20. How to teach social network analysis to social science students

Olga Chyzh

Social network analysis is a growing and constantly evolving field, with dedicated majors, specializations, departments, and institutes created at top universities in the US and internationally. The private sector, including the technology and financial sectors, has also exhibited a growing demand for network specialists. In this context, courses on network analysis draw high and diverse enrollments at all levels of instruction. Eager to gain an edge over their competition on the job market, students that sign up for network courses tend to come from a diverse set of majors and with varying levels of statistical and mathematical backgrounds.

In this chapter, I will draw on my teaching experience to discuss the unique pedagogical advantages and challenges of teaching a course on network analysis, compared to teaching other statistical methods. Teaching network analysis is different from other quantitative methods in several ways.

First, even beginner network students are not novices to understanding network processes—they have been a part of multiple networks, online and offline, their entire life. Educators can capitalize on this inherent familiarity to build the confidence that may counter and offset the fear of math that is so common among social science students. At the same time, too much confidence may result in resistance to learning new concepts, especially the ones that challenge students' preconceived notions of how things work.

Second, network analysis is closely linked to visualization. Whereas visualization is a powerful diagnostics tool for working with all types of data (Tufte 2001; Wickham et al. 2015), few students find a Q-Q plot as engaging as a network graph. Pedagogically, student fascination with network visualization can be channeled to getting students to think about the causal processes that resulted in the particular network configurations. The challenge, however, is to convey that network analysis is more than pretty graphs and comes with a whole set of sophisticated tools for both descriptive and inferential analysis.

Third, network analysis comes with its own unique terminology (nodes and edges), data structures (edge lists, adjacency and incidence matrices), and statistical inference tools (exponential random graph models, latent space models, local structure graph models). Unlike more traditional advanced methods topics that build on an existing statistical foundation, a course in network analysis, in many aspects, starts from square one. In this sense, a course in network analysis is somewhat of an entry-level advanced class. Unlike most advanced methods classes, where an instructor can reasonably assume the knowledge of the basics, such as the mean, the variance, and the Central Limit Theorem, in a network analysis class, you have to start from the beginning. A lack of background, however, has some advantages, for example, less time devoted to clarifying and correcting previous, incorrectly learned concepts.

After describing each of these unique advantages and challenges, I provide several examples, and conclude by sharing several broader ideas on teaching advanced methods to social science students.

BUILT-IN EXPERTISE

In my experience, teaching network analysis is unlike teaching any other methods course. No foot-dragging, no tooth-pulling, none of that "I am not that good at the maths" type of disclaimers that are so common to other quantitative methods courses (Williams and Sutton 2011). Students are actually excited to take the course. They are there to learn, and they come prepared with at least a basic understanding of the core concepts. Because unlike linear regression or Bayesian statistics, networks are in fact a natural part of our students' lives.

Always a keystroke away from their friends, classmates, co-workers, as well as myriads of strangers—politicians, advertisers, celebrities, and social media influencers—vying for their attention, our students have navigated these networks at their fingertips for as long as they remember. Natural networkers, at least as it relates to online social networks, our students have an intuitive understanding of networks and an interest in learning how to understand them better.

This built-in familiarity and interest translate in a pedagogical advantage for teaching network analysis compared to other advanced methods. The availability of everyday examples facilitates introduction of concepts at every level of difficulty, reinforcing and accelerating learning of otherwise complex material.

I like to start the first class with an exercise, adopted from Lave and March (1993). I ask the students to explain the known feature of the living arrangement in the student dormitories, specifically, that students who are friends with each other usually live on the same floor or even share a room. I ask the students to propose causal mechanisms that could lead to this living arrangement. What I am asking in this exercise—to provide a causal explanation for human friendship networks—is a non-trivial theoretical task. Yet, my new research assistants are also both the research subjects and the experts on the topic of the study. If anyone understands anything about the dynamics of an undergraduate dorm, it is the students who live there. Drawing on this inherent expertise, I help students feel confident in their network modeling skills. Before they know it, we develop a sophisticated theory of network formation that involves complex network dynamics, such as homophily, preferential attachments, and triadic closure. This exercise allows for introducing network analysis terminology in the context of simply naming the processes that have been familiar for a long time.

VISUALIZATION

Another pedagogical advantage of teaching network analysis is the close connection between networks and visualization. Teaching network analysis is virtually impossible without teaching network visualization. Network data are not easily displayed in tabular form or reducible to conventional numerical summary statistics. While some of the numerical equivalents to conventional statistical summaries do exist (e.g., density, modularity, clustering coefficient), network data are often best presented using graphics and graphical summaries. The emphasis on graphics in lieu of numbers and mathematical notation may help remove some of the known stressors that impede student learning (Bos and Schneider 2009).

