SIAM J. APPL. MATH.
Vol. 74, No. 1, pp. 167–190
© 2014 Society for Industrial and Applied Mathematics

Downloaded 03/30/15 to 134.102.186.160. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php

CORE-PERIPHERY STRUCTURE IN NETWORKS*

M. PUCK ROMBACH†, MASON A. PORTER‡, JAMES H. FOWLER§, AND PETER J. MUCHA¶

Abstract. Intermediate-scale (or "meso-scale") structures in networks have received considerable attention, as the algorithmic detection of such structures makes it possible to discover network features that are not apparent either at the local scale of nodes and edges or at the global scale of summary statistics. Numerous types of meso-scale structures can occur in networks, but investigations of such features have focused predominantly on the identification and study of community structure. In this paper, we develop a new method to investigate the meso-scale feature known as *core-periphery structure*, which entails identifying densely connected core nodes and sparsely connected peripheral nodes. In contrast to communities, the nodes in a core are also reasonably well connected to those in a network's periphery. Our new method of computing core-periphery structure can identify multiple cores in a network and takes into account different possible core structures. We illustrate the differences between our method and several existing methods for identifying which nodes belong to a core, and we use our technique to examine core-periphery structure in examples of friendship, collaboration, transportation, and voting networks.

Key words. core-periphery structure, networks, meso-scale structure

AMS subject classifications. 62H30, 91C20, 91D30, 94C15

DOI. 10.1137/120881683

1. Introduction.

Networks are used to model systems in which entities, represented by nodes, interact with each other. When representing a network as a graph, all of the connections are pairwise and hence represented by ties known as edges [5, 43]. Such a representation has led to numerous insights in the social, natural, and information sciences, and the study of networks has in turn borrowed ideas from all of these areas [3].

Networks can be described using a mixture of local, global, and intermediate-scale (meso-scale) perspectives. Accordingly, one of the key uses of network theory is the identification of summary statistics for large networks in order to develop a framework for analyzing and comparing complex structures [43]. In such efforts, the algorithmic identification of meso-scale network structures makes it possible to discover features that might not be apparent either at the local level of nodes and edges or at the global level of summary statistics.

In particular, considerable effort has gone into algorithmic identification and investigation of a particular type of meso-scale structure known as *community struc-*

*Received by the editors June 19, 2012; accepted for publication (in revised form) July 31, 2013; published electronically February 18, 2014. This work was funded by the James S. McDonnell Foundation (220020177) and the NSF (DMS-0645369) and was carried out in part at the Statistical and Applied Mathematical Sciences Institute in Research Triangle Park, North Carolina.
http://www.siam.org/journals/siap/74-1/88168.html

†Oxford Centre for Industrial and Applied Mathematics, Mathematical Institute, University of Oxford, Oxford OX2 6GG, UK. Current address: Department of Mathematics, UCLA, Los Angeles, CA 90095-1555 (rombach@ucla.math.edu).

‡Oxford Centre for Industrial and Applied Mathematics, Mathematical Institute, University of Oxford, Oxford OX2 6GG, UK; and CABDyN Complexity Centre, University of Oxford, Oxford OX1 1HP, UK (porterm@maths.ox.ac.uk).

§Department of Political Science and School of Medicine, University of California, San Diego, CA 92093 (fowler@ucsd.edu).

¶Department of Mathematics and Department of Applied Physical Sciences, University of North Carolina, Chapel Hill, NC 27599 (mucha@unc.edu).

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

Downloaded 03/30/15 to 134.102.186.160. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php

ture [21, 51], in which cohesive groups called "communities" consist of nodes that are connected densely to each other and the connections between nodes in different communities are comparatively sparse. Myriad methods have been developed to detect network communities [21, 26, 44, 51], and several of these methods allow communities to overlap with each other [1, 2, 47]. Investigations of community structure have led to insights in applications such as committee [50] and voting [40] networks in political science, friendship networks at universities [61] and other schools [29], protein-protein interaction networks [38], and mobile telephone networks [45].

Although (and arguably because) studies of community structure have been very successful [21, 51], other types of meso-scale structures—often in the form of different "block models" [18, 21]—have received much less attention than they deserve. The type of meso-scale network structure that we consider in the present paper is known as *core-periphery structure* [14]. The qualitative notion that social networks can have such a structure makes intuitive sense and has a long history in subjects like sociology [17, 36], international relations [10, 57, 59, 63], and economics [35]. Snyder and Kick [58] examined networks of international relations—which were based on, e.g., trade flow and diplomatic relations—and they provided evidence for a block-model core-periphery structure in their data (as was expected theoretically). Nemeth and Smith [41] and Smith and White [57] noted a core-periphery structure in international trade. Core-periphery structures have also been examined in national trade [54, 67], academic journals [17, 68], human social networks [31, 66], social networks in monkeys [13, 56], and networks of interactions between scientists [9, 24].

The most popular quantitative method for investigating core-periphery structure was proposed by Borgatti and Everett in the late 1990s [6]. Since then, various notions of core-periphery structure have been developed [14, 15, 16, 32, 55, 69], but most examinations of core-periphery structure still rely on implementations of the methods in [6] or [12] in the software package UCInet [7].

By computing a network's core-periphery structure, one attempts to determine which nodes are part of a densely connected core and which are part of a sparsely connected periphery. Core nodes should also be reasonably well connected to peripheral nodes, but the latter are not well connected to a core or to each other. Hence, a node belongs to a core if and only if it is well connected both to other core nodes and to peripheral nodes. A core structure in a network is thus not merely densely connected but also tends to be "central" to the network (e.g., in terms of short paths through the network). The goal of quantifying various notions of "centrality," which are intended to measure the importance of a node or other network component [43, 64], also helps to distinguish core-periphery structure from community structure. Additionally, networks can have nested core-periphery structure as well as both core-periphery structure and community structure [37, 69], so it is desirable to develop algorithms that allow one to simultaneously examine both types of meso-scale structures.

In Figure 1.1, we show images of the adjacency matrices of idealized block models that illustrate (a) community structure, (b) core-periphery structure, (c) a global core-periphery structure with a local community structure, and (d) a global community structure with a local core-periphery structure. By permuting rows and columns of the adjacency matrix, one can see that (c) and (d) are equivalent.

Several results underscore the importance of considering core-periphery structure in addition to community structure. For example, Chung and Lu [11] showed that power-law random graphs, in which the number of nodes of degree k is proportional to $k^{-\beta}$, almost surely contain a dense subgraph that has a short distance to almost all other nodes when the exponent $\beta \in (2, 3)$. This suggests that it is sensible for

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

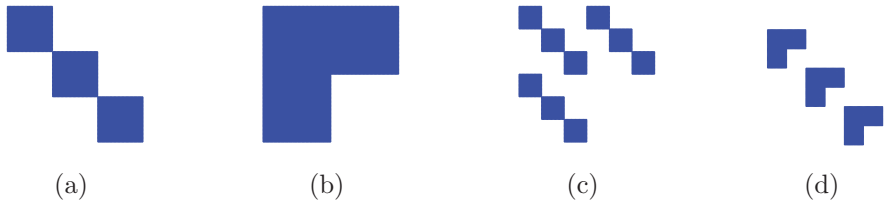


FIG. 1.1. *Examples of network block models.* (a) *Community structure,* (b) *core-periphery structure,* (c) *global core-periphery structure with local community structure, and* (d) *global community structure with local core-periphery structure. Note that* (c) *and* (d) *are equivalent.*

networks with heavy-tailed degree distributions to contain some sort of cohesive core, and there is strong evidence that this is indeed the case in many real-world networks (such as many social networks and the World Wide Web) [5, 43, 64]. Moreover, core-periphery structure and community structure provide complementary lenses through which to view meso-scale network structures [69].

Nodes of particularly high degree (which are sometimes called "hubs") occur in many real-world networks and can pose a problem for community detection, as they often are connected to nodes in many parts of a network and can thus have strong ties to several different communities. For instance, such nodes might be assigned to different communities when applying different computational heuristics using the same notion of community structure [30], and it becomes crucial to consider their strengths of membership across different communities (e.g., by using a method that allows overlapping communities) [1, 2]. In such situations, the usual notion of a community might not be ideal for achieving an optimal understanding of the meso-scale network structure that is actually present, and considering hubs to be part of a core in a core-periphery structure might be more appropriate [37]. For example, one can consider communities as tiles that overlap to produce a network's core [69].

The rest of this paper is organized as follows. We first describe several previously proposed methods for detecting core-periphery structure in networks before presenting our new method, which computes a continuous value along a core-periphery spectrum and thereby yields a centrality measure based on core-periphery structure. We illustrate our method using a set of synthetic (computer-generated) benchmark random networks with a planted core. We then apply our method to several real networks: the Zachary Karate Club, co-authorship networks of network scientists, a voting-similarity network of United States Senators, and the London Underground ("The Tube") transportation network. We conclude by summarizing our results, and we then present additional results and discussion in the appendix.

2. Detecting core-periphery structure.

2.1. Existing methods. Intuitively, one expects many real networks to possess some sort of core-periphery structure as part of their meso-scale structure. Perspectives proposed to examine core-periphery structure include block models [6], k-core organization [32], consideration of connectivity information and short paths through a network [15, 16, 55], and overlapping of communities [69].

The most popular notion of core-periphery structure in networks was developed by Borgatti and Everett [6], who proposed algorithms for detecting both discrete and continuous versions of core-periphery structure in weighted, undirected graphs. Their discrete notion of core-periphery structure is based on comparing a network to

Downloaded 03/30/15 to 134.102.186.160. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

Downloaded 03/30/15 to 134.102.186.160. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php

a block model that consists of a fully connected core and a periphery that has no internal edges but is fully connected to the core. Their method aims to find a vector C of length N whose entries can be either 1 or 0. The ith entry C_i equals 1 if the corresponding node is assigned to the core, and it equals 0 if the corresponding node is assigned to the periphery. Let $C_{ij} = 1$ if $C_i = 1$ or $C_j = 1$, and let $C_{ij} = 0$ otherwise. Define

$$\rho_C = \sum_{i,j} A_{ij} C_{ij}, \tag{2.1}$$

where the adjacency-matrix element A_{ij} represents the weight of the tie between nodes i and j and equals 0 if nodes i and j are not connected. This method of computing a discrete core-periphery structure seeks a value of ρ_C that is high compared to the expected value of ρ_C if C is shuffled such that the number of 1 and 0 entries is preserved but their order is randomized. The output is the vector C that gives the highest z-score for ρ_C.

As a variant discrete notion of core-periphery structure, Borgatti and Everett defined [6]

$$C_{ij} = \begin{cases} 1 & \text{if } C_i = C_j = 1, \\ a \in [0,1] & \text{if } C_i = 1 \text{ xor } C_j = 1, \\ 0 & \text{otherwise,} \end{cases} \tag{2.2}$$

where "xor" denotes an "exclusive or" operation. Borgatti and Everett also defined a continuous notion of core-periphery structure in which a node is assigned a "coreness" value of C_i and $C_{ij} = C_i \times C_j = a$. Our method for studying core-periphery structure in weighted, undirected networks (see section 2.2) is motivated by this continuous formulation of Borgatti and Everett.

