Core-periphery structure in networks: A statistical exposition

Eric Yanchenko and Srijan Sengupta[*]

Department of Statistics,
North Carolina State University,
SAS Hall, 2311 Stinson Dr,
Raleigh, NC 27607
e-mail: ekyanche@ncsu.edu*;* ssengup2@ncsu.edu

Abstract: Many real-world networks are theorized to have *core-periphery structure* consisting of a densely-connected core and a loosely-connected periphery. While this phenomenon has been extensively studied in a range of scientific disciplines, it has not received sufficient attention in the statistics community. In this expository article, our goal is to raise awareness about this topic and encourage statisticians to address the many open inference problems in this area. To this end, we first summarize the current research landscape by reviewing the metrics and models that have been used for quantitative studies on core-periphery structure. Next, we formulate and explore various inferential problems in this context, such as estimation, hypothesis testing, and Bayesian inference, and discuss related computational techniques. We also outline the multidisciplinary scientific impact of core-periphery structure in a number of real-world networks. Throughout the article, we provide our own interpretation of the literature from a statistical perspective, with the goal of prioritizing open problems where contribution from the statistics community will be most effective and important.

1. Introduction

In its simplest form, a network[1] is a mathematical representation of a set of objects (e.g., social actors, entities) which interact (i.e., have some relationship) with each other. The interacting objects are called *nodes* and their interactions are called *edges.* In our highly-interconnected world, networks represent a powerful model to study complex systems. As a result, many fields study networks, e.g., social interpersonal networks, where each node is a person and each edge represents some interaction like friendship or social contact [110, 83, 60, 31, 51]; infrastructural networks like airport networks, where each node is an airport

---

[*]Corresponding author.
[1]Throughout this work, we use the term "network" to refer the mathematical object that is also known as a "graph" or "network-graph." The term "network" can also mean the collection of interacting entities in the real world, but the meaning should be clear from context.
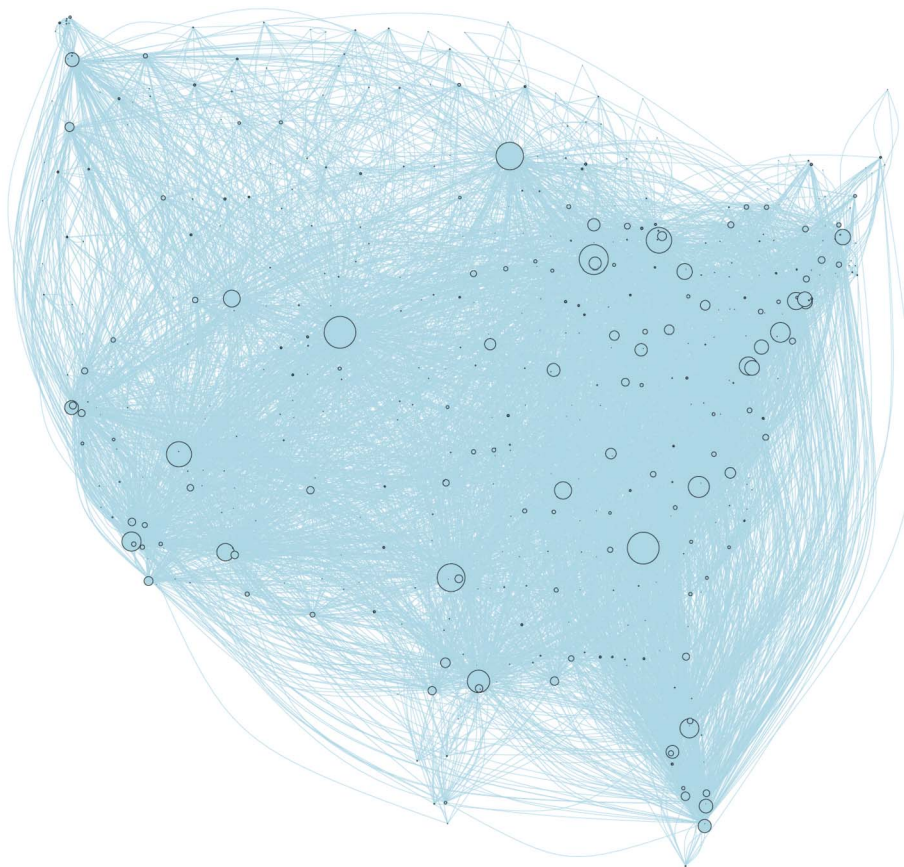
FIG 1. *Airport network of the contiguous United States where each node is an airport and an edge represents a flight between two airports [27]. The size of each node corresponds to its degree.*

