ARTICLE IN PRESS

European Management Journal xxx (2016) 1–10

Contents lists available at ScienceDirect

European Management Journal

journal homepage: www.elsevier.com/locate/emj

Exponential random graph models for management research: A case study of executive recruitment

Mary-Paz Arrieta-Paredes*, Bruce Cronin

Centre for Business Network Analysis, University of Greenwich Business School, Park Row, Greenwich, London SE10 9LS, United Kingdom

ARTICLE INFO

Article history:
Received 9 August 2015
Received in revised form
13 June 2016
Accepted 15 August 2016
Available online xxx

JEL classification:
G32
J24
L25

Keywords:
Executive recruitment
Industry origin
Managerial performance
Shareholder value
Social network analysis
Exponential random graph models

ABSTRACT

We introduce a recent development in the statistical analysis of relational data that offers rigorous discrimination of a variety of structural and behavioural effects of interest to management research. Exponential random graph models account for the highly interdependent nature of network data that are problematic for the predominant inferential statistical analysis used in management research. We illustrate the value of the approach with an application focused on executive recruitment by large UK firms, modelling migrations of managers among firms as a network of relationships. We find rigorous statistical support for the influences of industry origin in executive recruitment, particularly in relation to legal and accounting activities. The flexibility and sophisticated relational variables available in the models offer considerable analytical power of value to a wide range of management applications.

1. Introduction

The network paradigm has become increasingly prominent in contemporary management research (Carpenter, Li, & Jiang, 2012; Kilduff & Brass, 2010). Organizations and their participants are now frequently conceptualized as socially embedded, with organizational activity seen to have extra-organizational influences and impacts, often via persistent repeated social interactions or networks. However, as the interdependence among and between actors and organizational settings has become more recognized, questions of causality have become more complex. To what extent do people shape networks or are people shaped by them? If there is co-evolution of people and networks, how does this unfold (Tasselli, Kilduff, & Menges, 2015)?

Management research has given little attention to these questions when considering social embeddedness, instead normally simply testing the effect of the presence of network ties via a count of the number of ties (e.g., Li, Popo, & Zhou, 2008). Alternatively, structural approaches give attention to the characteristics of the network in which the actor or organization is embedded and their position within it, providing opportunities and constraints that may determine organizational activity or channel effects of the activity outward, such as the diffusion of innovation. Social network analytic techniques are typically employed to derive a metric indicator of network position for use as an independent variable; many studies, for example, have found actor or organizational performance in various forms associated with the network metric betweenness centrality (e.g., Burt, 1992; Flynn & Wiltermuth, 2010).

The use of indicators of network position as independent variables in standard inferential statistics is problematic, however, because of the highly interdependent nature of relational data; by definition, such data are not independent nor are the distributions of network ties generally known. Standard inferential statistics,

Abbreviations: ERGM, exponential random graph model; GoF, goodness of fit; NACE, statistical classification of economic activities in the European Community (Nomenclature statistique des activités économiques dans la Communauté européenne).

* Corresponding author.
E-mail addresses: M.ArrietaParedes@greenwich.ac.uk (M.-P. Arrieta-Paredes), b.cronin@greenwich.ac.uk (B. Cronin).

http://dx.doi.org/10.1016/j.emj.2016.08.002

widely used in management research, infers properties of a population from those of a sample on the assumption that both the population and the sample meet a known distribution and each variable has independent effects on the variable of interest. Thus, network data require distinctive analytical treatment both in terms of statistical analysis and in distinguishing structural (network-originating) and selection (actor-originating) effects.

In this paper, we report a recently developed method that appears well positioned to meet the challenges in analyzing relational data in management research. Exponential random graph models (ERGMs) comprise an increasingly popular method in social network analysis to more rigorously model associations among interdependent data than conventional approaches. ERGMs provide an alternative approach to statistical analysis that can accommodate interdependent data and does not require a priori knowledge or assumptions about sample and population distributions. The models also provide for explicit testing of distinct structural and selection effects.

Following an introduction and description of ERGMs, we provide a case study applying the method to a problem in managerial research to model determinants of the movements of executives from one company to another. This problem, traditionally addressed as executive recruitment by an autonomous firm, is modelled as a network of movements of executives among firms — recruitment simultaneously filling a vacancy but creating another elsewhere.

2. The ERGM framework

A characteristic of the highly interdependent data found in networks is that small changes in connections can have large effects on network structure and the distribution of ties varies greatly from network to network. Outside some attempts to model 'scale-free' distributions approached in very large networks (Barabási, 2000), no a priori assumptions can be made about the distribution of ties in a network. Thus, unlike standard inferential statistics, the characteristics of a network as a whole or the significance of relationships amongst variables cannot be inferred by comparing patterns in the observed data with those in a standard distribution.

