Dana Warmsley

Math 4740 – Prof. Eldredge

Final Paper

Wasserman, S. (1980). "Analyzing Social Networks as Stochastic Processes." Journal of the American Statistical Association, 75(370), 280 - 294.

A social network is a group of people connected to one another via relationships and/or interactions. Examples of social networks are families, schools, organizations, and more recently, emerging online structures such as Myspace and Facebook. Social network analysis, which began to come to prominence in the 1950s, allows for the study of social networks using mainly graph theory, statistics and algebra. The field's importance lies in the fact that the world is made of social networks of all different sizes, and so social network analysis can give light to almost any question involving human interaction. Whether it is understanding how information is spread through an educational social network, how outbreaks of diseases turn into epidemics, or how economic events or news affect an entire financial system, social network analysis is able to shed light.

Much of the major research in social network analysis thus far has been aimed at understanding 1) how social networks grow/decline, and 2) how information and innovations flow through these networks. Wasserman's "Analyzing Social Networks as Stochastic Processes" (1980) was one of the first articles to propose a methodology for studying the change in network structure over time via stochastic processes, building on the ideas of authors such as Katz and Proctor (1959), Kauffman (1969), Holland and Leinhardt (1975). Wasserman's major criticism of the discourse on social networks at the time was that they were being studied as static objects, and that any conclusions regarding longitudinal data were usually qualitative in nature. Assuming a network reaches equilibrium, he sought to understand 1) the ways in which the present structure of a network influences its future structure, and 2) the movement of networks toward their equilibrium states.

Wasserman proposed two stochastic models for social networks – the Reciprocity Model and the Popularity Model – based on Holland and Leinhardt's (1977) modeling framework: Let $X(t)$ be a binary matrix-valued stochastic process describing relationship choices in a social network. Its elements, $X_{ij}(t)$, are equal to 1 if node i chose node j (i→j), and 0 otherwise. As outlined by Wasserman, there are two major assumptions in this framework. First, $X(t)$ is a continuous-time Markov chain with finite state space. This assumption guarantees us that the future state of the network, as characterized by its relationship ties, depends only on its current state and not past states. Second, the changes in relationship ties in the network are independent over small periods of time. This assumption makes it unlikely that more than one relationship change will occur at the same time during the extremely small time interval (t, t+h), especially as amount of time h approaches 0. As there are only two possible changes that can occur, the creation or destruction of a relationship, the transition probability matrix is fully defined by the following equations:

$$P\{X_{ij}(t+h)=1 \mid X(t)=x, X_{ij}(t)=0\} = h\lambda_{0ij}(x,t)+o(h)$$
$$P\{X_{ij}(t+h)=0 \mid X(t)=x, X_{ij}(t)=1\} = h\lambda_{1ij}(x,t)+o(h)$$

, where h represents the small

amount of time passed in each interval, λ represents the infinitesimal transition rates for $X(t)$, and o(h) represents the probability of more than one change occurring simultaneously in the network.

Wasserman proposed two models based on this framework, the first of which is the Reciprocity Model. This model best applies to networks in which relationships are not symmetric – an individual may choose to be friends with someone, but this friendship may or may not be reciprocated. In order to quantify this aspect of relationships, Wasserman defined $D_{ij}(t) = (X_{ij}(t), X_{ji}(t))$, the dyad for nodes (i,j), to be a continuous-time Markov chain with state space D= {(0,0),(1,1),(0,1),(1,0)}, where (1,0), for instance, means that node i chose node j, but node j did not choose node i. Each dyad is generated by Q, the matrix of infinitesimal transition rates between the states of the network, as defined by the following time-homogenous equations: $\begin{matrix} \lambda_{0ij}(x,t) = \lambda_0 + \mu_0 x_{ji} \\ \lambda_{1ij}(x,t) = \lambda_1 + \mu_1 x_{ji} \end{matrix}$. From these transition rate equations, we can see that a choice i→j depends only on whether or not j→i already exists, how "important" that reciprocated tie is (as measured by the parameters μ_0, μ_1), and the overall rate of change of the network (as measured by the parameters λ_0, λ_1). The parameters of this model can be estimated by analyzing the change in dyadic ties over time in your actual social network. Since X(t) is stationary in time (due to its time-homogenous transition rates), Wasserman was able to use properties of stochastic processes, such as the reversibility equations, to calculate the vector of equilibrium probabilities $\pi(\infty)$ for the four states of the reversible dyad process.

The second model proposed by Wasserman was the Popularity Model. The time-homogenous transition rates for the Popularity Model, $\begin{matrix} \lambda_{0ij}(x,t) = \lambda_0 + \pi_0 x_{+j} \\ \lambda_{1ij}(x,t) = \lambda 1 + \pi_1 x_{+j} \end{matrix}$, indicate that a node j gets chosen based only on the amount of nodes that have already chosen j (j's indegree, as indicated by x_{+j}), how "important" j's popularity is (as measured by the parameters π_0, π_1), and the overall rate of change of the network (as measured by the parameters λ_0, λ_1). Since the indegree of a node is representative of a node's "social status" in the network, this allows for the understanding of how changes in relationships differ for the more popular nodes. Instead of studying the dyads of the network as a stochastic process, this model focuses on X(t)'s column processes. More specifically, Wasserman considers each column of the binary matrix X(t) as an independent and identically distributed stochastic process and extracts information about indegree from each column. This indegree process, $I_j(t) = \sum_i X_{ij}(t)$, is itself a Birth-Death chain for which the equilibrium probabilities are well known and were multiplied in order to get the equilibrium probabilities of the entire process.