Unlike other areas of visualization, interpreting and understanding network graphs does not require much statistical background. Network visualizations are always simple, even if the network itself is very complex. At the same time, visualizations of even the simplest networks are intriguing. Students can spend a lot of time brainstorming theoretical processes that may have resulted in a particular network realization. In fact, there is no better way to introduce a network concept, however complex, than by showing a visualization.

Emphasis on visualization helps break up tedious mathematical exposition, turning learning networks into a fun activity, and one with immediate gratification—using the igraph package in R (Csardi and Nepusz 2006), for example, you can teach students how to make rather sophisticated network graphs in under an hour.

Figure 20.1 shows an example. The figure depicts the largest connected component of Coleman's (1964) network of interactions among high-school students. The goal of this visualization is to show the consecutive waves of contagion, should one of the network nodes (in this example, node 21) be infected with a contagious disease. If this disease is spread on contact, then the first round of the affected students will contain the direct contacts of node 21, that is, every student that shares an edge with 21. The second round will affect the nodes that are removed from 21 by the shortest path of length two, and so on.

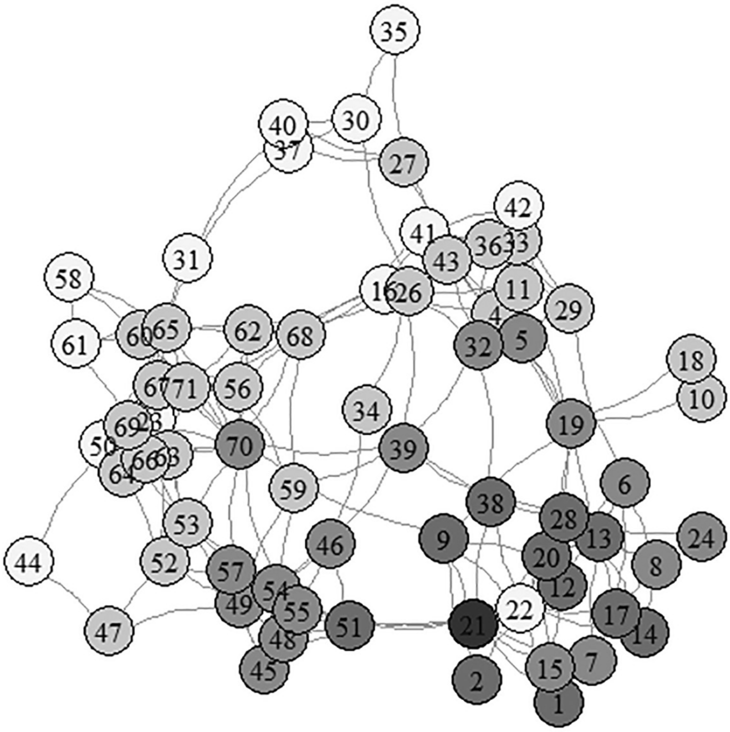

Figure 20.1 Contagion in the Coleman Interaction Network

Working through this intuitive example provides opportunities to introduce some rather complex concepts and operations. For instance, identifying the set of nodes that are removed from a given node by the shortest path of length two involves squaring the adjacency matrix. Emphasis on the visualization helps soften the shock of the impromptu use of matrix algebra, while connecting unfamiliar mathematical operations to direct practical application helps demystify their function.

Visually engaging network graphics help maintain and reinforce the initial interest as we progress towards more advanced material. The ease of visualizing concepts, such as modularity, incentivizes learning the theory behind these concepts. A visualization instantly clarifies the subtle distinctions between, say, betweenness and closeness centrality. As a result, by the end of the first week students are equipped with the basic network graphics tools that will help them visualize virtually any kind of network or network concept.

NETWORKS AS A DISTINCT AREA

As a methodological approach, network analysis is somewhat distinct from the rest of the field of quantitative methodology, with a closer connection to mathematics and computer science than statistics or econometrics. It relies on its own terminology, data structures, and statistical packages. Even network visualization is based on tailor-made network visualization packages.

This disconnect between network tools and the more conventional statistical tools learned in introductory classes has a somewhat equalizing effect, placing beginner students on a more equal footing with the students with a lot of statistics background.

The inability to directly draw on concepts learned in previous statistics classes is both a pedagogical advantage and a challenge. On one hand, starting from the beginning limits the scope of the material that can be covered in a single course. On the other hand, the foundational concepts are fresh in the students' minds throughout the course, and less class time is spent bringing everyone on the same page or reminding students of the material covered in previous classes.