In UCInet [7], the suggested heuristic for computing continuous core-periphery scores is the MINRES method [8, 12]. MINRES seeks a vector C such that the adjacency matrix is approximated by CC^T. The approximation minimizes the off-diagonal sums of squared differences. It thus seeks to find a vector C that minimizes $\sum_i \sum_{j \neq i} [A_{ij} - C_i C_j]^2$. Taking a partial derivative with respect to each element of C gives

$$C_i = \frac{\sum_{j \neq i} A_{ij} C_j}{\sum_{j \neq i} C_j^2}, \tag{2.3}$$

which in turn yields an iterative process for computing the MINRES vector. In many cases, this vector will be similar to the leading eigenvector of the adjacency matrix.

Holme defined a core-periphery coefficient [32]

$$c_{cp}(G) = \frac{C_C\left(V_{\text{core}}(G)\right)}{C_C\left(V(G)\right)} - \left\langle \frac{C_C\left(V_{\text{core}}(G')\right)}{C_C\left(V(G')\right)} \right\rangle_{G' \in \mathcal{G}(G)}, \tag{2.4}$$

where V is the set of nodes of an unweighted and undirected graph G, the angled brackets indicate averaging, and $\mathcal{G}(G)$ is an ensemble of graphs with the same degree sequence as G. Additionally,

$$C_C(U) = \left(\left\langle \langle P(i,j) \rangle_{j \in V \setminus \{i\}} \right\rangle_{i \in U} \right)^{-1}, \tag{2.5}$$

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

Downloaded 03/30/15 to 134.102.186.160. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php

and $P(i,j)$ is the distance (i.e., number of edges in the shortest path) between nodes i and j. A k-core of the graph G is a maximal connected subgraph in which all nodes have degree at least k, and V_{core} is the k-core with maximal $C_C(U)$. Using k-cores to examine core-periphery structure is computationally fast (and we note that one could, in principle, generalize Holme's method for weighted graphs using some kind of weighted k-core [25]), but it entails extremely strong restrictions on the notion of a network core. Philosophically, we view it as analogous to requiring a network community to be a clique.

One expects a core of a network to have high connectivity to other parts of the network, so Da Silva, Ma, and Zeng [15] introduced a measure of connectivity known as network *capacity*:

$$K = \sum_{l=1}^{M} P_l^{-1}\,, \tag{2.6}$$

where M is the total number of connected pairs of nodes and P_l is the length of the shortest path between the lth pair of nodes. They then defined a core coefficient as $cc = N'/N$, where N is the total number of nodes in the network, N' satisfies $\sum_{m=0}^{N'} K_m = 0.9 \sum_{u=0}^{N} K_u$, and K_m is the capacity of the network after the removal of m nodes. (One could define a more general notion using a parameter instead of the specific value 0.9.) The nodes are removed in order of closeness centrality, which is defined as the mean shortest path from a node to each of the other nodes in a network [15]. Note that in the remainder of this paper, we will use the following definition for the closeness centrality of a node j (there are several different definitions available in the literature [43]):

$$CC_j = \frac{1}{N} \sum_{i \in V} P(i,j)\,,$$

where V denotes the set of nodes and $P(i,j)$ is the sum of edge weights in a shortest path in the context of weighted networks. Da Silva, Ma, and Zeng considered only binary networks, but their method can be generalized straightforwardly to weighted networks.

Other recent ideas for examining core nodes in a network include the computation of "knotty centrality" [55] (which attempts to discover nodes that have high geodesic betweenness centrality but which need not have high degree), the identification of cores based on collections of nodes in overlapping communities [69], and the use of random walkers [16].

2.2. Our method. Our method for studying core-periphery structure in weighted, undirected networks is motivated by the continuous formulation of Borgatti and Everett [6] that we described above. However, our method takes cores of different sizes and shapes into account. It thereby gives credit to all nodes that take part in a core, and it weights this credit by the quality of the associated core. As we discuss below, we employ a *transition function* to interpolate between core and peripheral nodes. Additionally, we construct elements C_{ij} of a *core matrix* to compute the quality of a core. We will present several viable choices for both the transition function and the core matrix.

We define the *core quality*

$$R_\gamma = \sum_{i,j} A_{ij} C_{ij}\,, \tag{2.7}$$

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

Downloaded 03/30/15 to 134.102.186.160. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php

where γ is a vector that parametrizes the core quality (see the discussion below), the elements C_{ij} of the core matrix are given by $C_{ij} = f(C_i, C_j)$, and $C_i \geq 0$ is the *local core value* of the ith node. The local core values are elements of a *core vector* C. Our example calculations in this paper usually use a product form

$$C_{ij} = C_i C_j, \tag{2.8}$$

but we discuss other viable choices in section 2.2.1.

We seek a core vector C that maximizes R_γ and is a normalized (so that its entries sum to 1) shuffle of the vector C^* whose components $C_i^* = g(i)$ are determined using a *transition function* g. The number of components of the vector C^* is equal to the number of nodes in the network, and C_i^* gives the local core value of the ith node. Our example calculations in this paper usually use the transition function given by the sharp (because it has a discontinuous derivative) function

$$C_i^*(\alpha,\beta) = g_{\alpha,\beta}(i) = \begin{cases} \frac{i(1-\alpha)}{2\lfloor \beta N \rfloor}, & i \in \{1,\dots,\lfloor \beta N \rfloor\}, \\ \frac{(i-\lfloor \beta N \rfloor)(1-\alpha)}{2(N-\lfloor \beta N \rfloor)} + \frac{1+\alpha}{2}, & i \in \{\lfloor \beta N \rfloor + 1,\dots,n\}. \end{cases} \tag{2.9}$$

(When $\beta = 0$, only the top case in (2.9) applies; when $\beta = 1$, only the bottom case applies.) The parameter β sets the size of the core: as β varies from 0 to 1, the number of nodes included in the core varies from N to 0. The parameter α sets the size of the score jump between the highest-scoring peripheral node and the lowest-scoring core node. In the limit in which $\alpha = 1$, this yields a discrete classification (discontinuous function) into a unique core and unique periphery that assigns each node to either the core or the periphery.

With the transition function (2.9) and the product form (2.8) for the core-matrix elements, the core quality is given by

$$R_\gamma = R_{\alpha,\beta} = \sum_{i,j} A_{ij} C_{ij} = \sum_{i,j} A_{ij} C_i C_j. \tag{2.10}$$

For a given value of $\gamma = (\alpha, \beta)$, we seek a shuffle C of C^* such that R_γ is maximized.

For any choice of core matrix and transition function, we define the aggregate *core score* of each node i as

$$CS(i) = Z \sum_\gamma C_i(\gamma) \times R_\gamma, \tag{2.11}$$

where the normalization factor Z is chosen so that $\max_k[CS(k)] = 1$, where $k \in \{1,\dots,N\}$ indexes the nodes. A core score gives a notion of network centrality [43, 64]. As discussed above, our usual choice in this paper is to maximize the core quality (2.10) that uses the product form (2.8) for the core matrix and the sharp transition function (2.9) to interpolate between core and peripheral nodes. See section 2.2.1 for a discussion of other choices for constructing the core matrix and section 2.2.2 for other choices of transition function.

In the results that we present in this paper, we assign the values of $C_i^*(\alpha,\beta)$ to the nodes to obtain a value $C_i(\alpha,\beta)$ that maximizes $R_{\alpha,\beta}$ using a simulated-annealing algorithm [34]. (See the appendix for details of the procedure.) Other computational heuristics can, of course, be employed to increase the method's speed or for any other reason. In all of our examples using a two-parameter transition function, we sample α and β uniformly over a discretization of the square $[0,1] \times [0,1]$. In particular,

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

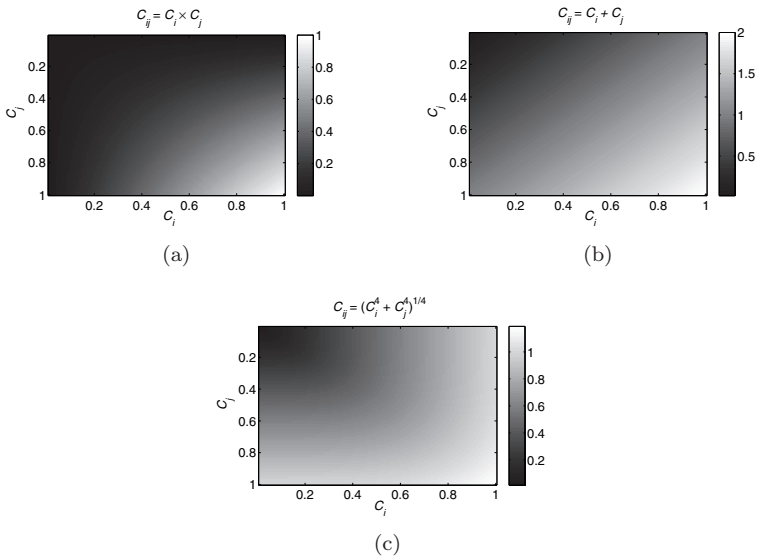


FIG. 2.1. *Several options for the core-matrix element C_{ij} include* (a) *the product form $C_{ij} = C_i \times C_j$,* (b) *the 1-norm $C_{ij} = \|(C_i, C_j)\|_1 = C_i + C_j$, and* (c) *the 4-norm $C_{ij} = \|(C_i, C_j)\|_4 = \sqrt[4]{C_i^4 + C_j^4}$.*

we always use $\alpha = \beta = [0.01{:}0.01{:}1]$ (in MATLAB notation). It is also interesting to consider the core quality of specific values of α and β, and one could in principle improve the speed of our general approach by developing procedures for choosing α and β selectively in a manner that takes advantage of the structure of particular networks or families of networks. Indeed, the a priori choice of which values of α and β to sample is a difficult but interesting question. The purpose of this paper is to introduce a novel notion of core-periphery structure and to demonstrate why it is interesting using a variety of examples, so we leave the aforementioned issues for future consideration.

2.2.1. Functional forms for elements of a core matrix. In most of the calculations in this paper, we construct the core-matrix elements C_{ij} using a product form $C_{ij} = C_i C_j$. However, other choices are also viable.

An idealized core-periphery structure entails that core nodes are well-connected to other core nodes as well as to peripheral nodes and that peripheral nodes are not well-connected to each other. Let v_1 and v_2 be core nodes, and let w_1 and w_2 be peripheral nodes. We then want $C_{w_1 w_2}$ to be small and $C_{v_1 v_2}$ and $C_{v_i w_j}$ (where $i, j \in \{1, 2\}$) to be large. For example, the block structure in panel (b) of Figure 1.1 satisfies these conditions.