and each edge represents a flight between them, as in Figure 1 [50, 75]; citation networks where each node is a paper and each edge represents a citation [72, 100, 18, 24]; and biological networks where each node is a cell or molecule and each edge represents an interaction in a biological process [46, 3, 84].

Statistics is a key component of today's data-driven science of networks. Quoting from a high-profile joint editorial in the multidisciplinary journal Network Science: "*Statistics is often defined as the study of data, involving anything from its collection, preparation, and management to its exploration, analysis, and presentation. In this view, our definition of network science delineates a subarea of statistics concerned with data of a peculiar format*" [19]. Fittingly, the last two decades have seen excellent development of the statistical literature on networks, covering several aspects of statistical analysis, such as: data-generating models [53, 52, 95, 113]; consistency of estimation [11, 138, 73]; spec-

tral decomposition [93, 126, 112], anomaly detection [97, 67, 58], and hypothesis testing [12, 10, 79, 130].

Statistical formalism is especially important in the *meso-scale* or medium-scale analysis of networks, i.e., analyzing the properties of groups of nodes, instead of analyzing individual nodes (local-scale) or the entire network (system-scale). In the statistics literature, by far the most studied meso-scale property is *community structure* [92, 91, 41]. A community is a group of nodes, and a network is considered to have a community structure if nodes are heavily connected within communities but only loosely connected between communities, implying that nodes in the same community share certain fundamental characteristics.

The topic of this review is another meso-scale feature, *core-periphery* (CP) structure [15, 28], which appears prominently in network science but has received little attention from the statistics community. Under this property, the set of nodes consists of two groups, a *core* and a *periphery*. Core nodes are densely connected to each other as well as to periphery nodes, whereas periphery nodes are only sparsely inter-connected. An alternative definition is that core nodes are a short distance from all other nodes. CP structure has been observed in a wide variety of real-world networks. For example, in global trading networks, countries with large economies trade with both large and small economies, forming the core, while small economies are less likely to trade with each other, forming the periphery [71, 81]. In airport networks, major airports (corresponding to large cities or airline hubs) have flights to other major airports as well as regional airports, but smaller airports have few flights between themselves [77, 78]. Academic citation networks also exhibit a CP structure as high-profile papers receive citations from many types of papers whereas obscure papers are less likely to cite each other [135, 111, 129]. In each of these examples, the groups of nodes (representing the core or the periphery) share some fundamental, underlying characteristic, which makes their assignment to the correct group an important task.

Studying the CP structure of a network can be important for understanding individual nodes as well. Nodes in the core are likely more "influential" to the network in some sense. For example, in a power grid network, core power plants are more vital to the health of the grid than those in the periphery [e.g., 131]. If a core power plant stops being operational due to a storm or a nefarious actor, the consequences on the power grid are much greater than if a peripheral plant went offline. Additionally, nodes on the boundary between the core and periphery may play a unique, mediatory role between the two groups [23]. Understanding the CP structure of a network helps the researcher determine which nodes are the most significant and worthy of further investigation.

