Research in Social Stratification and Mobility 41 (2015) 56–65

Contents lists available at ScienceDirect

Research in Social Stratification and Mobility

journal homepage: http://www.elsevier.com/locate/rssm

Communities of classes: A network approach to social mobility

David Melamed *

The Ohio State University, Department of Sociology, Columbus, OH, 43210, USA

ARTICLE INFO

Article history:
Received 19 December 2014
Received in revised form 28 April 2015
Accepted 8 May 2015
Available online 16 May 2015

Keywords:
Communities
Modularity
Network
Social mobility
Stratification

ABSTRACT

Based on recent insights in network analysis, a new approach to the analysis and interpretation of social mobility data is presented. The approach advocates using community detection methods to identify communities of classes within which classes share members at above expected rates and between which classes share members at below expected rates. This approach, when applied to mobility data, offers novel interpretations of mobility patterns and may be used to substantially improve the fit of models of social mobility. To illustrate, the community structure of social mobility is analyzed using data from the General Social Survey. Several models are employed to demonstrate both the interpretation of the community structure of social mobility as well as how the community structure may be implemented to improve model fit.

1. Introduction

Analysts of mobility processes – be it intergenerational social mobility, religious mobility, or educational mobility, etc. – have a variety of log-linear and log-multiplicative methods at their disposal with which to analyze the structures and patterns embedded within mobility tables (e.g., Hout, 1983). Often, the structure in mobility tables is sufficiently complicated that parsimonious models do not capture the observed patterns. That is, simple models that account for the marginals, symmetrical movement in the structure, and/or inheritance effects may not be able to fit the observed data. In such circumstances, there are ever more complicated models that may be fit to the data. For example, one may add dimensions to Goodman's (1985) RCII model until a model fits the data. Likewise, if a preferred model (e.g., quasi-symmetry) does not fit the data, one can estimate a correspondence analysis on the residuals to "see" the associations left over in the data (Van der Heijden, de Falguerolles, & de Leeuw, 1989). In both of these cases, however, the results are particularly complicated and an understanding of the underlying mobility processes may be obscured by the complexity of the model.

In this paper, an alternative method is introduced for understanding and fitting mobility processes. Specifically, when a model of mobility processes (e.g., quasi-independence, symmetry, etc.) does not fit the data, the residuals from that model can be analyzed using community detection methods (e.g., Newman, 2010). By thinking about class categories (or any categories in a mobility table) as nodes, and people as the weighted relations between them, novel insights are drawn that illuminate the structure of social mobility tables. The results from the community detection analysis parsimoniously inform the analyst of associations that remain after the log-linear model has been estimated. Specifically, a community detection analysis identifies which categories share members at rates above chance (and hence belong together). Subsequently, community membership may be

* Corresponding author. Tel.: +1 4407596943.
E-mail address: melamed.9@osu.edu

http://dx.doi.org/10.1016/j.rssm.2015.05.001

included in the log-linear model to improve fit (provided the original model did not fit the data).

The approach described above offers a couple of advantages relative to other similar methods. Van der Heijden and colleague's (1989) correspondence analysis of the residuals from an ill-fitting log-linear model, like all instances of correspondence analysis, suffers from ambiguity of the results. While correspondence analysis may paint a pretty picture of the results (provided two dimensions are sufficient), it is unclear how to proceed with the analysis aside from just describing the associations. Likewise, the interpretation of results from Goodman's (1985) RCII model is particularly complicated. The intrinsic association parameter, and the row and column scores each need to be interpreted, for each dimension that is fit to the data. The strength of the community structure analysis that is detailed below is that an objective function is maximized to identify the "best" way to combine categories, and subsequently a single within-community term may be added to any log-linear model. The results are therefore clear, unambiguous, and relatively straightforward.