As one can see from Figure 2.1, one can try to approximate such an idealized block structure using various ways of constructing C_{ij}. For example, in addition to the product form (2.8), one can instead use a p-norm and write

$$C_{ij} = \|(C_i, C_j)\|_p = \sqrt[p]{C_i^p + C_j^p}\,. \tag{2.12}$$

As one considers progressively larger p, this will look increasingly like an ideal core-periphery block model (in which core-core edges and core-periphery edges produce a

Downloaded 03/30/15 to 134.102.186.160. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

Downloaded 03/30/15 to 134.102.186.160. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php

value of 1 in a network adjacency matrix, but periphery-periphery edges produce a value of 0).

2.2.2. Transition function. Our methodology for computing core-periphery structure entails choosing a transition function to interpolate between core and peripheral nodes. In most of the calculations in this paper, we use the sharp two-parameter function (2.9) to illustrate our approach. However, there are many other viable choices for the transition function.

One variant is to construct the vector C^* using a smooth transition function $g(i)$. For example, one possibility is

$$C_i^*(\alpha,\beta) = g_{\alpha,\beta}(i) = \frac{1}{1+\exp\{-(i-N\beta)\times\tan(\pi\alpha/2)\}}\,, \tag{2.13}$$

which has parameters $\alpha \in [0,1]$ and $\beta \in [0,1]$. The parameter α sets the sharpness of the boundary between the core and the periphery. The value $\alpha = 0$ yields the fuzziest boundary, and $\alpha = 1$ gives the sharpest transition: as α varies from 0 to 1, the maximum slope of C^* varies from 0 to ∞. The parameter β sets the size of the core: as β varies from 0 to 1, the number of nodes included in the core varies from N to 0.

Another option, which allows our method to be significantly faster, is a transition function that has only one parameter. One can choose such a parameter to control the size of the core, the sharpness of the boundary, or some combination of the two. For example, one possibility is

$$C_i^*(\alpha) = g_\alpha(i) = \frac{1}{2}\tanh\left(8\exp\left\{\frac{-10*(\alpha-N/2)^2}{2}\right\}(i-\alpha)+1\right)\,. \tag{2.14}$$

We plot (2.14) for various values of α in Figure 2.2. One can then average $C_i^*(\alpha)$ over multiple values of α to produce aggregate core scores.

In this paper, we calculate aggregate core scores using formulations with both two-parameter and one-parameter transition functions. In the former case, we always average over 10000 values of (α,β) that are sampled uniformly from $[0,1]\times[0,1]$. (In particular, we use $\alpha = \beta =$ [0.01:0.01:1].) In the latter case, we always average over 10000 values of α that are uniformly sampled from $[0,1]$. (In particular, we use $\alpha =$ [0.0001:0.0001:1].)

2.2.3. Interpreting core scores. There are several ways to use and interpret the results of our approach for studying core-periphery structure. One can average over a set of parameter values—e.g., in the (α,β) parameter plane if one uses a two-parameter transition function—and obtain a set of aggregate core scores that yield a continuous centrality measure for the nodes in a network. Alternatively, one can determine a core-periphery structure at a single point in parameter space—such as the point that produces the largest value of the core quality R in (2.7). (See the discussion of the Zachary Karate Club network in section 4.1.) Sometimes, as with the London Underground network in section 4.2, one can observe a clear dichotomy between core and peripheral nodes after calculating continuous core scores. Finally, it can sometimes be useful to impose a specific core size in advance (and thereby dichotomize core and peripheral nodes), as we do with the synthetic benchmark networks in section 3.1.

The flexibility described in the above paragraph is a beneficial feature of our method, which can be used either to produce a continuum of core scores or a discrete

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

Downloaded 03/30/15 to 134.102.186.160. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php

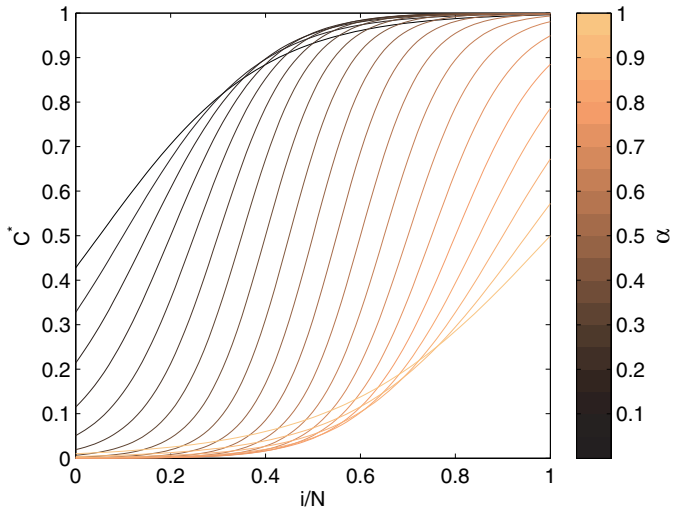


FIG. 2.2. *An example of a one-parameter transition function in which the parameter α controls both the size of a network core and the sharpness of the boundary between core and peripheral nodes.*

classification of core versus periphery. The utility of both of these perspectives, and hence the desirability of the development of methods for studying core-periphery structure that have such flexibility, was recognized more than two decades ago [6, 10, 57]. For example, studies of international relations have included vehement arguments as to whether countries should be classified discretely (e.g., into core, semiperipheral, and peripheral countries) or along a continuum [57], and methods that allow both discrete and continuous perspectives on core-periphery structure ought to be helpful for studying such applications.

3. Synthetic benchmark networks. In this section, we examine our method using an ensemble of random networks with an imposed core-periphery structure to demonstrate that it performs well at detecting the kind of core-periphery structure envisioned by Borgatti and Everett [6]. We then consider lattice networks, which do not have any meaningful core-periphery structure.

3.1. Imposed core-periphery structure. We develop a family of synthetic networks that have a planted core-periphery structure, and we use $CP(N, d, p, k)$ to denote this ensemble of networks. (We will consider networks with both core-periphery structure and community structure when we examine real networks. For example, see the London Underground network in section 4.2 and the network of network scientists in section 4.5.) Each network in the ensemble $CP(N, d, p, k)$ has N nodes, where dN of the nodes are core nodes, $(1 - d)N$ of the nodes are peripheral nodes, and $d \in [0, 1]$. The edges are assigned independently at random. The edge probabilities for periphery-periphery, core-periphery, and core-core pairs are kp, kp, and k^2p, respectively. Note that $p \in [0, 1]$ and $k \in [1, (1/p)^{1/2}]$. We fix $N = 100$, $d = 1/2$, and $p = 1/4$ and compute the core-periphery structure averaged over 100 different instances of $CP(N, d, p, k)$ for each of the parameter values $k = 1, 1.05, 1.1, \ldots, 2$. In Figure 3.1, we show our results from determining core nodes by computing the aggregate core score (2.11) with core quality (2.10) and transition function (2.9). The syn-

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

Downloaded 03/30/15 to 134.102.186.160. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php

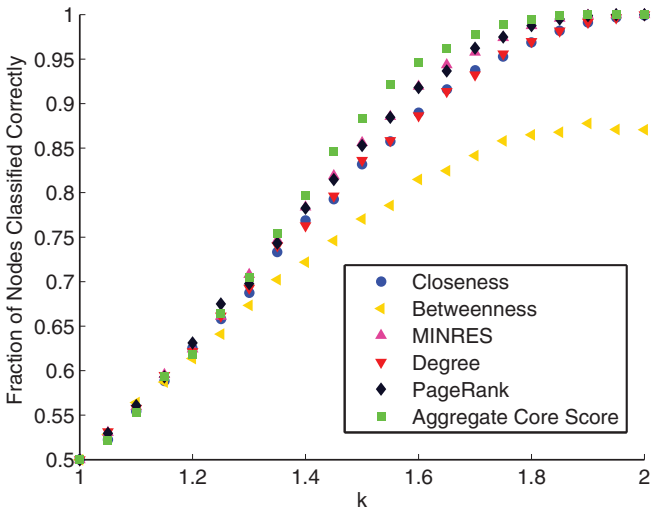


FIG. 3.1. *Fraction of core nodes correctly identified by computing aggregate core score averaged over* 100 *realizations of networks in the ensemble* $CP(100, .5, .25, k)$. *We compute the aggregate core score* (2.11) *using the core quality* (2.10) *and the transition function* (2.9).

thetic networks in $CP(N, d, p, k)$ possess a discrete core-periphery structure, whereas our method produces a continuous ranking, which we recall makes the aggregate core score a type of centrality.

We also examine the results of attempting to determine the core nodes using various types of centrality that we described in section 2.1—closeness, degree, PageRank [46], geodesic node betweenness [43], and MINRES [12]—which are designed to measure notions of node importance. We only test continuous node-ranking notions, which we evaluate by counting how many of the 50 core nodes—recall that the networks have $Nd = 50$ core nodes by construction—are placed in the top 50 according to each method. (Alternatively, one can use information-theoretic diagnostics to evaluate the results of comparisons like this one.)

In Figure 3.1, we show the fraction of nodes that are correctly identified as one of the top 50 core nodes. When testing the methods, we used a random permutation (using the MATLAB command `randperm`, which produces a uniform permutation) of the labels of the nodes to prevent any bias based on node order. In this case, none of the tested methods should suffer from such a bias. (Note that our method starts the optimization with a random permutation of the vector C^*. We again employed the command `randperm`.) We used our own implementation of MINRES for the calculation in this figure.

As we have indicated, our method examines core-periphery structure as a type of centrality. Nodes are more likely to be part of a network's core if (1) they have high strength (i.e., weighted degree) *and* (2) they are connected to other core nodes. Neither notion of importance is sufficient on its own. Nodes with high degree are construed as important in many situations, and the second idea is reminiscent of quantities like eigenvector centrality and PageRank centrality [46], which recursively define nodes as important based on having connections to other nodes that are important [43]. We will also compare core scores with notions of centrality when we discuss political voting-similarity networks in section 4.4.

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

Downloaded 03/30/15 to 134.102.186.160. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php

3.2. Lattices. As another example of a synthetic network, consider a lattice, which does not have any meaningful core-periphery structure. (A lattice also does not have any meaningful community structure.) All nodes in a lattice have the same degree if one uses periodic boundary conditions. Moreover, lattices are *symmetric*: for any two nodes, there exists a network automorphism that swaps the labeling of these two nodes. Thus, if one node is placed in the core and the other is placed in the periphery, then one could relabel the network in a way that would swap those assignments. Thus, for such networks, any assignment of core-periphery structure is arbitrary. The aggregate core score of *every* node in a lattice converges to the same value (which is equal to 1) as one applies our method with increasingly high precision (i.e., by using more values of (α, β)).