There are some key distinctions between a CP structure and community structure. Community structure implies low edge density between communities and high edge density within communities. In contrast, a CP structure implies two groups, typically one small group (core) and one large group (periphery), with high edge density within the first group, medium edge density between the two groups, and low edge density within the second group. Under community structure, all nodes prefer within-group connections over between-group connec-
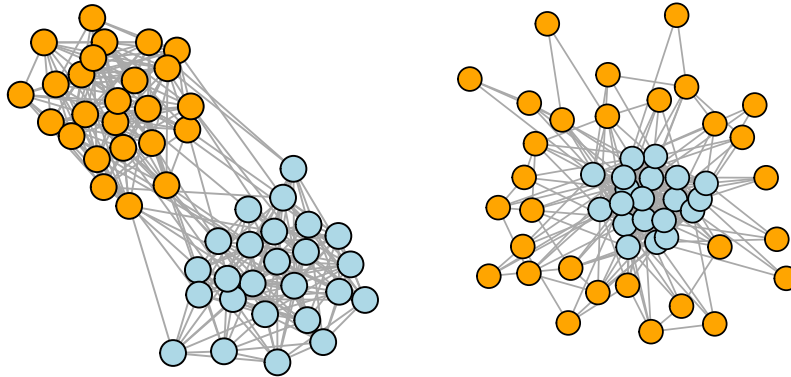
Fig 2. *Comparison of networks generated with a community structure (left) versus a core-periphery structure (right).*

tions. Under a CP structure, core nodes prefer within-group connections over between-group connections, but periphery nodes prefer between-group connections over within-group connections. Another fundamental difference between these two structures is that a network with communities can be broken down into separate, minimally interacting, self-contained sets of nodes. This is in contrast to a CP structure where core nodes influence and interact with the entire network. In Figure 2, we see how community structure is composed of two densely connected sets of nodes which have minimal edges between them. A CP structure, alternatively, has a dense core (blue nodes) that is highly connected to the periphery (orange nodes) but has few intra-periphery edges. It is also possible for a network to have multiple communities and/or CP structures.

The goal of this review is to survey the research landscape of CP structures from a statistical perspective. We are not only interested in methods to identify this network feature, but also in understanding how these structures are generated, their significance and what they mean in the context of the specific application. Our work builds off the reviews in [28] and [120] but has a greater emphasis on important statistical concepts related to CP structures, such as hypothesis testing and Bayesian inference. Another contribution of this review is that, in each section, we include a careful discussion of the existing methods, comparing and contrasting them and presenting the strengths and weakness of the literature in this area. We begin this work by introducing standard notation and random graph models. In Section 2, we describe a number of CP metrics from the network science literature. Statistical inference tasks for a CP structure are covered in Section 3, where we consider generative models, consistency, hypothesis testing, and Bayesian inference. Sections 4 and 5 explore computational techniques and real-world examples, respectively. We conclude in Section 6 by proposing important avenues for future research.

1.1. Network basics

We begin with a brief overview of notations and statistical models used to describe networks. Consider a network of n nodes and let A be its $n \times n$ *adjacency matrix*. Then $A_{ij} = 1$ if nodes i and j are connected by an edge and $A_{ij} = 0$ otherwise. If network interactions are mutual in nature, e.g., being friends on Facebook, then we say that the network is *undirected* and $A_{ij} = A_{ji}$ for all $i \neq j$. If network interactions are not mutual in nature, e.g., a paper citing another paper, then $A_{ij} = 1$ if there is an edge *from* node i *to* node j and we say that the network is *directed*. If a node has an edge with itself, e.g., sending an email to oneself, then we call this a *self-loop* and $A_{ii} = 1$. Lastly, edges need not be binary and can instead have a *weight*. A *weighted* network allows edges to take values in $\mathbb{R}$, e.g., an airport network where the nodes are airports and the weight of an edge is the number of daily flights between airports. Unless otherwise noted, for the remainder of this work, we only consider *simple* networks which are unweighted, undirected networks and contain no self-loops, and we define $m = \sum_{i<j} A_{ij}$ to be the total number of edges. Note that any network can be transformed to a simple network by symmetrizing the edges and removing weights and self-loops.

