce their own biases and misconceptions. With this understanding, asking and answering specific questions about fixed graphs is not necessarily the most appropriate approach for understanding real world systems, as overgeneralizing from a specific realization of the data as a network can lead to misleading conclusions.

The uses of randomized models between the fields also helps to highlight this difference. From a combinatorial perspective, the Erdös–Renyí model is a beautiful construction, where the independence of edges vastly simplifies the computation of expected properties, and whose tractability has led to many exciting results in Ramsey theory. From a networks perspective however, this model does both a poor job of explaining the way that networks form[1] as well as capturing the topological features that are most relevant to networks questions. Instead, networks research tends to prefer more complex models that represent a contextually appropriate generative story as well as more practical topological features such as transitivity or community structure. The tradeoff is again tractability and elegance with applicability and interpretability.

How then does one tell a complex network from a graph? While the author is fond of the daddy long–legs test[2] or the trivial automorphism group test[3], in practice, context determines the relevant questions. Thus, this question is similarly ill–posed. It is not the topological properties that determine which questions are relevant but rather the modeling approach and application that determine the usefulness of a particular nomenclature.

[1] A new individual in a social network is unlikely to choose their friends by a sequence of independent, weighted coin flips.

[2] If it looks like a spider that has been whacked with a newspaper it is probably a network...

[3] except for trivial leaf permutations there are rarely interesting automorphisms of networks, whereas combinatorial graphs frequently have large automorphism groups. See [5] and [6] for further discussions of this statement.

References

[1] Albert-László Barabási and Réka Albert. Emergence of scaling in random networks. *Science*, 286:509–512, October 1999.

[2] B. Bollobás. Mathematical results on scale-free random graphs. In *Handbook of Graphs and Networks*, pages 1–37. Wiley, 2003.

[3] Anna D. Broido and Aaron Clauset. Scale-free networks are rare. *arXiv:1801.03400 [physics.soc-ph]*, January 2018.

[4] A. Clauset, C. Shalizi, and M. Newman. Power-law distributions in empirical data. *SIAM Review*, 51(4):661–703, November 2009.

[5] Diego Garlaschelli, Franco Ruzzenenti, and Riccardo Basosi. Complex networks and symmetry I: A review. *Symmetry*, 2(3):1683–1709, September 2010.

[6] Ben D. MacArthur, Rubén J. Sánchez-García, and James W. Anderson. Symmetry in complex networks. *Discrete Applied Mathematics*, 156(18):3525–3531, November 2008.

[7] Eugene P. Wigner. The Unreasonable Effectiveness of Mathematics in the Natural Sciences. *Communications in Pure Applied Mathematics*, 13, February 1960.