A possible concern about our methodology is that it might lead to false positives due to "forcing" different core-periphery structures on a network—especially given that we set the maximum aggregate core score to be 1, so every network will always have nodes with high aggregate core scores. However, as lattices illustrate, this does not necessarily lead to false positives. An aggregate core score is an average over many computational runs (using different values of α and β), and symmetry guarantees that each node has an equal probability of being assigned a high score in a given run. Therefore, by taking averages over many runs, we see that the aggregate core score of each node is similar, and there is convergence to equal core scores in the limit of averaging over infinitely many runs. Hence, our method correctly indicates that lattice networks have no meaningful core-periphery structure.

This example is simple, but it illustrates that one should examine not simply core-score magnitude but rather how core scores are distributed. Just as with other centrality measures, this can be done visually, by computing the variance, or by computing a centralization [64].

4. Real networks. In this section, we examine core-periphery structures in networks constructed using various real-world data sets.

4.1. The Zachary Karate Club. We first consider the infamous Zachary Karate Club network [70], which consists of friendship ties between 34 members of a university karate club in the United States in the 1970s. (In this paper, we use the unweighted version of this network.) A conflict led the club to split into two new clubs, and the (unweighted) Zachary Karate Club network has become one of the standard benchmark examples for investigations of community structure [21, 51].

We visualize the network in Figure 4.1, where we have identified the nodes according to the split that occurred as a result of a longstanding disagreement between the instructor (Mr. Hi) and the club president (John A.).[1] These two primary actors are represented, respectively, by nodes 1 and 34.

In Table 4.1, we the show the nodes along with their aggregate core scores (2.11) computed using the core quality (2.10) and the transition function (2.9). We also show the node degrees, which have a high positive correlation with the aggregate core scores. Unsurprisingly, the main actors (nodes 1 and 34) have the highest aggregate core scores. One can see additional structure by considering all values of the parameters α and β rather than averaging over them. (Recall that we consider $\alpha = \beta =$ [0.01:0.01:1].) In particular, the fact that node 1 has the highest aggregate core score does not imply that it has the highest value of $C_1^*(\alpha, \beta)$ for all α and β. In Figure 4.2, we show how the top node varies as a function of α and β. Node 1 has the highest

[1] These names are pseudonyms introduced in [70].

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

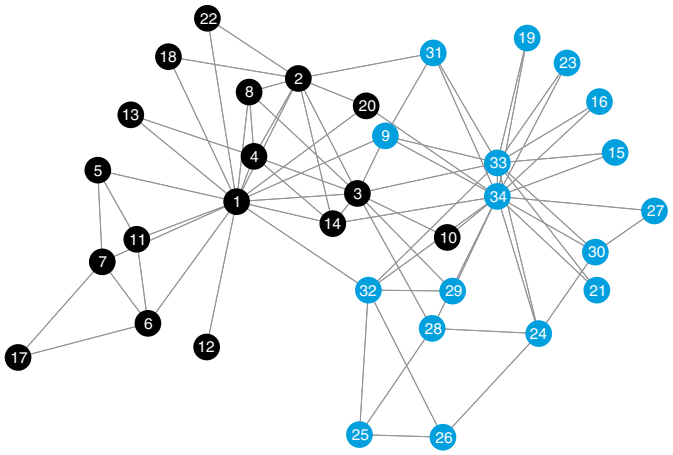


FIG. 4.1. *The Zachary Karate Club network* [70]*, which we visualize using the implementation of the Kamada–Kawai algorithm* [33] *in* [60]*. The colors represent the two groups into which the club split while it was under study.*

TABLE 4.1

Nodes in the Zachary Karate Club network along with their aggregate core scores (2.11) *computed using the core quality* (2.10) *and the transition function* (2.9)*. We also give the node degrees.*

Node	Core score	Degree	Node	Core score	Degree
1	1.0000	16	19	.2255	2
34	.9951	17	15	.2254	2
3	.9702	10	21	.2254	2
33	.8719	12	23	.2244	2
2	.8577	9	16	.2244	2
9	.7755	5	26	.2196	3
14	.7546	5	25	.2038	3
4	.7537	6	7	.1840	4
8	.6441	4	6	.1840	4
31	.5849	4	18	.1787	2
32	.5377	6	22	.1785	2
24	.4661	5	11	.1580	3
20	.4499	3	5	.1579	3
30	.4152	4	13	.1425	2
28	.3957	4	27	.1050	2
29	.3784	3	12	.0477	1
10	.2506	2	17	.0343	2

core value only about 20% of the time, whereas node 34 is the top node about 74% of the time. However, the values of α and β for which node 34 is the top node have lower core qualities R from (2.10) on average than do those for which node 1 is at the top. Such nuances are invisible if one attempts to examine coreness using only the notion of degree. Figure 4.2 also illustrates that we obtain different cores for different values of α and β.

Some of the nodes (e.g., 15, 16, 19, 21, and 23) in the Zachary Karate Club network are automorphs of each other (such nodes are *role equivalent* [18, 20]), as one can swap their labels without changing the network structure. In the limit as the number of runs in computing core-periphery structure becomes infinite, such nodes will be assigned the same aggregate core score. See our discussion of lattice networks

Downloaded 03/30/15 to 134.102.186.160. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

Downloaded 03/30/15 to 134.102.186.160. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php

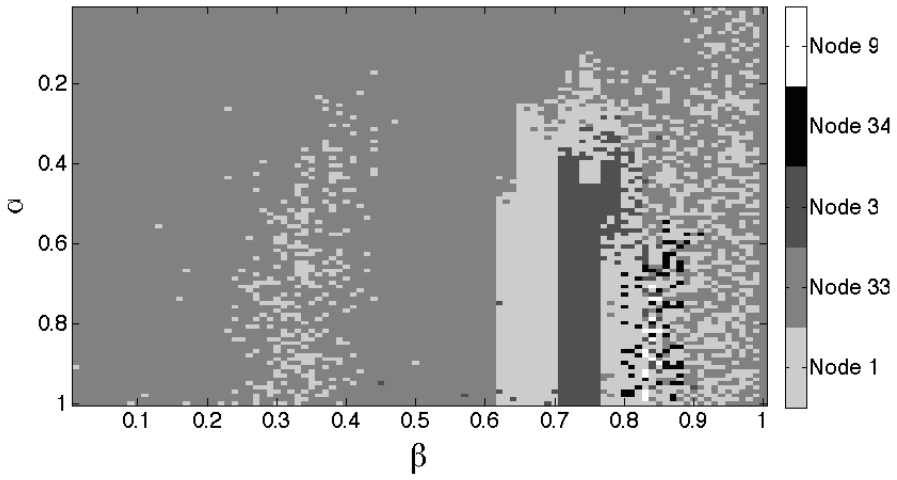


FIG. 4.2. *The node of the Zachary Karate Club network that has the top core score (i.e.,* $\arg\{\max_k(C)\}$*, where* $k \in \{1, \ldots, 34\}$ *indexes the nodes) as a function of* α *and* β*. We computed core scores using the core quality* (2.10) *and the transition function* (2.9).

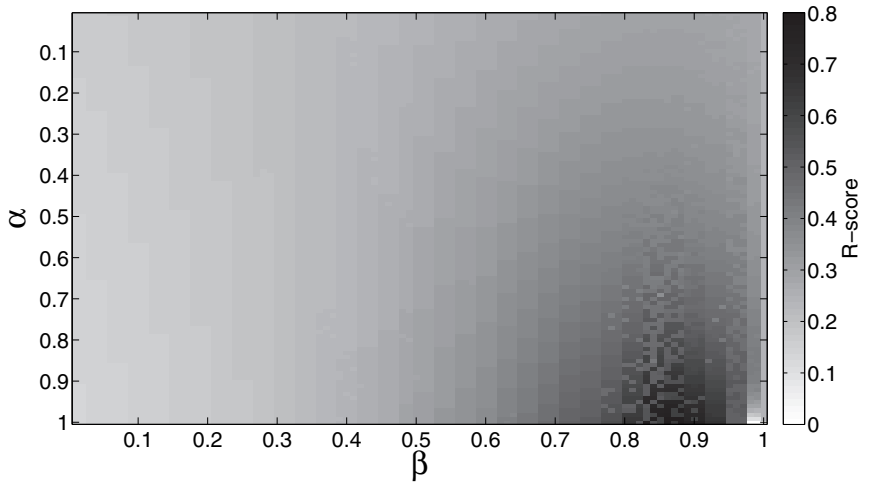


FIG. 4.3. *Core quality* R *from* (2.10) *of nodes in the Zachary Karate Club as a function of the parameters* α *and* β*. We used the transition function* (2.9).

in section 3.2.

We plot the core quality R from (2.10) as a function of α and β in Figure 4.3. The landscape of top core nodes can be complicated, especially as one considers larger networks, but examining it in a small network like the Zachary Karate Club is convenient for illustrating both how our method works and how it can expose multiple possible core-periphery structures in a network.

4.2. The London Underground. One expects many metropolitan (metro) and subway transportation networks to exhibit a core-periphery structure [53]. To

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

TABLE 4.2

The ten most core-like nodes in the London Underground network along with their aggregate core scores (2.11)*, which we obtained using the core quality* (2.10) *and the transition function* (2.9).

Node	Core score
King's Cross St. Pancras	1.0000
Farringdon	0.9773
Barbican	0.9751
Paddington	0.9693
Great Portland Street	0.9692
Moorgate	0.9663
Embankment	0.9653
Euston Square	0.9632
Edgware Road	0.9546
Baker Street	0.9490

illustrate this, we compute core scores for the London Underground ("The Tube") transportation network, which exhibits a strong core-periphery structure and a weak community structure. We collected the data for this example using the website for the London Underground (http://www.tfl.gov.uk). The Tube network that we assembled has 317 nodes (one for each station) and weighted edges that represent the number of direct, contiguous connections between two stations. For example, Baker Street and Edgware Road share an edge of weight 2, as they are adjacent stations on both the Circle Line and the Hammersmith & City Line. They are also connected by the Bakerloo Line; however, they are not adjacent stations on that line, so this does not affect the weight of the edge between them.

We partitioned the network into communities algorithmically by optimizing the modularity quality function [21, 43, 51] using the Louvain [4] computational heuristic. This splits the network into 21 communities, and the largest community that we obtained contains 19 nodes.[2] Most of these communities consist of groups of stations on a single line.

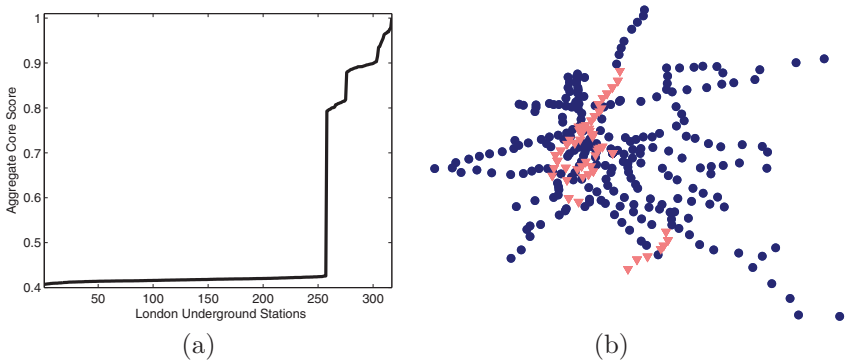
In Table 4.2, we show the results that we obtained for the London Tube network by computing aggregate core scores (2.11) using the core quality (2.10) and the transition function (2.9). We list the top ten stations and their corresponding aggregate core scores.

