Some Elegant Probability and Non-Probability Models for Social Networks

John Levi Martin
Fall 2019
Tuesday 3:30 – 6:20 PM; HM 102

Overview:

This is a course in response to student interest in learning the fundamentals of certain inspiring approaches to social structure which are in danger of being sidelined by the monoculture of the garbage can exponential-linear model. We look at both probability models and non-probability (deterministic) models for structure, concentrating on those that build strong theory at the cost of inflexibility. We then move to development of the exponential random graph model, and then shift gears to try to collectively examine the robustness of different approaches to using maximum likelihood estimates from such models to make conclusions about data. The "collective" portion comes in that we will try to build some simulations and learn from those who know how to fit these properly to make sure we are giving all models a fair chance.

Where You Should Be:

All you need to know for this class is

1) Basic matrix operations (multiplication, transpose, inverse)
2) Familiarity with the idea of a graph as a set of nodes and edges
3) Basic programming facility with a flexible language like R
4) Willingness to learn what you don't already know.

Set up

Each class will have two portions, with a ten minute break in between. We will go over the reading, trying to do two things: first, to derive key aspects that may have been done too quickly for us to follow in the source material, and second, to emphasize the key insights and implications. We will also examine results from applying the method (where applicable) to a data set. Students will take turns being the one to attempt an application for each unit. There is likely to be an informal klatsch afterwards on most occasions.

Requirements:

Part of this class is doing exercises on data sets; we'll have one or two examples ready, but it would be great if you have one that you are personally interested in also. Towards the end, we switch to examining the robustness of various approaches, and that might involve a bit more programming. You'll be asked to simulate some network data, and then to try the different approaches we've learned about; your write up is your final assignment.

Outline:

I. INTRODUCTION (**Tuesday, October 1**)

A. *Overview of the Class.*
Required Readings: None

Other Readings: If you want an overview, take a look at Ronald L. Breiger, "The Analysis of Social Networks," in Melissa Hardy and Alan Bryman (eds.), *Handbook of Data Analysis* (London: Sage, 2003).

B. *Probability and Non-Probability Models*
Here we are going to walk through the simplest examples of probability and non-probability models. You may know these. If not, read something on each. For the first, we're going to consider the Gilbert (Bernoulli) and Erdős–Rényi graph.

Required Readings: Dorwin Cartwright and Frank Harary. 1956. "Structural Balance," *Psychological Review* 63:277-293; For random graphs, the most common citation is to P. Erdős and A. Rényi, "On Random Graphs. I" *Publicationes Mathematicae* (1959) 6: 290–297. But most folks learn this from other pieces. I'm not sure if you should read this, or find a synthetic source. It's pretty dense, actually.

Other Readings: On balance classic works are: Fritz Heider. 1946. "Attitudes and Cognitive Organization," *Psychological Review* 52:358-374; Frank Harary. 1955. "On Local and N-balance of Signed Graphs," *Michigan Mathematical Journal* 5:37-41. Noah Friedkin and Craig Rawlings keep this tradition alive! We'll follow it into triad analysis later. On random graphs, the classic work is E.N. Gilbert, "Random Graphs," *Annals of Mathematical Statistics* (1959) 30/4: 1141–1144.

II. ANATOMY OF KINSHIP (**Tuesday, October 8**)

A. *The Idea of Marriage Classes and Prescriptive Structures*
Required Readings: I don't feel like I can really dump Lévi-Strauss on you when we also have to read White's book, but I am going to really suggest that if you never have read it, you do. And you can look at Andre Weil's mathematical contribution in the middle. That's where Harrison White takes off.

Other Readings: Claude Lévi-Strauss, *The Elementary Structures of Kinship*, pp. 12-83, 119-133, 146- 220, 232-309. (Chs. 2-6, 9, 11-13, 15-17).

B. *The Derivation*
Required Readings: Harrison White, *An Anatomy of Kinship*. Start on in…

Other Readings: …and we will see how far we get. I definitely want us to get through the first few chapters.

III. ROLE ALGEBRAS (**Tuesday, October 15**)

A. *Role Structures.*

Presumably you all get the idea of structural equivalence, but if not, there are readings below. But we'll jump into the idea of role structures.

Required Readings: Scott Boorman and Harrison White. 1976. "Social Structure from Multiple Networks II: Role Structures." *American Journal of Sociology* 81:1384-1446.

