The Annals of Applied Statistics
2010, Vol. 4, No. 1, 1–4
DOI: 10.1214/10-AOAS346

INTRODUCTION TO PAPERS ON THE MODELING AND ANALYSIS OF NETWORK DATA

BY STEPHEN E. FIENBERG

Carnegie Mellon University

In today's world, networks seem to appear everywhere. There are social networks, communication networks, financial transaction networks, gene regulatory networks, disease transmission networks, ecological food networks, mobile telephone and sensor networks and more. We, our professional colleagues, our friends and family, and especially our students, are often part of online networks such as *Facebook*, *LinkedIn* and now *Google Buzz*. Some network structures are static and others are dynamically evolving. Networks are usually represented in terms of graphs with the nodes representing entities, for example, people, and the edges representing ties or relationships. Edges may be directed or undirected depending on the application and substantive question of interest. In terms of statistical science, a network model is one that accounts for the structure of the network ties in terms of the probability that each network tie exists, whether conditional on all other ties, or as considered part of the distribution of the ensemble of ties.

Ideas and language from graph theory abound in the technical literature on networks. A typical representation involves a network with N nodes, having $\binom{N}{2}$ unordered pairs of nodes, and hence $2\binom{N}{2}$ possible directed edges. If the labels on edges reflect the nodes they link, as (i, j), Y_{ij} represents the existence of an edge from individual i to j, and $\{\mathbf{Y}\} = \{Y_{12}, Y_{13}, \ldots, Y_{(N-1)N}\}$ represents the ties in the graph. The simplest network models assume the edges to be independent, while a statistically more interesting class of models treats the dyadic structures for pairs of nodes to be independent.

In an extensive review of the statistical literature on network modeling, Goldenberg et al. (2010) note:

> Almost all of the "statistically" oriented literature on the analysis of networks derives from a handful of seminal papers. In social psychology and sociology there is the early work of Simmel (1950) at the turn of the last century and Moreno (1934) in the 1930s, as well as the empirical studies of Milgram (1967) and Travers and Milgram (1969) in the 1960s; in mathematics/probability there is the Erdös–Rényi work on random graph models [Erdös and Rényi (1959, 1960), and a closely related *Annals of Mathematical Statistics* paper by Gilbert (1959)]. There are of course other papers that dealt with these topics contemporaneously or even earlier. But these are the ones that appear to have had lasting impact.

Received March 2010.