In Figure 4.4(a), we plot the aggregate core scores for the stations in order of ascending values. This reveals a sharp jump in aggregate core score and thereby suggests that the London Tube has a core group of (about) 60 stations and a periphery of 257 stations. Additionally, we note that considering core-periphery structure also makes it possible to distinguish between peripheral stations with the same degree centrality. (In the ordering from largest to smallest degree, stations 240–287 all have the same degree.) In Figure 4.4(b), we plot the stations using their geographical locations. The ▼ symbol designates the 60 most important stations, and the ● symbol designates the 257 other stations. The figure illustrates that it is reasonable to construe the network as dichotomized into (about) 60 core nodes and (about) 257 peripheral nodes. The large set of ▼ nodes in the middle constitute the stations in Central London (e.g., King's Cross/St. Pancras and Paddington, which are both associated with major train stations). The ▼ nodes that are farther towards the bottom right constitute the sta-

[2]The Louvain method is stochastic, so one can get slightly different network partitions from different runs of the algorithm. We simply wanted a reasonable community structure as a means of comparison, so we used a single run of the algorithm in each situation for which we computed community structure.

Downloaded 03/30/15 to 134.102.186.160. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.



(a) (b)

FIG. 4.4. (a) *The ordered list of aggregate core scores* (2.11) *for the London Underground stations suggests that there are* 60 *important stations. (We use the core quality* (2.10) *and the transition function* (2.9)*.)* (b) *We plot the stations using their geographical locations. The* ▼ *symbol designates the* 60 *most important stations, and the* ● *symbol designates the* 257 *other stations.*

tions around Waterloo, which is another major train station in London. A possible explanation for the split core is that the two clusters of core stations are separated geographically by the river Thames, which runs through central London. Most of the historical landmarks (e.g., Buckingham Palace, Trafalgar Square, and the Tower of London) are north of the Thames. The so-called "South Bank" (which is centered around Waterloo) is a 1960s arts hub containing the Royal Festival Hall, the National Theatre, and the London Eye.

4.3. Networks of network scientists. We now consider co-authorship networks among scholars who study network science. We study two such networks—one from 2006 [42] and another from 2010 [19]. These networks (which both concentrate on papers written by physicists) have 379 and 552 nodes, respectively, in their largest connected components. The nodes correspond to scholars working in the field of network science, and an edge between two of them has a weight based on the number of papers that they have co-authored. (Note that the 2006 network is not a subset of the 2010 network.)

In Table 4.3, we show the names of the scholars from both 2006 and 2010 with the top thirty aggregate core scores (2.11) using the core quality (2.10) and the transition function (2.9). In Table 4.4, we give the top 30 aggregate core scores for the 2010 network using three variant computations: the single-parameter transition function (2.14) with the product form (2.8) for the core-matrix elements; the smooth transition function (2.13) with the product form (2.8); and the usual transition function (2.9) with the p-norm (2.12) with $p = 2$ for the core-matrix elements. The ordering of the top 30 scholars is similar across different variations of the methodology, although there are some differences.

The networks of network scientists have both a sensible community structure and a sensible core-periphery structure. (Recall the block model in Figure 1.1(c) and (d).) We illustrate this point in our visualization of the 2010 network in Figure 4.5. Each pie represents a community, which we computed by optimizing modularity using the Louvain algorithm [4]. Each pie is composed of the nodes in a single community, and each node is represented by a segment colored according to its aggregate core score (2.11) computed using the core quality (2.10) and the transition function (2.9). One can plainly see that the network's core nodes are distributed throughout the various

Downloaded 03/30/15 to 134.102.186.160. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

TABLE 4.3

The 30 *nodes with the top aggregate core scores* (2.11) *for the (left)* 2006 *and (right)* 2010 *networks of network scientists. We used the core quality* (2.10) *and the transition function* (2.9).

NNS2006 Node	Core score	NNS2010 Node	Core score
Barabási, A.-L.	1.00	Barabási, A.-L.	1.00
Oltvai, Z. N.	0.97	Newman, M. E. J.	0.94
Jeong, H.	0.96	Pastor-Satorras, R.	0.93
Vicsek, T.	0.95	Latora, V.	0.93
Kurths, J.	0.88	Arenas, A.	0.93
Neda, Z.	0.87	Moreno, Y.	0.92
Ravasz, E.	0.86	Jeong, H.	0.92
Newman, M. E. J.	0.86	Vespignani, A.	0.91
Pastor-Satorras, R.	0.85	Díaz-Guilera, A.	0.90
Schubert, A.	0.85	Guimerà, R.	0.90
Boccaletti, S.	0.85	Watts, D. J.	0.89
Vespignani, A.	0.84	Vazquez, A.	0.89
Farkas, I.	0.84	Viczek, T.	0.89
Derenyi, I.	0.83	Amaral, L. A. N.	0.89
Holme, P.	0.82	Solé, R. V.	0.88
Crucitti, P.	0.81	Albert, R.	0.87
Albert, R.	0.80	Kahng, B.	0.87
Schnitzler, A.	0.80	Boccaletti, S.	0.86
Solé, R.	0.80	Oltvai, Z. N.	0.86
Rosenblum, M.	0.79	Barthelemy, M.	0.85
Tomkins, A.	0.79	Kurths, J.	0.84
Moreno, Y.	0.78	Fortunato, S.	0.84
Latora, V.	0.78	Marchiori, M.	0.83
Rajagopalan, S.	0.78	Kertész, J.	0.83
Raghavan, P.	0.77	Caldarelli, G.	0.82
Pikovsky, A.	0.76	Dorogovtsev, S. N.	0.81
Kahng, B.	0.75	Boguñá, M.	0.80
Díaz-Guilera, A.	0.74	Goh, K. I.	0.80
Vazquez, A.	0.74	Crucitti, P.	0.80
Kim, B.	0.74	Strogatz, S. H.	0.80

communities and that many communities have both core and peripheral nodes.

We calculated community structures in which the 2006 network is split into 19 communities and the 2010 network is split into 25 communities, although different community-detection methods yield somewhat different partitions of the networks [30]. For example, one previous examination [52] of community structure in the 2006 network of network scientists using a spectral tripartioning method identified three large groups: one in which A.-L. Barabási is the key node (in the sense of having the largest "community centrality" [42] in the group), one in which M. E. J. Newman is the key node, and one in which A. Vespignani and R. Pastor-Satorras are the two key nodes. As shown in Table 4.3, all four of these nodes have very high aggregate core scores.

Individual communities in both the 2006 and 2010 networks exhibit a core-periphery structure. As indicated above, the core nodes are distributed throughout the communities. In the 2006 network, 12 of the 19 communities contain at least one node among those with the top 30 aggregate core scores in Table 4.3. In the 2010 network, 9 of the 25 communities contain at least one node in the top 30 from Table 4.3. Additionally, each of the communities in the two networks includes one or two highly connected (i.e., high-strength) nodes and several other nodes with low strengths. In the 2006 network, the mean strength is 4.8, and 17 of the 19 communities contain a node with a strength of at least 9. (There are 43 such nodes in

Downloaded 03/30/15 to 134.102.186.160. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

Downloaded 03/30/15 to 134.102.186.160. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php

TABLE 4.4

The 30 nodes with the top aggregate core scores from the 2010 network of network scientists. From left to right, we computed these scores using the single-parameter transition function (2.14) and the product normalization (2.8) (using the parameter values $\alpha = [0.0001{:}0.0001{:}1]$ in MATLAB notation), the smooth two-parameter function (2.13) and the product normalization (using the parameter values $\alpha = \beta = [0.01{:}0.01{:}1]$), and the sharp two-parameter function (2.9) and the 2-norm normalization (i.e., (2.12) with $p = 2$, again using $\alpha = \beta = [0.01{:}0.01{:}1]$). Note that the second column in Table 4.3 uses the sharp two-parameter function and the product norm.

NNS2010 node	SP&PN	NNS2010 node	SmF&PN	NNS2010 node	ShF&2N
Barabási, A.-L.	1.0000	Barabási, A.-L.	1.0000	Barabási, A.-L.	1.0000
Jeong, H.	.9868	Moreno, Y.	.9702	Newman, M. E. J.	.9954
Vespignani, A.	.9859	Vespignani, A.	.9536	Pastor-Satorras, R.	.9932
Pastor-Satorras, R.	.9851	Jeong, H.	.9361	Jeong, H.	.9910
Newman, M. E. J.	.9788	Newman, M. E. J.	.9176	Vespignani, A.	.9888
Arenas, A.	.9765	Arenas, A.	.9129	Moreno, Y.	.9888
Moreno, Y.	.9762	Guimerà, R.	.8942	Díaz-Guilera, A.	.9862
Latora, V.	.9649	Díaz-Guilera, A.	.8809	Latora, V.	.9829
Guimerà, R.	.9638	Pastor-Satorras, R.	.8755	Arenas, A.	.9819
Vazquez, A.	.9616	Boccaletti, S.	.8686	Solé, R. V.	.9812
Díaz-Guilera, A.	.9604	Vicsek, T.	.8355	Amaral, L. A. N.	.9773
Vicsek, T.	.9491	Amaral, L. A. N.	.8341	Boccaletti, S.	.9768
Amaral, L. A. N.	.9470	Latora, V.	.8130	Vicsek, T.	.9737
Albert, R.	.9415	Barthelemy, M.	.8107	Guimerà, R.	.9712
Boccaletti, S.	.9379	Vazquez, A.	.8069	Vazquez, A.	.9689
Watts, D. J.	.9346	Kurths, J.	.7714	Kahng, B.	.9679
Solé, R. V.	.9321	Kahng, B.	.7633	Kurths, J.	.9676
Kahng, B.	.9309	Oltvai, Z. N.	.7616	Kertész, J.	.9624
Kurths, J.	.9241	Caldarelli, G.	.7462	Bornholdt, S.	.9577
Oltvai, Z. N.	.9197	Kertész, J.	.7096	Dorogovtsev, S. N.	.9554
Barthelemy, M.	.9183	Albert, R.	.7023	Marchiori, M.	.9549
Marchiori, M.	.9167	Watts, D. J.	.6861	Watts, D. J.	.9526
Caldarelli, G.	.9022	Porter, M. A.	.6842	Albert, R.	.9493
Kertész, J.	.8914	Solé, R. V.	.6823	Barthelemy, M.	.9488
Fortunato, S.	.8883	Fortunato, S.	.6761	Oltvai, Z. N.	.9478
Goh, K. I.	.8852	Kaski, K.	.6752	Caldarelli, G.	.9474
Kim, D.	.8836	Tomkins, A. S.	.6648	Havlin, S.	.9458
Danon, L.	.8773	Boguñá, M.	.6584	Mendes, J. F. F.	.9443
Boguñá, M.	.8747	Goh, K. I.	.6458	Stauffer, D.	.9408
Strogatz, S. H.	.8690	Kim, D.	.6411	Tomkins, A. S.	.9401

the entire network.) In the 2010 network, the mean strength is 4.7, and 20 of the 25 communities contain a node with a strength of at least 10. (There are 50 such nodes in the entire network.) The 2006 and 2010 networks of network scientists are examples that contains both an identifiable community structure and an identifiable core-periphery structure. However, methods to detect core-periphery structure need not indicate anything about community structure and vice versa. As we discussed previously, community structure and core-periphery structure provide different lenses through which to view a network [69]. There can be examples in which a core and a periphery are describable as separate communities, but community structure and core-periphery structure are different concepts.