Now, let us consider data generating models for simple networks. Let $A \sim P$ mean that $A_{ij} \sim \mathsf{Bernoulli}(P_{ij})$ independently for $1 \leq i < j \leq n$. Under this general framework, a number of statistical models have been proposed with different specifications of P, the data-generating matrix. The simplest such model is the *Erdös-Rényi* (ER) model [40] where $P_{ij} = p$ for all $i \neq j$, i.e., each node has the same probability, p, of connecting with each remaining node. While this is a very simple model and unlikely to be observed in reality, it can be useful as a benchmark or null model. A generalization of the ER model to allow for heterogeneous edge probabilities is the Chung-Lu (CL) model [25], where the model parameters consist of a *weight vector* $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_n)$, and we have $P_{ij} \propto \theta_i \theta_j$. Nodes with larger θ_i are more likely to have edges in general, and, in particular, more likely to have edges with other nodes with large weights. A CL model may make sense for networks with *degree heterogeneity* where the *degree* of a node is the number of edges for that node, defined as $d_i = \sum_{j=1}^{n} A_{ij}$ for node i. The degree of a node is useful and simple measure of node importance.

To generate networks with community structure, a standard approach is to use a stochastic block model (SBM) [54]. Let $\boldsymbol{c} = (c_1, \ldots, c_n)$ be the community assignment vector where $c_i = b$ means that node i is in community $b \in \{1, \ldots, B\}$, and B is the number of communities. Let Ω be the $B \times B$ connectivity matrix where Ω_{uv} is the probability of an edge between a node in community u and community v, and we have $P_{ij} = \Omega_{c_i c_j}$. The name "block" comes from the fact that the adjacency matrix has an approximate block structure since nodes belonging to the same community are stochastically equivalent. A generalization of the SBM is the Degree-Corrected Stochastic Block Model (DCSBM) [61] which combines the ideas of the CL model and the SBM. Using the same notation as in previous models, we have $P_{ij} = \theta_i \Omega_{c_i c_j} \theta_j$. This model keeps the same block structure as an SBM but allows for "hubs", i.e., nodes

with high degree. The popularity-adjusted blockmodel [113, 94] further extends the DCSBM to allow nodes to be central in one community but peripheral in other communities. Throughout this paper we will consider how each of these data-generating models apply to CP structures.

2. Metrics

In this section, we review metrics that are already popular in the scientific literature for quantifying the CP structure of a network. We believe that examining these approaches statistically can make significant contributions to the CP literature, and it is therefore helpful for statisticians to become familiar with them.

One of the primary challenges of studying CP structures is that a universal, formal definition does not exist. This is akin to the problem of defining communities or clusters in multivariate data analysis. Because of this, there are many descriptors to identify and quantify a CP structure, but the general form is that of a highly-connected *core* with a loosely-connected *periphery* [15, 28]. Each method implicitly defines a different CP structure, and most of these definitions can be categorized into two frameworks: the block-model framework and the transport framework. Another important point to consider when defining a CP structure is whether the core-to-periphery edges are dense or not (most authors take it to be dense). Additionally, the core size should be much smaller than the rest of the network. If, for example, 95% of the nodes were in the core, this should not constitute a CP structure.

There are several other concepts that are closely related to CP structures. Rich clubs [139] are characterized by interconnected high-degree nodes and bow-tie structures for directed networks [80, 118] have core nodes with many incoming and outgoing edges. For the sake of time and space we limit this discussion to CP structures. Lastly, we define a "core-periphery assignment" to be an integer-valued vector c of length n, such that $c_i = 1$ if node i is in the core and $c_i = 0$ otherwise. We define the size of the core as $K = \sum_{i=1}^{n} c_i$.

