Available online at www.sciencedirect.com
SCIENCE DIRECT®
Social Networks 28 (2006) 165–178
SOCIAL NETWORKS
www.elsevier.com/locate/socnet

Computing core/periphery structures and permutation tests for social relations data ☆

John P. Boyd *, William J. Fitzgerald, Robert J. Beck

University of California at Irvine, CA 92697, USA

Abstract

The core/periphery structure is ubiquitous in network studies. The discrete version of the concept is that individuals in a group belong to either the core, which has a high density of ties, or to the periphery, which has a low density of ties. The density of ties between the core and the periphery may be either high or low. If the core/periphery structure is given a priori, then there is no problem in finding a suitable statistical test. Often, however, the structure is not given, which presents us with two problems, searching for the optimal core/periphery structure, and devising a valid statistical test to replace the one invalidated by the search. *UCINET* [Borgatti, S.P., Everett, M.G., Freeman, L.C., 2002. *UCINET* for Windows, Version 6.59: Software for Social Network Analysis. Analytic Technologies, Harvard], the oldest and most trusted network program, gives incorrect answers in some simple cases for the first problem and does not address the second. This paper solves both problems with an adaptation of the Kernighan–Lin search algorithm, and with a permutation test incorporating this algorithm.

Keywords: Core; Periphery; Permutation test; Algorithm

1. Introduction

The idea that some groups or organizations have core/periphery structures has enjoyed considerable historical mention in social network analysis (e.g., Laumann and Pappi, 1976;

☆ This material is based on work supported under the United States Department of Education Grant # PT3 447625-23470-03.
* Corresponding author. Tel.: +1 949 824 5427; fax: +1 949 824 8651.
E-mail address: jpboyd@uci.edu (J.P. Boyd).

0378-8733/$ – see front matter © 2005 Elsevier B.V. All rights reserved.
doi:10.1016/j.socnet.2005.06.003

Alba and Moore, 1978; Mintz and Schwartz, 1981). This notion continues to be employed in studies including scientific citation networks (Doreian, 1985), world economy (Smith and White, 1992), corporate structure (Barsky, 1999), and small groups (Beck et al., 2003, submitted for publication; Cummings and Cross, 2003).

Borgatti and Everett (1999) presented several formal models for core/periphery structures, which were incorporated into the most widely used network analysis program, *UCINET* (Borgatti et al., 2002), for general application. Given a square data matrix, *UCINET* finds a core/periphery structure in two possible ways: either it computes the degree of "coreness" for each node or actor in a matrix (continuous model), or it finds a labeled bipartition of these nodes or actors, core and periphery subgroups (discrete model). We shall restrict our attention in this paper to the discrete model, revisiting the continuous model in another paper.

The output for each model in *UCINET* also includes an overall measure of "fitness" that indicates how well the observed data approximates an ideal core/periphery structure. A high fitness measure implies a good agreement with the model, while a lower fitness measure suggests that the model should be rejected. However, since there is no test for the statistical significance of fitness, *UCINET* users are left without the benefit or comfort of a *p*-value.

In other words, as Borgatti and Everett (1999) point out, this overall fitness value does not tell us the likelihood of obtaining, by chance alone, a value as high as the one actually observed. This is a problem when the core/periphery bipartition is not given a priori, but is instead selected to be optimal, or even close to optimal. This is similar to the statistical problem of determining the difference between two groups that is the output from a clustering program—obviously, running a simple *t*-test would be wrong. Although a solid statistical underpinning for these optimal core/periphery models would significantly strengthen both their usefulness and substantive interpretation, Borgatti and Everett (1999) did not specify a statistical test of significance. They maintain that this is because it is necessary to first have a theory about how ties are formed, and then to generate a null model specific to the substantive context and to the type of data at hand. Therefore, because each unique dataset demands its own null model, a universal permutation test would not be appropriate (1999:393–394).

From a pragmatic standpoint, this situation leaves researchers somewhat frustrated. The concept of a core/periphery structure is one of a number of viable a priori models, which if consistent with data, may have important substantive consequences and theoretical implications. Thus, we would like a more decisive and efficient way to assess the basic applicability of this model than is currently available. At this initial stage, it would be desirable to have some kind of generic permutation test (Good, 1994) that could be employed in most, if not all, situations to obtain the significance of an observed fitness value. With this solution, we would readily be able to tell whether we should move on to other models, or if it might be more fruitful to develop more specific and more powerful statistical tests based on models of tie formation as called for by Borgatti and Everett (1999). In other words, it is better to have a statistical test for even a simple null hypothesis, even though it may not be as complex as one would like. We will point out possible directions for more complex models in the discussion section.

In this paper, we propose and develop such a universal permutation test for the discrete core/periphery model. We believe that this test would be useful in most situations where