In Figure 4.6, we zoom in on the largest community (which contains 53 nodes) in the 2010 network of network scientists. This community includes the node (A.-L. Barabási) with the highest aggregate core score. This figure illustrates that nodes with high core scores occupy well-connected positions inside of their communities as well as in the entire network.

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

Downloaded 03/30/15 to 134.102.186.160. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php

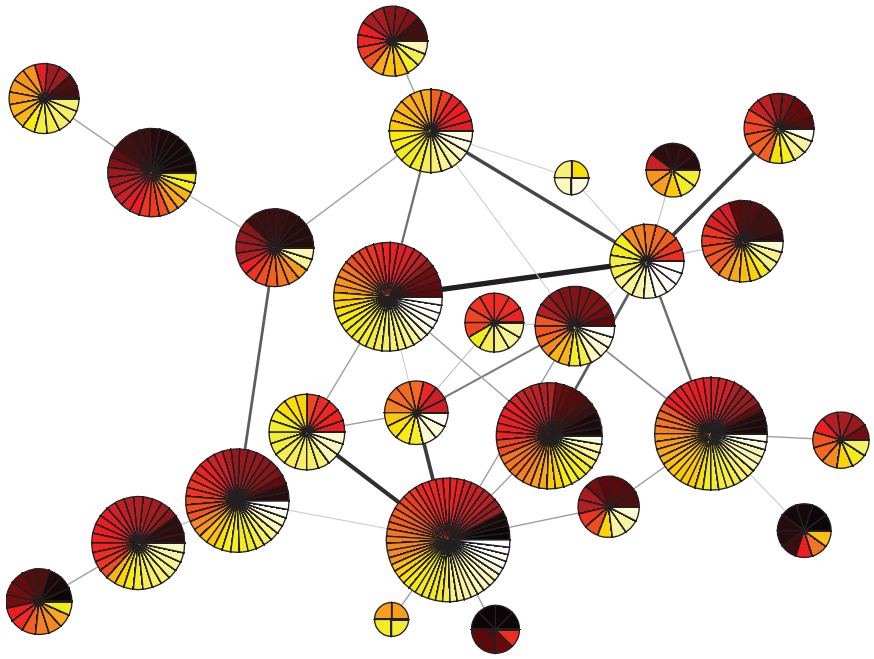


FIG. 4.5. *Visualization of the* 2010 *network of network scientists. Each pie represents a community, and the shading indicates the rank order of the nodes' aggregate core scores* (2.11)*, which we computed using the core quality* (2.10) *and the transition function* (2.9)*. Darker colors indicate higher rankings; the colors are spaced evenly over all (aggregate) core scores and contain no information about the score distribution. Each wedge represents a single node, and larger pies contain more nodes. The darkness of the edges represents the total strength of connections between communities. We produced this visualization using code described in* [60] *that uses the Kamada–Kawaii algorithm* [33] *to locate the centers of the pies. We then tweaked the center locations by hand.*

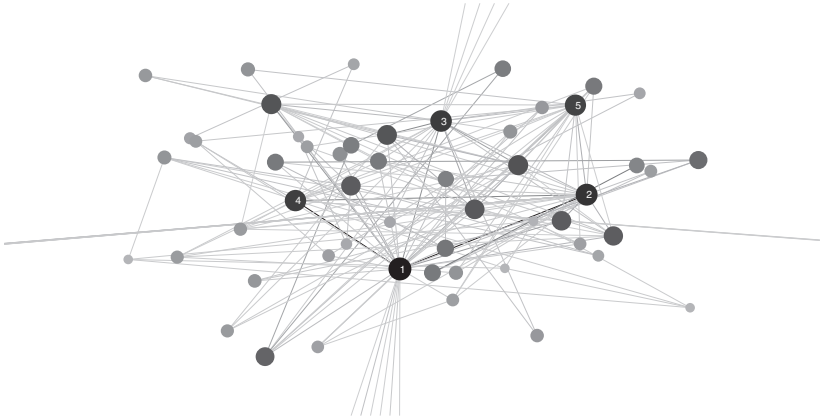


FIG. 4.6. *Magnification of the largest community in the* 2010 *network of network scientists. The darkness of the edges corresponds to the strength of the edges, and the size and darkness of the nodes represent the aggregate core score. (Edges that leave the picture are connected to nodes in other communities.) The five labeled nodes and their corresponding core scores are A.-L. Barabási* (1)*, H. Jeong* (.9181)*, T. Vicsek* (.8856)*, R. Albert* (.8737)*, and Z. N. Oltvai* (.8550)*.*

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

Downloaded 03/30/15 to 134.102.186.160. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php

TABLE 4.5

Senators in the 108*th Congress along with their aggregate core scores* (2.11) *and the percentage of bills for which they voted in line with their political parties (PV). We determined the core scores using the core quality* (2.10) *and the transition function* (2.9).

Node	Core score	PV	Node	Core score	PV
■ Chuck Grassley [R - IA]	1	97%	■ Olympia Snowe [R - ME]	0.3667	82%
■ Thad Cochran [R - MS]	0.9864	98%	■ Lincoln Chafee [R - RI]	0.3580	78%
■ Mitch McConnell [R - KY]	0.9628	98%	• Ben Nelson [D - NE]	0.3512	72%
■ Pete Domenici [R - NM]	0.9476	96%	• John Breaux [D - LA]	0.3448	74%
■ Bill Frist [R - TN]	0.8943	97%	• Max Baucus [D - MT]	0.3378	82%
■ Pat Roberts [R - KS]	0.8712	97%	• Patty Murray [D - WA]	0.3339	95%
■ Conrad Burns [R - MT]	0.8595	96%	• Mary Landrieu [D - LA]	0.3322	85%
■ Jim Bunning [R - KY]	0.8472	97%	• Blanche Lincoln [D - AR]	0.3274	87%
■ Saxby Chambliss [R - GA]	0.8132	97%	• Tim Johnson [D - SD]	0.3192	94%
■ Orrin Hatch [R - UT]	0.7969	97%	• Mark Pryor [D - AR]	0.3172	89%
■ Bob Bennett [R - UT]	0.7966	97%	• Evan Bayh [D - IN]	0.3169	86%
■ Jim Talent [R - MO]	0.7625	97%	• Kent Conrad [D - ND]	0.3038	88%
■ Kit Bond [R - MO]	0.7481	96%	• Byron Dorgan [D - ND]	0.3036	91%
■ Ted Stevens [R - AK]	0.7177	96%	• Debbie Stabenow [D - MI]	0.3023	96%
■ John Cornyn [R - TX]	0.6890	96%	• Tom Carper [D - DE]	0.3021	86%
■ Mike Crapo [R - ID]	0.6819	96%	• Barbara Mikulski [D - MD]	0.2994	96%
■ Liddy Dole [R - NC]	0.6739	96%	• Harry Reid [D - NV]	0.2960	93%
■ Sam Brownback [R - KS]	0.6736	96%	• Tom Daschle [D - SD]	0.2927	94%
■ Lamar Alexander [R - TN]	0.6676	97%	• Ron Wyden [D - OR]	0.2904	93%
■ Larry Craig [R - ID]	0.6540	96%	• Bill Nelson [D - FL]	0.2899	93%
■ George Allen [R - VA]	0.6323	96%	• Maria Cantwell [D - WA]	0.2836	95%
■ Richard Shelby [R - AL]	0.6094	95%	• Chuck Schumer [D - NY]	0.2774	94%
■ James Inhofe [R - OK]	0.5977	96%	• Jeff Bingaman [D - NM]	0.2732	92%
■ Richard Lugar [R - IN]	0.5918	96%	• Herb Kohl [D - WI]	0.2704	94%
■ Trent Lott [R - MS]	0.5822	95%	• Dianne Feinstein [D - CA]	0.2616	92%
■ Chuck Hagel [R - NE]	0.5732	95%	• Mark Dayton [D - MN]	0.2522	93%
■ Craig Thomas [R - WY]	0.5525	95%	• Hillary Clinton [D - NY]	0.2261	95%
■ Wayne Allard [R - CO]	0.5357	95%	• Jay Rockefeller [D - WV]	0.2254	93%
• Zell Miller [D - GA]	0.5327	38%	• Chris Dodd [D - CT]	0.2209	94%
■ Gordon Smith [R - OR]	0.5324	94%	• Carl Levin [D - MI]	0.2181	95%
■ Lisa Murkowski [R - AK]	0.5277	95%	• Joseph Lieberman [D - CT]	0.2154	93%
■ Norm Coleman [R - MN]	0.5189	94%	• Joe Biden [D - DE]	0.2140	92%
■ John Warner [R - VA]	0.5145	94%	• Patrick Leahy [D - VT]	0.2028	94%
■ Lindsey Graham [R - SC]	0.5055	94%	• James Jeffords [I - VT]	0.1964	88%
■ Jeff Sessions [R - AL]	0.5009	94%	• Daniel Inouye [D - HI]	0.1921	93%
■ Mike Enzi [R - WY]	0.4885	94%	• Paul Sarbanes [D - MD]	0.1765	96%
■ Rick Santorum [R - PA]	0.4741	94%	• Dick Durbin [D - IL]	0.1732	95%
■ Ben Campbell [R - CO]	0.4658	93%	• Barbara Boxer [D - CA]	0.1718	95%
■ Peter Fitzgerald [R - IL]	0.4492	93%	• Jon Corzine [D - NJ]	0.1686	95%
■ Donald Nickles [R - OK]	0.4420	93%	• Edward Kennedy [D - MA]	0.1643	94%
■ Kay Bailey Hutchison [R - TX]	0.4387	92%	• Daniel Akaka [D - HI]	0.1593	94%
■ George Voinovich [R - OH]	0.4305	92%	• Russ Feingold [D - WI]	0.1505	91%
■ Mike DeWine [R - OH]	0.4290	91%	• John Edwards [D - NC]	0.1386	96%
■ Jon Kyl [R - AZ]	0.4169	93%	• Jack Reed [D - RI]	0.1378	95%
■ John Sununu [R - NH]	0.3996	91%	• John Kerry [D - MA]	0.1246	98%
■ John Ensign [R - NV]	0.3923	90%	• Tom Harkin [D - IA]	0.1138	94%
■ Arlen Specter [R - PA]	0.3898	85%	• Fritz Hollings [D - SC]	0.1097	88%
■ Judd Gregg [R - NH]	0.3821	90%	• Frank Lautenberg [D - NJ]	0.1006	94%
■ Susan Collins [R - ME]	0.3789	84%	• Robert Byrd [D - WV]	0.0997	90%
■ John McCain [R - AZ]	0.3687	84%	• Bob Graham [D - FL]	0.0559	93%

4.4. Voting-similarity network of the United States Senate. Finally, let's consider similarity networks constructed using roll-call votes from the United States Congress. One can build such a network from a single 2-year Congress of either the Senate or the House of Representatives [48, 49, 65]. For each House and Senate, one constructs a complete (or almost complete) weighted network in which each node represents a legislator and a weighted edge between two legislators indicates the similarity of their voting patterns. In our calculation, each adjacency-matrix element A_{ij} is equal to the number of times that legislators i and j voted in the same way divided by the total number of bills on which both i and j cast a vote. This type of network is called a "similarity network," because the weights of the edges give a measure of similarity between the nodes to which they are incident. (As was recently discussed in the context of resolutions in the United Nations General Assembly [39], one can also construct networks from voting data in several other ways.)