2.1. Block model metrics

First, we consider metrics which assume a block CP structure, i.e., the adjacency and/or data-generating matrix have a block structure. The defining characteristic of this approach is that the core is dense and the periphery is sparse. Borgatti and Everett [15] took the first, and by far most cited, foray into studying CP structures. Their method measures how well the observed network agrees with an "ideal" CP structure. In particular, let $\Delta(c) := \Delta$ be the ideal CP structure where $\Delta_{ij} = 1$ if node i or node j is in the core, and $\Delta_{ij} = 0$ otherwise, i.e., $\Delta_{ij} = c_i + c_j - c_i c_j$. Then the Borgatti and Everett metric is

$$\rho = \mathsf{Cor}(A, \Delta) \tag{1}$$

where $\mathsf{Cor}(\cdot, \cdot)$ is the Pearson correlation.

If the CP assignments are known *a priori*, then one simply computes ρ for this assignment. For example, in a world trade network where countries are nodes and edges represent inter-country trade, we might consider the nations with large economies as the core and compute ρ with these labels [70]. More often, however, we do not have any information on the nodes so we want to determine whether *any* assignment has a CP structure. In this more realistic and challenging case, ρ is maximized over the space of all possible assignments using a combinatorial optimization routine such as genetic algorithms [6, 108]. See Section 4 for further discussion on algorithms.

To interpret ρ from a statistical perspective, we write it out in more detail to see that

$$\rho = \frac{M_{\boldsymbol{c}}(K) - a_K}{b_K}, \text{ where} \tag{2}$$

$$M_{\boldsymbol{c}}(K) = \sum_{i<j} A_{ij}\Delta_{ij}, \ a_K = \binom{n}{2}\text{mean}(A)\text{mean}(\Delta), \ b_K = \{\text{var}(A)\text{var}(\Delta)\}^{1/2}$$

where $M_{\boldsymbol{c}}(K)$ is the number of core-core and core-periphery edges and $\text{mean}(\cdot)$ and $\text{var}(\cdot)$ are the sample mean and variance operators, respectively. Conditional on A, a_K and b_K depend only on the number of nodes in the core, K, and increase as K increases. Thus, maximizing ρ over the set of CP assignments is equivalent to maximizing the number of core-core and core-periphery edges with a penalty for the size of the core. This definition agrees with our understanding of a CP structure since it yields large values for networks with a dense core and sparse periphery. It also means that core-core and core-periphery edges are treated equivalently. This descriptor is statistically tractable because it can be easily computed and has a natural interpretation with respect to the adjacency matrix. Other authors have generalized this work [e.g., 102, 68] but we omit the discussion of these extensions here.

In a similar vein, Brusco [20] finds the CP labels which maximize

$$Z(A, \boldsymbol{c}) = \sum_{i<j, c_i=c_j=1} \mathbb{I}(A_{ij}=1) + \sum_{i<j, c_i=c_j=0} \mathbb{I}(A_{ij}=0)^2 \tag{3}$$

where $\mathbb{I}(Q) = 1$ if Q is true and 0 if false. In other words, this method maximizes the number of core-core edges while minimizing the number of periphery-periphery edges. A perfect CP structure occurs when every core node has an edge with every other core node and there are no edges between periphery nodes. This method is similar to that of Borgatti and Everett in that it maximizes core-core edges with a "penalty" term, but here the penalty is the number of periphery-periphery edges, as opposed to the core size. Additionally, (3) does not account for core-periphery edges.

