Assessing Degeneracy in Statistical Models of Social Networks [1]

Mark S. Handcock
University of Washington, Seattle

Working Paper no. 39
Center for Statistics and the Social Sciences
University of Washington

December 31, 2003

[1]Mark S. Handcock is Professor of Statistics and Sociology, Department of Statistics, University of Washington, Box 354322, Seattle WA 98195-4322. E-mail: handcock@stat.washington.edu; Web: www.stat.washington.edu/handcock. This research supported by Grant DA012831 from NICHD and Grant HD041877 from NIDA. The author is grateful to Carter Butts, Martina Morris, Pip Pattison, Garry Robins, Tom Snijders, Jim Moody, Julian Besag and Steve Goodreau for useful comments and discussions.

Abstract

This paper presents recent advances in the statistical modeling of random graphs that have an impact on the empirical study of social networks. Statistical exponential family models (Wasserman and Pattison 1996) are a generalization of the Markov random graph models introduced by Frank and Strauss (1986), which in turn are derived from developments in spatial statistics (Besag 1974). These models recognize the complex dependencies within relational data structures. A major barrier to the application of random graph models to social networks has been the lack of a sound statistical theory to evaluate model fit. This problem has at least three aspects: the specification of realistic models, the algorithmic difficulties of the inferential methods, and the assessment of the degree to which the graph structure produced by the models matches that of the data. We discuss these and related issues of the model degeneracy and inferential degeneracy for commonly used estimators.

KEY WORDS: Random graph models; log-linear network model; Markov fields; Markov Chain Monte Carlo; Statistical Exponential Families; Psuedolikelihood.

1. INTRODUCTION

Networks are a useful device to represent "relational data", that is, data with properties beyond the attributes of the individuals (nodes) involved. Relational data arise in many social science fields and graphical models are a natural approach to representing the regular pattern of the relations between nodes. In typical applications, the nodes in a graph represent individuals, and the ties (edges) represent a specified relationship between individuals. Nodes can also be used to represent larger social units (groups, families, organizations), objects (airports, servers, locations), or abstract entities (concepts, texts, tasks, random variables). This framework has many applications, including the structure of social networks, the dynamics of epidemics, the interconnectedness of the WWW, and long-distance telephone calling patterns.

etc.