Below I present this new approach to interpreting social mobility that draws from recent advances in identifying communities in social networks. Aside from aiding in interpretation, the community procedure may also be leveraged to substantially improve model fit. Then the approach is applied to social mobility tables that were derived from the General Social Survey (GSS; Smith, Marsden, Hout, & Kim, 2005). The community structure of multiple models of social mobility is identified for female, male, and all respondents, the results of which reveal interesting substantive findings. Last, the community structure is then leveraged to improve model fit, resulting in at least one model fitting the data for female, male, and all respondents. The paper concludes with limitations and a general discussion.

2. Communities of social classes

Within network science, a mode refers to a set of objects for which relations may be measured. For example, a person-to-person network is a one-mode network, while a person-to-groups network is a two-mode network (an affiliation network). An intergenerational mobility table may be productively understood to be a two-mode network, where the first mode is parental social class and the second mode is respondent social class. What are shared between these modes are members, or the count of people who have a given social class and have parents of a given social class. Though it is not necessary, it is often assumed that the same social classes are represented for both parents and respondents, which implies that both modes share the same number of nodes, or in this case, categories.

One common analytic approach for the analysis of social networks is the identification of communities or cohesive subgroups (Newman, 2010). Here, a community refers to a subset of nodes (categories) which share relations (people) at above expected rates. Community detection has a rich history in computer science and the social sciences (e.g., Fielder, 1973; Jackson, 2008; Wasserman & Faust, 1994), with several early sociological methods developed to detect subsets of similarly embedded actors (e.g., Breiger, Boorman, & Arabie, 1975; Burt, 1978; Lorrain and White, 1971), but this literature has seen an explosion of development since Girvan and Newman (2002) brought this problem to the attention of the general scientific community (Porter, Onnela, & Mucha, 2009).

In this paper, I focus on Newman's (2006a,b) eigenspectrum decomposition approach because it is easily applicable to mobility tables, has been generalized to multimode networks (Melamed, Breiger, and West, 2013), such as mobility tables, and is highly efficient and accurate relative to other solutions to the community finding problem.[1] Newman's (2006a) eigenspectrum approach is elegantly simple. One begins with a relational matrix that is denoted by **A**, defines a matrix of expected cell counts that is denoted by **P** (which is typically an "independence" model, though see below), subtracts **P** from **A** to yield **B**, which is called the modularity matrix, and finally one computes the eigenspectrum decomposition of the modularity matrix. Thus, **B** is a matrix of residuals between the observed data and the expected frequencies under some model. The eigenspectrum of the residual matrix, **B**, sheds light on the community structure of **A** (Newman, 2006a,b). Newman has shown that the signs of the entries in the eigenvector associated with the largest eigenvalue partition the nodes into an optimal two community split. Subsequent splits into more than two communities may be determined by examining the signs of the entries in the second leading eigenvector, and so on (Newman, 2006b: 9–10).[2]

Another development with respect to community structures made by Newman and Girvan (2004) was to define the quality function that is unfortunately also called modularity. Modularity (denoted by Q) is a quality function (i.e., goodness-of-fit) that indicates the strength of, or variance explained by, a community structure discovered by the community finding algorithms. That is, it provides a benchmark with which to compare possible solutions to the community structure. Larger values of Q indicate that larger shares of the relations in the data are within community ties; hence larger values indicate a better fit to the data. The specific formula for modularity developed by Newman and Girvan was generalized for the eigenspectrum approach by Newman (2006b). Thus, the eigenspectrum decomposition of the modularity matrix can be used to identify possible solutions to the community structure, and the quality function modularity can be used to identify which solution is "best."

The general formula for the modularity function is the most intuitive. Define a number-of-communities by number-of-communities matrix, which is denoted **e**. The i,jth entry in **e** is the proportion of ties in the network that go from the community in row i to the community

[1] The global optimum of a community solution is known to be NP-hard. Hence one can never know if a community solution is the optimal solution. The eigenspectrum approach is both efficient and accurate. For reviews of other approaches to identify communities, see Fortunato (2010), Newman (2010), or Porter, Onnela, and Mucha (2009).

[2] It should be noted that this strategy is typically implemented on binary networks, but there is no reason for this restriction. Newman (2004) has pointed out that the eigenspectrum decomposition easily generalizes to weighted graphs.