For an approach with more of a statistical flavor, we consider Zhang et al. [137]. The authors consider an SBM with two blocks where p_{00}, p_{11} and p_{01}

---

[2]In the original paper, they minimize $Z'(\boldsymbol{c}) = \sum_{i<j, \boldsymbol{c}_i=\boldsymbol{c}_j=1} \mathbb{I}(A_{ij} = 0) + \sum_{i<j, \boldsymbol{c}_i=\boldsymbol{c}_j=0} \mathbb{I}(A_{ij} = 1)$ but we note that these are equivalent formulations

are the intra-community probability of an edge for communities 0 and 1 and the inter-community probability of an edge, respectively. Nodes are randomly assigned to community 0 with probability γ_0 and community 1 with probability $\gamma_1 = 1 - \gamma_0$. Then the likelihood is

$$P(A|p,\gamma) = \sum_{\boldsymbol{c}} \prod_{i<j} p_{c_i c_j}^{A_{ij}} (1 - p_{c_i c_j})^{1-A_{ij}} \prod_i \gamma_{c_i} \tag{4}$$

where the sum is taken over all possible CP assignments, $\boldsymbol{c}$. This expression is then maximized over p_{ij} and γ_k using an Expectation-Maximization (EM) algorithm [34]. This optimization problem yields q_r^i, the probability that node i is assigned to group $r \in \{0, 1\}$, as well as estimated probabilities $\hat{p}_{00}, \hat{p}_{01}, \hat{p}_{11}$. Mathematically,

$$q_r^i = \sum_{\boldsymbol{c}} q(\boldsymbol{c}) \mathbb{I}(c_i = r)$$

where

$$q(\boldsymbol{c}) = \frac{\prod_{i<j} p_{c_i c_j}^{A_{ij}} (1 - p_{c_i c_j})^{1-A_{ij}} \prod_i \gamma_{c_i}}{\sum_{\boldsymbol{c}'} \prod_{i<j} p_{c_i' c_j'}^{A_{ij}} (1 - p_{c_i' c_j'})^{1-A_{ij}} \prod_i \gamma_{c_i'}}$$

To assign binary CP labels $\boldsymbol{c}$, each node is assigned to the group for which it has a higher probability of membership, i.e., $c_i = \mathbb{I}(q_1^i > 0.5)$. Thus, this method yields a continuous model for CP labeling.

Now, if $\hat{p}_{11} > \hat{p}_{01} > \hat{p}_{00}$ then these labels correspond to a CP structure as this result directly relates to the CP definition. If the likelihood is maximized with $\hat{p}_{11} > \hat{p}_{00} > \hat{p}_{01}$, then this is strong evidence in favor of community structure and against a CP structure. Therefore, this approach yields a sense of the statistical significance of the CP structure. This model, however, requires that the inter- and intra-community probabilities be homogenous, something likely violated in practice. Additionally, the maximization of (4) could return arbitrarily large cores, but this rarely occurs in practice.

Next, we consider a method which builds on the seminal work of Caron and Fox [22]. In [22], the authors consider networks from a point-process framework in order to achieve sparsity as well as exchangeability. Let Z be a point-process on $\mathbb{R}_+^2$ and

$$Z = \sum_{i,j} z_{ij} \delta_{t_i, t_j} \tag{5}$$

where $z_{ij} = 1$ if there is an edge between nodes i and j, $\delta_{x,y}$ is the Dirac delta function which takes the value ∞ at (x, y) and 0 elsewhere with $\int \delta_{x,y}\, dx\, dy = 1$ and $t_i, t_j \in \mathbb{R}_+$ represent a kind of time index. Each node has a *sociability parameter* $\theta_i > 0$ (similar to the weights in a CL model) such that the probability of an edge between nodes $i \neq j$ is

$$P(z_{ij} = 1 | t_i, t_j, \theta_i, \theta_j) = 1 - e^{-2\theta_i \theta_j}, \tag{6}$$

independent of the time indices. The authors further develop the theory and mechanics using Completely Random Measures [64].

