Available online at www.sciencedirect.com
SCIENCE DIRECT®
ELSEVIER
Social Networks 28 (2006) 269–274
SOCIAL NETWORKS
www.elsevier.com/locate/socnet

Book review

W. de Nooy, A. Mrvar, V. Batagelj, Exploratory Social Network Analysis with Pajek, Cambridge University Press, New York, 2005.

Indulging the language of my childhood, this is a rip snorter of a book. Sure, it has some irritating flaws but they pale in comparison to the excellence of the content and the caliber of the presentation. Here endeth the short review.

Modern social network analysis came of age in the era of personal computers and has developed rapidly with many conceptual, substantive, and technical advances in the last three decades. In Freeman's (2004) narrative, the modern form of the social network field is characterized by four features: first, we focus on social structure in the form of social ties among actors. This commitment to studying structure has always been a part of the field. Second, the field is grounded in *systematic empirical* structural data. Empirical information has often been present in the past but the extensive and systematic nature of contemporary data is new. Third, we draw heavily on graphical imagery. Fourth and finally, we rely heavily on mathematical models as well as sophisticated computational models. These features, when coupled, facilitate the transformation of network analysis from the realm of 'metaphor' to the realm of 'substance' (Wellman, 1988) and the creation and mobilization of rigorous and sophisticated technical procedures so that we have 'methods' to match structural substance.

The third and fourth features of Freeman's list are based upon, and demand, sound computer software—software that evolves as social network analysts evolve. It is no surprise, then, that there has been a proliferation of such software – some commercial and some freely available – to match the ambitions of the social network analytic community. Indeed the website of INSNA (The International Network of Social Network Analysts, a scholarly association) has a page http://www.insna.org/INSNA/soft_inf.html listing much of the available software. Also, there have been comparative reviews of social network software, for example, Huisman and van Duijn (2005), to help network analysts select software appropriate to their substantive goals, intentions and ambitions. Pajek (Batagelj and Mrvar, 2005) is one of the freely available packages. It is also a superb piece of software that, alas, appears to suffer from the reputation for having a manual that is 'not user friendly', for being a bit opaque in terms of what it offers, and for being difficult to use. Such an impression is reinforced by the review of Huisman and van Duijn. In this context, *Exploratory Social Network Analysis with Pajek* is a very welcome and important book. But it is far more than a corrective to the false impression that Pajek is user unfriendly. It is also a fine introduction to social network analysis.

While it is wonderful that so many scholars are turning to social network analysis in their research and seek out appropriate software, there is a downside to this upsurge of interest. I think a user of network analytic software needs to make some effort to understand the foundations of

0378-8733/$ – see front matter © 2005 Elsevier B.V. All rights reserved.
doi:10.1016/j.socnet.2005.12.002

the network analytic tools they use, to check that these tools are appropriate, and to have some understanding of what the output means both technically and substantively. From my reading of online help and discussion venues for social network software, many users seem unwilling to make that investment. They seem loathe to read the manuals that accompany the software they use, an unwillingness that extends to ignoring the technical literature completely. The notion of "user unfriendly" becomes chameleon and, in many ways, invoking it becomes a specious critique of software. The charge of "not user friendly" often translates as "not friendly to users who have not bothered to read the manual or anything about the tools". While the Pajek manual is not a simple read, the many examples that are included in it make the program straightforward to use. When applied to Pajek, the charge of being 'user unfriendly' is unfair. *Exploratory Social Network Analysis with Pajek* makes it clear that, when understood, the logical structure of Pajek ensures that it can be used easily.

The authors' introduction to network analysis has been restricted (by choice) in two ways. First, the book deals only with 'exploratory network analysis' having four parts: defining networks; manipulating networks; determining the structural features of networks; and (visually representing and) inspecting them. The second restriction is to eschew presenting mathematics and displaying formulae. One implication of these decisions is that different readers will get different rewards when reading the book. While everyone can get a clear sense of how to use Pajek from reading this book, mathematically inclined readers may be a little disappointed on the technical side. However, there is a large technical literature that such readers can consult. Similarly, readers wanting more on statistical network analyses will be disappointed because they are not there (by choice). Introductions are introductions and not everything can be covered. The two restrictions on the content of this book are eminently sensible. *Exploratory Social Network Analysis with Pajek* provides a lively introduction the network analytic ideas and a detailed discussion/demonstration of using them to analyze real social network data.