CHALLENGES

While visualizations help enhance student experience in any methods course, teaching network analysis requires a mastery in network data visualization using various statistical packages. Making visualizations takes up the bulk of class preparation time, as some concepts, for example, different types of modularity, are only made intuitive in visual terms.

There is an important distinction between teaching a course on network analysis with the help of visualizations and teaching a class on data visualization, but at times, an instructor of the former may find herself teaching the latter. Be prepared to answer very specific graphing questions, for example, creating colors using the RGB color space or controlling the aspect ratio of the graph.

At times, the focus on visualization may give some students a false impression that network analysis is less rigorous than other quantitative methods. This misperception, however, is quickly addressed by mixing in some more formal mathematical presentation.

Network analysis appeals to a wide variety of students from different fields of study and varying levels of statistical background. Teaching a diverse course like this is similar to simultaneously teaching multiple courses at different instruction levels and with applications to different fields. To enhance the learning experience of different groups of students, I like to collect some background information on students' areas of research interests using a short survey at the beginning of the first class. I use the information from this survey as a guide for selecting examples throughout the remainder of the class.

To account for varying levels of statistical preparation, I present the material using a layered approach: start by painting the big picture, appealing to intuitions rather than math, then go over the same material using more specialized terminology, and finally, showing a more formal mathematical derivation. I make sure to give explicit warning before the more advanced parts of the presentation, so that the students do not feel overwhelmed.

ADDITIONAL PEDAGOGICAL ADVICE

When teaching any quantitative methods courses, I find it useful to split up the course time into roughly equal amounts of lecture and practice. During the lecture portion, I present the material and go over the mathematical notation. During the practice component, I show how to re-create the lecture materials in R and ask students to complete several short exercises to reinforce what they learned. This way, the students go through material three times: passively when they listen to my initial presentation; actively when they follow along as I show how to implement the newly learned tools in R; and independently as they complete the short exercises. Repetition allows time to process any difficult material, while completing the exercises helps identify any areas that may need clarification.

To further reinforce learning, I use weekly or bi-weekly assessments. I prefer to give short but frequent assignments, which are promptly graded and returned with feedback. Quick turnaround allows for catching and clarifying any problem areas before moving on to the next topic. As I grade assessments, I take notes of these areas, so that I can pre-emptively address them in the next iteration of the course.

Frequent assessments and quick turnaround create a significant grading burden on the instructor, especially if assignments include a coding component. To streamline grading, I require all assignments to be submitted in Rmarkdown as well as html format. Rmarkdown functionality allows students to produce a single html document that includes the R code used to complete the assignment, the statistical output, and the text answers to any questions. The requirement to compile Rmarkdown documents into html format ensures that the students debug their own code before submitting the assignment, which further cuts down on grading time. This format allows the instructor to view and grade assignments online (e.g., on Canvas), rather than downloading, possibly debugging, and running each assignment on a local computer.

KEY TAKEAWAYS

- Teaching network analysis is different from other quantitative methods; even beginner network students are not novices to understanding network processes—they have been a part of multiple networks, online and offline, their entire life.
- The close link between network analysis and visualization facilitates the introduction of complex material.
- Network analysis comes with unique terminology and inferential statistics methods, which limits the parallels to other statistics courses.

REFERENCES

Bos, A. and Schneider, M. (2009). Stepping around the Brick Wall: Overcoming Student Obstacles in Methods Courses. *PS: Political Science and Politics*, 42(2), 375–383.

Coleman, J.S. (1964). *Introduction to Mathematical Sociology*. New York: Free Press.

Csardi, G. and Nepusz, T. (2006). "The igraph software package for complex network research." *InterJournal, Complex Systems*, 1695. https://igraph.org.

Lave, C.A. and March, J.G. (1993). *An Introduction to Models in the Social Sciences*. United Kingdom: University Press of America.

Tufte, E.R. (2001). *The Visual Display of Quantitative Information*. United Kingdom: Graphics Press.

Wickham, H., Cook, D., and Hofmann, H. (2015). Visualizing Statistical Models: Removing the Blindfold. *Statistical Analysis and Data Mining: The ASA Data Science Journal*, 8, 203–225. https://doi.org/10.1002/sam.11271

Williams, M. and Sutton, C. (2011). Challenges and Opportunities for Developing Teaching in Quantitative Methods. In Payne, G. and Williams, M. (Eds.), *Teaching Quantitative Methods: Getting the Basics Right* (pp. 66–85). Los Angeles: SAGE Publications.