In order to "propose simple but substantively interesting models that serve as benchmarks or null models against which data on groups with complicated structure can be compared" (pg. 285), Wasserman tested the Reciprocity model against the data used in Katz & Proctor (1959), and tested both models against the Newcomb (1961) data. Overall, analyses resulting from the Reciprocity model were found to support the conclusions of the papers from which the data originated. With the focus of finding the order and distribution properties of the Markov chain needed to study dyadic changes, Katz & Proctor (1961) asked 8th graders which 3 students they wanted to sit with at four different times during the year. After estimating model parameters based on the empirical data, Wasserman came to many of the same conclusions as the original authors, including the existence of a "sex cleavage" in which the majority of relationship choices made were between the same genders. Wasserman also offered a method to incorporate/model this sex cleavage into the original Reciprocity Model, in order to further understand the lack of a reciprocity effect on same-sex relationships.

In Newcomb (1961), two fraternities of 17 freshmen each were asked to rate each of the other members on a scale of 1 to 16. After performing analyses on the binarized form of this data, Wasserman confirmed many of the qualitative assessments made by Newcomb. In particular, Wasserman found that there was a small effect of reciprocated choices on the creation of new choices, but an even larger effect on the disappearance of choices. In other words, an existing choice is more likely to disappear if that choice is not reciprocated. In addition, Wasserman confirmed that there was a tendency in the network toward reciprocation, and that those high-attraction reciprocated pairs are the more stable pairs of the network. The Popularity model, tested against the Newcomb data only, also resulted in the confirmation of many of Newcomb's findings. Wasserman found that as networks approach equilibrium, the popular individuals (as defined by indegree) tend to remain so, and the less popular individuals might even become less popular.

Overall, both of these models seem to do well in fitting and explaining real social network data. Some major advantages of using stochastic processes to model network change include the available statistical measures that allow one to quantify long-term behavior of Markov chains, as well as the use of changeable parameters that can be adjusted to one's particular data. In addition, some of the more contemporary models of social network growth stem from these ideas of reciprocity and popularity. Wasserman's analysis of popularity offered early empirical evidence of what has come to be known as the "rich get richer" phenomenon or "preferential attachment", quantitatively represented by the Power-Law distribution. Barabasi and Albert's scale-free network model (1999), in particular, is based on this idea of preferential attachment, in which popular individuals tend not to just remain popular but become more popular over time. Snijders (1996, 2010) extended Wasserman's models such that changes in the network depend on both reciprocity and popularity, as well as activity (as measured by outdegree), balance (the extent to which friends share the same friends), transitivity (tendency of a node's mutual friends to also become friends), and a combination of other network effects.

Still, there are some major drawbacks. Many of the assumptions of the models limit their applicability to real world data. First, the use of a Markov chain to model a social network can be very limiting. It is often unrealistic to assume that the creation/destruction of a relationship between two people is based only on the current status of that relationship and not the history of the relationship. The continuous-time assumption of the Markov chain may prove difficult when using longitudinal network data collected at discrete points in time, and the time-homogenous assumption requires that the real social network have fixed transition rates. Another assumption that may be unrealistic is that dyads are independent and identically distributed, with only one change in a network allowed to occur in a small period of time. An alternate model that allows for multiple changes in dyads can be found in Mayer (1977). In real networks, changes to the network often happen simultaneously and not always independently. Indeed, the assumption of dyadic independence precludes us from understanding the importance of other network effects, such as transitivity. A final issue is that these models do not allow for the change in size of the overall network. While the networks here are considered dynamic over time, the population is considered static - an unrealistic assumption for real social networks that persists in many network models today. Wasserman acknowledges many of these limitations, noting that one must take extra care to make sure their network can be modeled by its parameters.

Bibliography

1. Barabási, A.L. & Albert, R. Emergence of scaling in random networks. Science. 286:509-512. 1999.

2. Holland, P.W., and Leinhardt, S. (1970), “A Method for Detecting Structure in Sociometric Data,” American Journal of Sociology, 70, 492-513.

3. Katz, L. and Proctor, C.H. (1959), “The Concept of Configuration of Interpersonal Relations in a Group As a Time-Dependent Stochastic Process,” Psychometrika, 24, 317-327.

4. Kauffman, S.A. (1969), “Metabolic Stability and Epigenesis in Randomly Constructed Genetic Nets,” Journal of Theoretical Biology, 22, 437-467.

5. Mayer, T.F. (1977), “Parties and Networks: Stochastic Models for Relationship Networks,” paper presented at the MSSB Advanced Research Symposium on Stochastic Process Models for Social Structure, Carnegie-Mellon University, Pittsburgh, Pa.

6. Newcomb, T.M. (1961), “The Acquaintance Process,” New York: Holt, Rinehart, & Winston.

7. Snijders, T.A.B., Stochastic actor-oriented dynamic network analysis. Journal of Mathematical Sociology, 21 (1996), 149-172.

8. Snijders, T.A.B., Steglich, C.E.G., and van de Bunt, G.G. (2010). Introduction to actor-based models for network dynamics . Social Networks, 32, 44-60. http://dx.doi.org/10.1016/j.socnet.2009.02.004

9. Wasserman, S. (1980). “Analyzing Social Networks as Stochastic Processes.” Journal of the American Statistical Association, 75(370), 280 - 294. http://www.jstor.org/stable/2287447?seq=1