The book is divided into five broad sections, each containing chapters and there are three useful appendices. Part I deals with fundamentals with a chapter on looking for social structure and a chapter on attributes and relations. (Both actors and relations can have attributes.) Part II discusses cohesion, with three chapters devoted to cohesive subgroups, sentiments and friendships, and affiliations. Part III, also comprising three chapters, deals with brokerage. The specific chapters cover center and periphery, brokers and bridges, and diffusion. Part IV concerns rankings and has three chapters covering prestige, ranking, and genealogies and citations. Part V is on social roles, with a single chapter on blockmodels.

In their preface, the authors note that their book integrates theory, applications, and professional software for social network analysis. The claim is correct. They also provide a useful diagram (Fig. 1, p. xxv.) showing how the chapters are linked in terms of their network analytic content. For neophytes, the first three chapters are essential reading while other readers can pick and choose. However, if the intent of a reader is to learn how Pajek is structured and how to use the program fruitfully, following the chapters in their presented order makes the most sense.

Every chapter is very useful overall, and each is peppered with helpful ideas and suggestions. Some suggestions are so useful that I would have bolded them. For example, the authors write (p. 11): "When you try to locate a command in Pajek, just consider which data objects you wish to use". Keeping track of networks, partitions, permutations, clusters, hierarchies, and vectors as data objects is an *essential* aspect of becoming fluent in Pajek.

Pajek has many drop-down menus that are used to create a sequence of operations to obtain something of value. So, suppose, to use one of the early examples, that an analyst wanted to draw a network where reciprocal binary arcs are replaced by edges. The analyst starts in the main menu

and clicks on Nets (for Networks) as this is the relevant data object. The command sequence is:

[Main]Net > Transform > Arcs → Edges > Bidirected only > Min value

The logical structure is compelling. (In this example, there are other options for the command sequence. Instead of choosing the minimum value, the analyst could select the maximum value (that would make no difference in this example) or the sum of values (which would make a difference) to the output.)

A useful display device used throughout the book is putting (abbreviated versions of) command sequences in the outside margins of the text. Of course, these are coupled to the network concepts that are discussed on the relevant page—but they are useful also for locating command sequences for obtaining specific output.

Many data objects can be available with the selected (highlighted) object being active. For networks, the highlighted network is the active network. Operations – for example, transforming reciprocal arcs to edges – are used to transform the network in some fashion. The authors note that after the use of such an operation (or a sequence of them), a dialogue box appears asking whether or not a new network is to be created. The authors write (p. 12) "If the answer is yes, which we advise, a new network . . . is added" to the available networks in the network window. I would have inserted 'strongly' and bolded the sentence. This way, the original network is left unchanged. Keeping track of the data objects that have been created and which ones are active is the critical responsibility of the user. In this regard, Pajek is uncompromising and my advice to all users is that they keep track of the data objects that have been created.

Pajek has some characteristic features that should be noted. It is designed to deal with very large networks where "large" is whatever can be stored in a computer memory. It follows that there is an emphasis on efficient programs. As a result, networks with millions of vertices can be studied. A second feature is that it contains many basic operations having a clear logical structure.

One consequence is that a sequence of these basic operations has to be used in many, if not all, analyses. I gave one example above. Another is the computation of "degree centrality" that is introduced in Chapter 6. From a 'one button' perspective, a network analytic package should allow us to compute degree centrality that is labeled directly as such. In Pajek, the command sequence to get degree centrality seems odd from this perspective: Nets > Partitions > Degree. Yet it makes eminent sense. The analyst is, in effect, creating a partition of the vertices according to their degree. The command does this and the result can be stored as partition and as a vector. You might ask "So what?" When stored as a vector any measure of centrality can be used to draw the network so that the size of the vertices in the networks reveals the relative magnitude of their centrality scores.

The design of Pajek is much broader than having a set of distinct tools. The ability to string together operations brings with it immense flexibility. This is an integral feature of Pajek—but it demands that users understand the potential command sequences that are available within the program. Let me emphasize that *this is not a criticism of Pajek.* Instead, it points to its huge range of available analyses. I suspect that the impression of Pajek being user unfriendly is driven, in part, by this feature which demands that users learn its underlying logical structure.

A major part of exploratory network analysis is visual, and considerable space in this book is devoted to visualization. The authors are acutely aware that pictures can be both compelling and misleading. They recommend, in the main, the use of the automatic procedures – involving spring embedders – but doing so with some experimentation. They are skeptical about the value of drawing networks from scratch manually if the goal of the drawing is *discovery* of structure:

analysts can build into the picture the visual structure of the features they want "to find". But the authors are not doctrinaire about this because users can manipulate visual images.