Other Readings: Francois Lorrain and Harrison White. 1971. "Structural Equivalence of Individuals in Social Networks." *Journal of Mathematical Sociology* 1:49-80; Harrison White, Scott Boorman, and Ronald Breiger, "Social Structure from Multiple Networks I: Blockmodels of Roles and Positions." *American Journal of Sociology* 81 (1976):730-779

B. *Second*.

Required Readings: R.L. Breiger and P.E. Pattison. 1986. "Cumulated Social Roles: The Duality of Persons and Their Algebras," *Social Networks* 8 : 215-256

Other Readings: Philippa Pattison. 1994. *Algebraic Models for Social Networks.* Cambridge. John Paul Boyd (1969) "The Algebra of Group Kinship" *Journal of Mathematical Psychology* 6:139-167.

IV. PROBABILITY AND NETWORKS (**Tuesday, October 22**)

A. *The Classic*.

This will probably take more than half the class. It is a really gorgeous piece of work.
Required Readings: M Newman, S Strogatz and D Watts. 2001. "Random Graphs with arbitrary Degree Distributions and their Applications." *Phys. Rev* E: 64.

Other Readings: Newman, "The Structure and Function of Complex Networks." *SIAM Review* 45:167-256. Steven H. Strogatz, "Exploring Complex Networks." *Nature* 410 (2001):268-276.

B. *Random Rewiring*.

This is conceptually simpler but represents a way of pursuing the generation of distributions with fixed characteristics, as opposed to stochastic generators.

Required Readings: Orsini, Chiara et al. [11 co-authors]. 2015. "Quantifying Randomness in Real Networks." *Nature Communications* 6:8627.

Other Readings: Duncan J. Watts. 1999. "Networks, Dynamics, and the Small-World Phenomenon." *American Journal of Sociology* 105: 493-527.*

V. TRIAD CENSUS (**Tuesday, October 29**)

A. *The U-MAN Distribution*.
Required Readings: Paul Holland and Samuel Leinhardt. 1970. "A Method for Detecting Structure in Sociometric Data" *American Journal of Sociology* 76: 492-513.

Other Readings: James Davis and Samuel Leinhardt. 1972. "The Structure of Positive Interpersonal Relations in Small Groups." In Joseph Berger, Morris Zelditch, Jr., and Bo Anderson (eds.), *Sociological Theories In Progress Volume 2*, Boston: Houghton Mifflin, pp. 218-251,

B. *Theories of Structural Tendencies*.
Required Readings: Paul Holland and Samuel Leinhardt. 1976. "Local Structure in Social Networks," *Sociological Methodology 1976*: 1-45

Other Readings: Paul Holland and Samuel Leinhardt. 1971. "Transitivity in Structural Models of Small Groups." *Comparative Group Studies* 2: 107-124. Eugene Johnsen. 1985. "Network Macrostructure Models for the Davis-Leinhardt Set of Empirical Sociomatrices." *Social Networks* 7: 203-224.

VI. THE p1 MODEL (**Tuesday, November 5**)

A. *A Parametric Representation*.
Required Readings: Paul Holland and Samuel Leinhardt, "An Exponential Family of Probability Distributions for Directed Graphs." *Journal of the American Statistical Association* 76(1981):33-50

Other Readings: None, but you can read the interesting discussion of the other folks who commented on the paper in this issue.

B. *A General Loglinear Approach*.
Required Readings: Stanley Wasserman and Dawn Iacobucci. "Statistical Analyses of Discrete Relational Data." *British Journal of Mathematical and Statistical Psychology* 39 (1986):41-64.

Other Readings: Well, the first major network methods tome, by Wasserman and Faust, *Social Network Analysis*, really spins this out into a system.

VII. HAMMERSLEY CLIFFORD THEOREM (**Tuesday, November 12**)

A. *The Autologistic Problem and The Möbius Inversion Lemma*.
Required Readings: J. E. Besag, "Nearest-Neighbour Systems and the Autologistic Model for Binary Data," *Journal of the Royal Statistical Society, Series B* 34(1972):75-83.

Other Readings: If I can find a good treatment of the lemma I will assign that.

B. *Proof of the Theorem*.
Required Readings: J. E. Besag, "Spatial Interaction and the Statistical Analysis of Lattice Systems," *Journal of the Royal Statistical Society, Series B* 36(1974):192-236.