In addition, network data, by definition, is highly interdependent and these dependencies are commonly maintained or reinforced by social processes such as reciprocation. It is difficult to isolate individual actor behaviour from the influences of those an actor is connected to. Thus, the observed variables typically lack independence, violating another assumption of standard inferential statistics. Analysis of network data by conventional inferential methods thus tends to generate spurious relationships among variables, resulting in type I errors (Borgatti, Everett, & Johnson, 2013; Snijders, van de Bunt, & Steglich, 2010).

ERGMs employ permutation-based approaches to overcome these limitations of conventional statistical analysis. Rather than assuming independence and a particular distribution of observations, each observed variable is compared with a large number of random permutations of values to determine the extent to which the observation persistently differs from randomness, providing a rigorous measure of statistical significance (Dekker, Krackhardt, & Snijders, 2007). The approach allows the combination of structural variables (network statistics such as directed ties, mutual ties, transitivity paths and particular microstructures) and actor variables (characteristics of individual actors such as industry, value added or revenue of firm and dyadic selection processes such as reciprocity) into a single model, thus allowing the discrimination of relative structural and selection effects (Robins, Pattison, Kalish, & Lusher, 2007).

Fig. 1 presents some examples of the structural configurations and selection processes that can be included in an ERGM alongside conventional actor variables, such as revenue. The first row provides examples of structural configurations. Inclusion of a) as a variable allows us to test the extent to which the configuration of the observed network is likely to have arisen from distribution of arcs in the network. The count of outward ties, 'outdegree' or 'sources'; inward ties, 'indegree' or 'sinks'; and isolates are important characteristics of a network structure. In a similar manner, we can look for other particular microstructures such as b), c) or d) and test the extent to which each of these is likely to have contributed to the observed network configuration.

The second row of Fig. 1 provides examples of dyadic social selection variables, where network ties arise from the characteristics of the actors. In e) the originator or 'sender' of the relationship possesses an attribute while the 'receiver' does not. With continuous variables, this can be interpreted as representing greater amounts of the attribute. In f), the receiver possesses an attribute that the sender does not, a social interaction described as 'popularity'. In g), the possessor of an attribute forms a relationship with another possessing the same attribute, 'homophily.' In h), a popular relationship is reciprocated.

The third row of Fig. 1 provides some extensions where attributes are continuous variables, which can nuance the selection effects just described. In i), the absolute difference between the sender and receiver is used as a parameter, indicating heterophily. In j), nodes with higher total values in an attribute tend to tie together. In k), nodes with a higher product of values tend to tie together. In l), nodes with a higher attribute tend to both receive ties and send ties. Attribute-based 'instar' and 'outstar' relationships are also readily modelled.

In this manner, quite elaborate configurations of social interaction can be constructed and tested. Elaborations include '4-step cycles', 'social circuits', 'attribute-based centralisation', 'degree assortivity' and cross-level effects. Wang, Robins, and Pattison (2009), in their PNet implementation of ERGMs, provide a comprehensive categorization of structural, dyadic and selection variables, with Wang, Robins, Pattison, and Lazega (2016) discussing extensions for modelling multilevel networks. Other implementations include the Statnet package for R (Handcock, Hunter, Butts, Goodreau, & Morris, 2003) and the PyMC library for Python (Fonnesbeck, Patil, Huard, & Salvatier, 2015).

ERGMs can be more formally considered as allowing the representation of a network as a graph G, in terms of summary measures $z(G)$ the network statistics. In mathematical terms (Robins & Lusher, 2013b), an ERGM assigns probabilities to a given graph G with respect to these statistics, such that the weighted average of the Zs can be stated as

$$P_{\boldsymbol{\theta}}(G) = ce^{\theta_1 z_1(G)+\theta_2 z_2(G)+\ldots+\theta_p z_p(G)} \tag{1}$$

Expression (1) tells us that the probability of a graph G depends on the number of configurations (network statistics) or some Z functions of them, where θs are their parameters and c is a normalized constant (Robins & Lusher, 2013b). Since the inferential goal is to find data's maximal support under $z(G)$, the estimation of (1) implies solving for the moments equation of $\boldsymbol{\theta}$ via maximum likelihood estimation. Because of natural data dependencies, this is usually done numerically; in our case, it was done by employing a stochastic approximation technique (the Robbins-Monro algorithm) as explained in Koskinen and Snijders (2013), which is solved via convergence to stable values.

It is also important to clarify that the goodness-of fit (GoF) in ERGMs is called a heuristic GoF, which is a simulation of how central or extreme non-fitted effects are in the distribution of the ERGM in equation (1) compared with fitted effects, such that if the