Until reading Chapter 1, I had not realized that Pajek has some built-in constraints on the amount of manual manipulation of a network, provided that they are switched on as an option. Using these options makes great sense—and will help prevent users from producing poorly drawn networks.

Another sentence that I would have bolded appears on page 16: "Note the difference between a command and an option in Pajek: a command is executed once while an option remains effective until it is turned off". Turning on the drawing constraints, as an option, is remarkably helpful, and I advise users to do so.

Some more advice: having drawn a network on the screen, do not assume that it (with its new coordinates) has been saved automatically. Save it explicitly. In fact, save *any* created data object that you may want later *explicitly*. Even more advice: *always* keep track of what you have done in an analysis sequence.

Readers of Chapter 1 are referred to Appendices 1 and 2 and both are critical for using Pajek. Appendix 1 is on 'getting started with Pajek' while Appendix 2 is on 'exporting visualizations'. I would have preferred an expanded coverage of these topics—especially those in Appendix 2. More could also have been made of Fig. 141 (p. 308) on setting options for drawing networks.

One feature of Pajek that is unsettling, at first, is that the image on the screen need not be the image that appears on a printed page. Trying to develop a printed image that differs from the screen image is difficult, and setting options in the options screen becomes essential. The authors recommend the use of Encapsulated Postscript (EPS) for figures, a bias that I share. They also recommend the use of GhostScript and GhostView as software for viewing EPS files. They note that most Windows applications cannot show EPS figures, but that has changed with Windows XP where EPS figures can be inserted easily into Word documents.

At face value, the set theoretic idea of a partition of the vertices of a network does not have a lot to offer graphically. Chapter 2 dispels that notion quickly and using partitions is essential in using Pajek to create compelling images. Partitions store discrete characteristics of vertices and these characteristics can be either structural or some property measured independently of the network. The authors (Fig. 17) show a dramatic image of a world system trading network. When it is drawn in color – with the colors of vertices attached to the attribute of world system position – the image is breathtaking. The gray scale of the book does not do justice to the colored images drawn in Pajek. However, the authors have made a website (http://vlado.fmf.uni-lj.si/pub/networks/book) available with the data sets used throughout the book and readers are encouraged to do the analyses (with the command sequences provided in the book) as they read. Again, partitions are a data object and a set of related data objects – that include networks – can be stored in a Pajek 'project file'. Doing so is very useful. Chapter 2 also makes clear how extracting parts of networks and shrinking parts of networks can be used effectively in delineating the structure of large networks.

The authors are clear that there is a sharp boundary between exploratory network analysis and testing hypothesis about the operation of social networks. It is an uneasy distinction and the authors occasionally stray over that boundary in an unconvincing fashion. For example, this happens in Chapter 3 with a discussion of cohesive subgroups. In part, the distinction is made to restrict the scope of the book, something that is very reasonable—but it remains problematic. Pajek does have some primitive 'statistical' analyses available and they are mobilized in this book. More important is the ability to export to the program R, mentioned by the authors, and points to sophisticated statistical analyses that are available through R. The authors argue that social

network analyses and statistical analyses are complementary activities. This is unarguable but I would have liked to see more on that critical interface.

Chapter 4 is devoted to sentiments and friendship with 'structural balance theory' playing a central role. The discussion is constructed well and is persuasive. The one thing that troubles me is the use of the term "error" and error scores. The theory that motivates this chapter predicts that if a signed social network is balanced – in the sense of Heider (1946) – then it has two clusters of vertices where the positive ties go between actors within a cluster and the negative ties go between actors in different clusters. If the network is 'clusterable' – in the sense of Davis (1967) – then there will be two or more clusters with positive ties 'within' the clusters and negative ties 'between' the clusters. Pajek can be used to locate partitions *as close as possible* to this 'ideal' state (of balanced signed networks) for networks that are neither exactly balanced nor clusterable. The authors, and Pajek, label ties that are inconsistent with an ideal pattern "errors". I would much prefer that they are called "inconsistencies" between the ideal partitioned world of balance theory and the partitioned empirical world in a network data set. (This applies also to Chapter 12 on blockmodeling.) Of course, this sounds like a trivial complaint and perhaps it is. However, when we have theories strong enough to permit a 'theory of errors' – and can estimate the statistical models based upon them – then I will not mind labeling the 'inconsistencies' as 'errors'. This would also allow us to straddle the boundary between exploratory network analyses and statistical analyses of network data. Chapters 8–12 in the edited collection of Carrington et al. (2005) describe some very promising avenues in this direction.