Other Readings: Everyone uses Besag for the Hammersley-Clifford Theorem, because it's a simpler proof, and it's still a beast! I think it's easier starting with the earlier paper. It also probably is easier to start with the Pattison work next week.

VIII. TO P* (**Tuesday, November 19**)

A. *Markov Graphs*.
Required Readings: Wasserman, Stanley, and Philippa Pattison. 1996. "Logit Models and Logistic Regressions for Social Networks: I. An Introduction to Markov Graphs and *p**." *Psychometrika* 61:401-425.

Other Readings: Pattison, Philippa, and Stanley Wasserman. 1999. "Logit models and logistic regressions for social networks: II. Multivariate relations." *British Journal of Mathematical and Statistical Psychology* 52: 169-193.

B. *A Pseudolikelihood Interpretation*.
Required Readings: Wasserman, Stanley, Carolyn J. Anderson, and Bradley Crouch. 1999. "A *p** Primer: Logit Models for Social Networks." *Social Networks* 21:37-66.

Other Readings: None yet, but I might add something, perhaps from Philip Leifeld who has made recent arguments for taking PL methods seriously.

IX. MCMC AND DEGENERACY (**Tuesday, November 26**)

A. *MCMC Fitting*.
Required Readings: On MCMC methods see Mark S. Handcock *in Dynamic Social Network Modeling and Analysis*, edited by Ronald Breiger, Kathleen Carley and Philippa Pattison (National Academies Press, 2003).

Other Readings: And also Tom A B Snijders. 2002. "Markov Chain Monte Carlo Estimation of Exponential Random Graph Models." *Journal of Social Structure* 3:2. I'd certainly be interested in anyone who wanted to figure out how Gelman's *STAN* works for large networks and report back! Word on the streets is that it is unbeatable.

B. *Degeneracy and new Approaches* .
Required Readings: Tom A B Snijders, Philippa Pattison, Gary Robins and Mark Handcock. 2006. "New Specifications for Exponential Random Graph Models." *Sociological Methodology* 36:99–153.

Other Readings: David R. Hunter. 2007. "Curved Exponential Family Models for Social Networks." *Social Networks* 29:216–230. This changed the way some of these effects are parameterized, if I remember correctly.

X. FRONTIERS (**Tuesday, December 3**)

NOTE: We may substitute readings discussing issues of fragility and pseudolikelihood estimation here if we find good ones. Note that we might also swap out a discussion of general methods of factoring matrices if that turns out to be of interest. That alternate module is given below.

A. *Robustness and/or SIENNA*

Required Readings: Tom A.B. Snijders. 2014. *Siena: Statistical Modeling of Longitudinal Network Data*. Pp. 1718-1725 in Reda Alhajj and Jon Rokne (eds.) Encyclopedia of Social Network Analysis and Mining. New York: Springer. DOI: http://dx.doi.org/10.1007/978-1-4614-6170-8_312. Yup! Grab it here from a U of C machine.

Other Readings: Per Block, Christoph Stadtfeld, and Tom A. B. Snijders. 2019. "Forms of Dependence: Comparing SAOMs and ERGMs from Basic Principles." *Sociological Methods and Research* 48: 202-239.

B. *TERGMS*.

Required Readings: Philip Leifeld and Skyler J. Cranmer. 2019. "A Theoretical and Empirical Comparison of the Temporal Exponential Random Graph Model and the Stochastic Actor-Oriented Model." *Network Science* 7(1): 20-51

Other Readings: Steve Hanneke, Wenjie Fu, and Eric P. Xing. 2010. "Discrete Temporal Models of Social Networks." *The Electronic Journal of Statistics* 4, 585-605. Note: this is what everyone cites, but I haven't read it and don't know how hard it is.

XA. FRONTIERS

This is the alternate module on rotation and reduction.

A. *Quantitative: Singular Value Decomposition*

Required Readings: One could start with Krishnan Namboodiri's "little green" Sage book *Matrix Algebra* and then…go to Wikipedia? And we'd derive in class?.

Other Readings: Bridgette Le Roux and Henry Rouanet, *Geometric Data Analysis*. It's rather abstract at the beginning.

B. *Qualitative*.

Required Readings: Ki Hang Kim. 1982. *Boolean Matrix Theory and Applications*, selection.

Other Readings: Philippa Pattison, *Algebraic Models for Social Networks*.