As an example, we consider the similarity network for the 108th Senate (2003–2005), which held office during the third and fourth years of George W. Bush's presidency. In Table 4.5, we give for each Senator the aggregate core score (2.11) computed using the core quality (2.10) and the transition function (2.9). In Figure 4.7, we show scatter plots relating the strength centrality and various other centrality measures

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

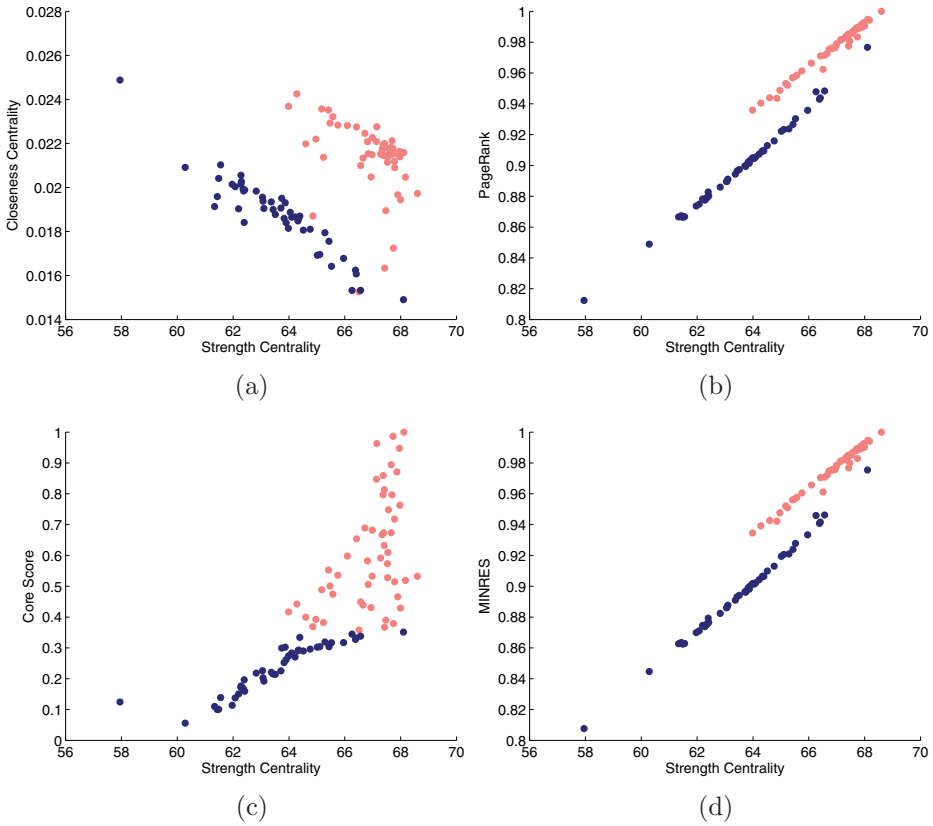


FIG. 4.7. *Scatter plots relating strength and various other centrality measures for the* 108*th Senate voting-similarity network. We show Republicans in red (lighter) and Democrats in blue (darker). In panel* (c)*, we computed aggregate core scores* (2.11) *using the core quality* (2.10) *and the transition function* (2.9).

for the 108th Senate network. We color Republicans in red and Democrats in blue. The strong similarity between the MINRES and the PageRank computation arises because (1) this example is a similarity network and (2) the aggregate core scores are relatively close together. (See the definition of MINRES in section 2.3.) They need not be similar in other examples.

Some of the centrality measures in Figure 4.7 have been used previously to study Senators and Representatives in legislation cosponsorship networks [22, 23], which have in turn been compared to modularity-based measures of political partisanship that were studied using roll-call voting networks [71]. As one can see from Figure 4.7, the different centrality measures do indeed measure different things. Methods for community detection are extremely good at separating the two main political parties—Democrats and Republicans—for modern Congresses, which suggests that voting coalitions align with party divisions [65]. In this setting, the aggregate core score separates the two communities completely, whereas none of the centrality measures by themselves are able to separate the communities well. Combinations of two such measures can sometimes distinguish a community of (mostly) Republicans from a community of (mostly) Democrats. Investigation of core-periphery structure us-

Downloaded 03/30/15 to 134.102.186.160. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

Downloaded 03/30/15 to 134.102.186.160. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php

ing aggregate core scores thus complements examination of community structure. As panel (c) illustrates, it also nicely complements existing centrality measures.

5. Conclusions and discussion. We have proposed a new family of methods for investigating core-periphery structure in networks. We generalized ideas from Borgatti and Everett [6] and designed an approach that gives nodes values (i.e., core scores) to nodes along a continuous spectrum between nodes that lie most deeply in a network core and those at the far reaches of a network periphery. Our approach can be used with a wide variety of different functions to transition between core and peripheral nodes, and it also allows one to use different ways to measure core quality. Such flexibility is important, and our method can be used to produce either a continuous measure of coreness or a discrete division of a network's nodes into a core and a periphery.[3] Moreover, sociologists have long recognized that it is important to consider both discrete and continuous core-periphery structures [57].

Our investigation of core-periphery structure complements studies of network community structure, which has been considered at great length and from myriad perspectives [21, 51]. By contrast, there are comparatively few methods for studying core-periphery structure, which we believe is just as important as community structure. As we have illustrated, networks can contain community structure, core-periphery structure, both, or neither. For example, the 2006 and 2010 networks of network scientists exhibit both types of meso-scale structures in a meaningful way. In these networks, investigating core-periphery structure reveals a global "infrastructure" that remains invisible if one searches only for community structure.

In contrast to the wealth of attention given to community structure over the last decade, the development of methods for examining core-periphery structure is in its infancy. The purpose of the present paper is conceptual development, and our current implementation of the method is slow because we use simulated annealing. Additionally, when using two-parameter transition functions, we used 10000 different (and uniformly spaced) values of (α, β), and one can improve speed considerably by considering fewer parameter values, designing schemes to sample values of α and β intelligently, or employing a one-parameter transition function. Further investigation of how to choose core-matrix elements is important, and one can also investigate core-periphery structure using perspectives that are rather different from our viewpoint in this paper.

Many networks contain meso-scale structures in addition to (or instead of) community structure, and the pursuit of methods to investigate them should prove fruitful. As we have illustrated, core-periphery structure provides one example that is worth further attention.

Appendix.

Simulated annealing. The MATLAB code that we used for simulated annealing was written by Vandekerckhove [62]. It uses the following parameters: an initial temperature of 1, a final temperature of 10^{-8}, a cooling schedule of $.8 \times T$ (where T represents the temperature), a maximum number of consecutive rejections of 1000, a maximum of 300 tries at one given temperature, and a maximum of 20 successes at one given temperature.

Acknowledgments. We thank Alex Arenas, Charlie Brummit, Mihai Cucuringu, Valentin Danchev, Colin McDiarmid, Sergey Dorogovtsev, Andrew Elliott, Martin Ev-

[3] It can also produce an assignment of nodes into more than two groups.

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

Downloaded 03/30/15 to 134.102.186.160. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php

erett, Des Higham, Sang Hoon Lee, José Mendes, Jim Moody, Alex Pothen, Gesine Reinert, Stan Wasserman, and two anonymous referees for helpful comments. We also thank Andrew Elliott for extensive discussions about code. We thank Christian Lohse for the suggestion of using the p-norm as a functional form. We thank Mark Newman for providing the data for the Zachary Karate Club network and the 2006 network of network scientists, Martin Rosvall for providing the data for the 2010 network of network scientists, and Keith Poole and Howard Rosenthal for maintaining the Congressional voting data at www.voteview.com [48]. The MATLAB code that we used for simulated annealing was written by Joachim Vandekerckhove [62]. The MATLAB code that we used for finding PageRank centrality was written by David Gleich [27, 28].