Much of network analysis has been focused on one-mode networks. More and more I am coming to think of two-mode networks as more fundamental than one-mode networks.[1] Given this bias, I was delighted to find Chapter 5 devoted to affiliation networks as two-mode data. Pajek has some very nice tools – especially *m*-slices – for exploring two-mode data and they are mobilized nicely in Chapter 5. Every two-mode network can be used to construct two one-mode networks for further analyses. Pajek makes this transparently easy to do. If readers have their own data sets, these too can be analyzed in parallel to the narrative in the text—an activity that increases the value of the book.

Chapter 6 brings into focus a problem with the book. Readers can read this chapter and not see a reference to the Freeman (1979) paper on centrality that stands as one of the conceptual high points of the field. (It is mentioned later in the chapter on citations.) Because 'textbooks' are tightly focused on the materials being introduced – and not their history – there is an excuse for being rather autistic with regard to what has gone before. While *Exploratory Social Network Analysis with Pajek* has many characteristics of a textbook, it is not just a textbook. And to the extent that it is not, the 'textbook autism' is problematic. Each chapter has a section on further reading. My sense is that the authors could have provided more than they did on the provenance of network ideas.

While on the topic of textbook features, I strongly compliment the authors on their use of exercises and questions. They are excellent in every chapter. The exercises involve data analyses based on the materials of the chapter – with data sets provided by the authors – and the questions are more conceptual. I urge readers to spend time with these questions and exercises *before* going

[1] One-mode network data are defined for one set of social objects. Most often the social objects are individuals and the data represent information on social relations between them. For example, the data can be about relations such as friendship, communicates with, or dominates. Another example of this data type is trading relations between nations or societies. Two-mode data are defined for two sets of social objects. Examples include people attending social events, organizations employing people, and nations belonging to alliances.

to the answers provided at the end of each chapter. These items have been constructed well and are thought provoking—they merit conscientious attention.

Researchers interested in genealogies and citation networks will find a lot to interest them in Chapter 11. But all readers will benefit from consulting this chapter even if they do not share these interests. Pajek has a wonderful ability to handle time in a dynamic fashion. Networks as links can be indexed by time and networks can be extracted for arbitrary periods of time. This adds immense flexibility. While genealogies and citation networks must be indexed by time, all network processes unfold through time. Pajek allows us to capture this and if networks are generated by event sequences occurring through time (Doreian, 2002), then Pajek facilitates looking at networks in this fashion. This is truly exciting.

The authors provide a nice introduction to blockmodeling in the final chapter. Now, if we can couple the temporal features of Pajek and blockmodeling, also implemented in Pajek, I would be in network heaven.

Wait a minute; we *can* do this—which may help explain why I am so enamored with this book and the software it presents. And when the network ideas and Pajek operations introduced in different chapters are combined, our ability to analyze networks is greatly expanded. Pajek[2] is not just a large toolbox. It is much more general and flexible. *Exploratory Social Network Analysis with Pajek* makes this abundantly clear.

References

Batagelj, V., Mrvar, A., 2005. Pajek: Package for Large Networks, Version 1.10 (October 25, 2005). University of Ljubljana, Ljubljana.

Carrington, P., Scott, J., Wasserman, S. (Eds.), 2005. Models and Methods in Social Network Analysis. Cambridge University Press, New York.

Davis, J.A., 1967. Clustering and structural balance in graphs. Human Relations 20, 181–187.

Doreian, P., 2002. Network sequences as generators of network evolution. Social Networks 24, 93–119.

Freeman, L.C., 1979. Centrality in social networks: I. Conceptual clarification. Social Networks 1, 215–239.

Freeman, L.C., 2004. The Development of Social Network Analysis: A Study in the Sociology of Science. Empirical Press, Vancouver.

Heider, F., 1946. Attitudes and cognitive organization. Journal of Psychology 21, 107–112.

Huisman, M., van Duijn, M.A.J., 2005. Software for Social Network Analysis. In: Carrington, P., Scott, J., Wasserman, S. (Eds.), Models and Methods in Social Network Analysis. Cambridge University Press, New York, pp. 270–316.

Wellman, B., 1988. Structural analysis: from method and metaphor to theory and substance. In: Wellman, B., Berkowitz, S.D. (Eds.), Social Structures: A Network Approach. Cambridge University Press, Cambridge.

Patrick Doreian *
Department of Sociology, University of Pittsburgh
2603 WWPH, Pittsburgh, PA 15260, USA

* Tel.: +1 412 648 7584.
E-mail address: pitpat@pitt.edu

[2] It is useful to consult the Pajek website often. This program is updated frequently to add new features and to remove bugs. A history of pajek is available also on the website and provides a listing of the changes that have been made.