REFERENCES

[1] Y.-Y. Ahn, J. P. Bagrow, and S. Lehmann, *Link communities reveal multiscale complexity in networks*, Nature, 466 (2010), pp. 761–764.
[2] B. Ball, B. Karrer, and M. E. J. Newman, *An efficient and principled method for detecting communities in networks*, Phys. Rev. E, 84 (2011), 036103.
[3] A.-L. Barabási, *Taming complexity*, Nature Phys., 1 (2005), pp. 68–70.
[4] V. D. Blondel, J.-L. Guillaume, R. Lambiotte, and E. Lefebvre, *Fast unfolding of communities in large network*, J. Statist. Mech., 10 (2008), P10008.
[5] S. Boccaletti, V. Latora, Y. Moreno, M. Chavez, and D.-U. Hwang, *Complex networks: Structure and dynamics*, Phys. Rep., 424 (2006), pp. 175–308.
[6] S. P. Borgatti and M. G. Everett, *Models of core/periphery structures*, Social Networks, 21 (1999), pp. 375–395.
[7] S. P. Borgatti, M. G. Everett, and L. C. Freeman, *UCINet*, v. 6.289, software package, available online at http://www.analytictech.com/ucinet/ (2011).
[8] J. P. Boyd, W. J. Fitzgerald, M. C. Mahutga, and D. A. Smith, *Computing continuous core/periphery structures for social relations data with MINRES/SVD*, Social Networks, 32 (2010), pp. 125–137.
[9] R. L. Brieger, *Career attributes and network structure: A blockmodel study of a biomedical research specialty*, Amer. Sociolog. Rev., 41 (1976), pp. 117–135.
[10] C. Chase-Dunn, *Global Formation: Structures of the World-Economy*, Basil Blackwell, Oxford, UK, 1989.
[11] F. Chung and L. Lu, *The average distances in random graphs with given expected degrees*, Proc. Natl. Acad. Sci. USA, 99 (2002), pp. 15879–15882.
[12] A. Comrey, *The minimum residual method of factor analysis*, Psych. Rep., 11 (1962), pp. 15–18.
[13] C. Corradino, *Proximity structure in a captive colony of Japanese monkeys (Macaca fuscata fuscata): An application of multidimensional scaling*, Primates, 31 (1990), pp. 351–362.
[14] P. Csermely, A. London, L.-Y. Wu, and B. Uzzi, *Structure and dynamics of core-periphery networks*, J. Complex Networks, 1 (2013), pp. 93–123.
[15] M. R. da Silva, H. Ma, and A.-P. Zeng, *Centrality, network capacity, and modularity as parameters to analyze the core-periphery structure in metabolic networks*, Proc. IEEE, 96 (2008), pp. 1411–1420.
[16] F. Della Rosso, F. Dercole, and C. Piccardi, *Profiling core-periphery network structure by random walkers*, Sci. Rep., 3 (2013), 1467.
[17] P. Doreian, *Structural equivalence in a psychology journal network*, J. Amer. Soc. Inform. Sci., 36 (1985), pp. 411–417.
[18] P. Doreian, V. Batagelj, and A. Ferligoj, *Generalized Blockmodeling*, Cambridge University Press, Cambridge, UK, 2004.
[19] D. Edler and M. Rosvall, *The Map Generator Software Package*, available online at http://mapequation.org/apps/MapGenerator.html (2010).
[20] M. G. Everett and S. B. Borgatti, *Regular equivalence: General theory*, J. Math. Sociol., 19 (1994), pp. 29–52.
[21] S. Fortunato, *Community detection in graphs*, Phys. Rep., 486 (2010), pp. 75–174.
[22] J. H. Fowler, *Connecting the Congress: A study of legislative cosponsorship networks*, Political Anal., 14 (2006), pp. 454–465.
[23] J. H. Fowler, *Legislative cosponsorship networks in the U.S. House and Senate*, Social Networks, 28 (2006), pp. 456–487.

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

[24] L. C. Freeman, *The impact of computer based communication on the social structure of an emerging scientific specialty*, Social Networks, 6 (1984), pp. 201–221.
[25] A. Garas, F. Schweitzer, and S. Havlin, *A k-shell decomposition method for weighted networks*, New J. Phys., 14 (2012), 083030.
[26] M. Girvan and M. E. J. Newman, *Community structure in social and biological networks*, Proc. Natl. Acad. Sci. USA, 99 (2002), pp. 7821–7826.
[27] D. F. Gleich, *PageRank*, software package, available online at http://www.mathworks.com/matlabcentral/fileexchange/11613-pagerank (2006).
[28] D. F. Gleich, *Models and Algorithms for PageRank Sensitivity*, Ph.D. thesis, Institute for Computational & Mathematical Engineering, Stanford University, Stanford, CA, 2009, Chapter 7.
[29] M. C. González, H. J. Herrmann, J. Kertész, and T. Vicsek, *Community structure and ethnic preferences in school friendship networks*, Phys. A, 379 (2007), pp. 307–316.
[30] B. H. Good, Y.-A. de Montjoye, and A. Clauset, *Performance of modularity maximization in practical contexts*, Phys. Rev. E, 81 (2010), 046106.
[31] M. Granovetter, *The strength of weak ties: A network theory revisited*, Sociolog. Theory, 1 (1983), pp. 201–233.
[32] P. Holme, *Core-periphery organization of complex networks*, Phys. Rev. E, 72 (2005), 046111.
[33] T. Kamada and S. Kawai, *An algorithm for drawing general undirected graphs*, Inform. Process. Lett., 31 (1988), pp. 7–15.
[34] S. Kirkpatrick, C. D. Gelatt, Jr., and M. P. Vecchi, *Optimization by simulated annealing*, Science, 220 (1983), pp. 671–680.
[35] P. Krugman, *The Self-Organizing Economy*, Oxford University Press, Oxford, UK, 1996.
[36] E. O. Laumann and F. U. Pappi, *Networks of Collective Action: A Perspective on Community Influence*, Academic Press, New York, 1976.
[37] J. Leskovec, K. J. Lang, A. Dasgupta, and M. W. Mahoney, *Community structure in large networks: Natural cluster sizes and the absence of large well-defined clusters*, Internet Math., 6 (2009), pp. 29–123.
[38] A. C. F. Lewis, N. S. Jones, M. A. Porter, and C. M. Deane, *The function of communities in protein interaction networks at multiple scales*, BMC Systems Biol., 4 (2010), 100.
[39] K. T. Macon, P. J. Mucha, and M. A. Porter, *Community structure in the United Nations General Assembly*, Phys. A, 391 (2012), pp. 343–361.
[40] P. J. Mucha, T. Richardson, K. Macon, M. A. Porter, and J.-P. Onnela, *Community structure in time-dependent, multiscale, and multiplex networks*, Science, 328 (2010), pp. 876–878.
[41] R. J. Nemeth and D. A. Smith, *International trade and world-system structure: A multiple network analysis*, Rev. Fernand Braudel Center, 8 (1985), pp. 517–560.
[42] M. E. J. Newman, *Finding community structure in networks using the eigenvectors of matrices*, Phys. Rev. E, 74 (2006), 036104.
[43] M. E. J. Newman, *Networks: An Introduction*, Oxford University Press, Oxford, UK, 2010.
[44] M. E. J. Newman and M. Girvan, *Finding and evaluating community structure in networks*, Phys. Rev. E, 69 (2004), 026113.
[45] J.-P. Onnela, J. Saramäki, J. Hyvönen, G. Szabó, D. Lazer, K. Kaski, J. Kertész, and A.-L. Barabási, *Structure and tie strengths in mobile communication networks*, Proc. Natl. Acad. Sci. USA, 104 (2007), pp. 7332–7336.
[46] L. Page, S. Brin, R. Motwani, and T. Winograd, *The Pagerank Citation Ranking: Bringing Order to the Web*, Technical Report 422, Stanford InfoLab, Stanford, CA, 1999; available online at http://ilpubs.stanford.edu:8090/422/.
[47] G. Palla, I. Derenyi, I. Farkas, and T. Vicsek, *Uncovering the overlapping community structure of complex networks in nature and society*, Nature, 435 (2005), pp. 814–818.
[48] K. T. Poole, *Voteview*, software and data sets on political leanings of congressional representatives, available online at http://voteview.com (2014).
[49] K. T. Poole and H. Rosenthal, *Congress: A Political-Economic History of Roll Call Voting*, Oxford University Press, Oxford, UK, 1997.
[50] M. A. Porter, P. J. Mucha, M. E. J. Newman, and C. M. Warmbrand, *A network analysis of committees in the U.S. House of Representatives*, Proc. Natl. Acad. Sci. USA, 102 (2005), pp. 7057–7062.
[51] M. A. Porter, J.-P. Onnela, and P. J. Mucha, *Communities in networks*, Not. Amer. Math. Soc., 56 (2009), pp. 1082–1097, 1164–1166.
[52] T. Richardson, P. J. Mucha, and M. A. Porter, *Spectral tripartitioning of networks*, Phys. Rev. E, 80 (2009), 036111.
[53] C. Roth, S. M. Kang, M. Batty, and M. Barthelemy, *A long-time limit for world subway networks*, J. Roy. Soc. Interface, 9 (2012), pp. 2540–2550.

Downloaded 03/30/15 to 134.102.186.160. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

[54] W. G. Roy, *The unfolding of the interlocking directorate structure of the United States*, Amer. Sociol. Rev., 48 (1983), pp. 248–257.
[55] M. Shanahan and M. Wildie, *Knotty-centrality: Finding the connective core of a complex network*, PLoS One, 7 (2012), e36579.
[56] P. E. Simonds, *Outcast males and social structure among bonnet macaques (Macaca radiata)*, Amer. J. Physical Anthropol., 38 (1973), pp. 599–604.
[57] D. A. Smith and D. R. White, *Structure and dynamics of the global economy: Network analysis of international trade*, Social Forces, 70 (1992), pp. 857–893.
[58] D. Snyder and E. L. Kick, *Structural position in the world system and economic growth,* 1955–1970: *A multiple-network analysis of transnational interactions*, Amer. J. Sociol., 84 (1979), pp. 1096–1126.
[59] S. Steiber, *The world system and world trade: An empirical explanation of conceptual conflicts*, Sociol. Quart., 20 (1979), pp. 23–26.
[60] A. L. Traud, C. Frost, P. J. Mucha, and M. A. Porter, *Visualization of communities in networks*, Chaos, 19 (2009), 041104.
[61] A. L. Traud, E. D. Kelsic, P. J. Mucha, and M. A. Porter, *Comparing community structure to characteristics in online collegiate social networks*, SIAM Rev., 53 (2011), pp. 526–543.
[62] J. Vandekerckhove, *General Simulated Annealing Algorithm*, software package, available online at http://www.mathworks.de/matlabcentral/fileexchange/10548 (2008).
[63] I. Wallerstein, *The Modern World-System*, Academic Press, New York, 1974.
[64] S. Wasserman and K. Faust, *Social Network Analysis: Methods and Applications*, Structural Analysis in the Social Sciences, Cambridge University Press, Cambridge, UK, 1994.
[65] A. S. Waugh, L. Pei, J. H. Fowler, P. J. Mucha, and M. A. Porter, *Party Polarization in Congress: A Network Science Approach*, arXiv:0907.3509, 2013.
[66] H. C. White, S. A. Boorman, and R. L. Breiger, *Social structure from multiple networks. I. Blockmodels of roles and positions*, Amer. J. Sociol., 81 (1976), pp. 730–780.
[67] S. W. Williams, *Internal colonialism, core-periphery contrasts and devolution: An integrative comment*, Area, 9 (1977), pp. 272–278.
[68] C. L. Willis and S. J. McNamee, *Social networks of science and patterns of publication in leading sociology journals,* 1960 *to* 1985, Sci. Commun., 11 (1990), pp. 363–381.
[69] J. Yang and J. Leskovec, *Structure and overlaps of communities in networks*, ACM Trans. Intell. Systems Tech., to appear.
[70] W. W. Zachary, *An information flow model for conflict and fission in small groups*, J. Anthropol. Res., 33 (1977), pp. 452–473.
[71] Y. Zhang, A. J. Friend, A. L. Traud, M. A. Porter, J. H. Fowler, and P. J. Mucha, *Community structure in Congressional cosponsorship networks*, Phys. A, 387 (2008), pp. 1705–1712.

Downloaded 03/30/15 to 134.102.186.160. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.