Naik et al. [89] take this novel approach of modeling networks and apply it to CP structures. The authors define a CP structure as a sparse network with dense core sub-network. A network is said to be *dense* if the number of edges $m = O(n^2)$ and *sparse* if $m = o(n^2)$. The probability of an edge between two nodes is formulated to be a slight generalization of (6). Specifically,

$$P(A_{ij} = 1|\boldsymbol{\theta}) = 1 - e^{-2(\theta_{i1}\theta_{j1} + \theta_{i2}\theta_{j2})} \tag{7}$$

where $\theta_{i1} \geq 0$ is a core parameter and θ_{i2} is the core-periphery parameter. The model parameters $(\theta_{i1}, \theta_{i2}, t_i)_{i=1,\ldots,n}$ are the points from a Poisson process on $\mathbb{R}^3_+$ with mean measure $\nu(d\theta_1, d\theta_2)dt$ where ν is a σ-finite measure on $\mathbb{R}^2_+$, concentrated on $\mathbb{R}^2_+ \setminus \{(0,0)\}$, where $\int_{R^2_+} \min(1, \theta_1 + \theta_2)\nu(d\theta_1, d\theta_2)$ is finite. The network is sparse if

$$\int_{\mathbb{R}^2_+} \nu(d\theta_1, d\theta_2) = \infty \tag{8}$$

and dense otherwise. Using compound Completely Random Measures [49], the authors achieve $\theta_{i1} \geq 0$ and $\theta_{i2} > 0$. Then a CP structure is enforced with the following assumptions on ν:

$$\int_{(0,\infty) \times \{0\}} \nu(d\theta_1, d\theta_2) = 0 \tag{9}$$

$$\int_{\{0\} \times (0,\infty)} \nu(d\theta_1, d\theta_2) > 0 \tag{10}$$

$$0 < \int_{(0,\infty)^2} \nu(d\theta_1, d\theta_2) < \infty. \tag{11}$$

The nodes with $\theta_{i1} > 0$ are labeled as core nodes and $\theta_{i1} = 0$ as periphery nodes. We refer the reader to the paper for full details.

By characterizing a CP structure as a dense core within a sparse graph, the authors implicitly enforce the core to be small, since a large core would make the entire network dense. The concepts of dense/core graphs are also intuitive and connected to conceptual definition of a CP structure. This approach, however, does not allow for the core-periphery edges to be dense and there is no explicit estimator, making comparison to other methods difficult.

2.2. Transport metrics

We now consider transport-based methods where the formulation depends on distance or centrality measures. The central idea is that core nodes are a short distance from all other nodes. We start with the work of Holme [55]. The author states that for a network to have a CP structure it must have a high *closeness centrality* in addition to a well-defined cluster. The closeness centrality [107] of a node is a well-known measure of centrality, defined as the inverse of the average distance between one node and all other nodes. One can also find the closeness centrality of a subset of nodes. Let V be the set of all nodes of a network and

$U \subseteq V$ be a subset of nodes with $|U| = n_U$. Then the closeness centrality of U, $C_C(U)$, is defined as

$$C_C(U) = \left(\frac{1}{n_U(n-1)} \sum_{i \in U} \sum_{j \in V \setminus \{i\}} d_{ij} \right)^{-1} \tag{12}$$

where d_{ij} is the distance between node i and j (number of edges it takes to go from node i to node j). If the average distance between nodes in U and all other nodes is small, then the terms in the sum will be small. Taking the inverse will then yield a large value of $C_C(U)$ and thus the nodes in U are deemed to be more important. On average, a short distance between nodes is a desirable characteristic of a core. Now, consider a sequence of potential cores, $V_1, V_2, \ldots$, where V_k is the *k-core*, defined as the maximal subset of nodes such that each node is connected to at least k other nodes in the subset [128, 90]. The core, V_{core}, is the k-core that maximizes the closeness centrality, i.e., $V_{core} = \arg \max_{V_k} \{C_C(V_k)\}$. Then the proposed CP metric, c_{cp}, is defined as

$$c_{cp} = \frac{C_C(V_{core})}{C_C(V)} - \frac{1}{|\mathcal{G}|} \sum_{G' \in \mathcal{G}} \frac{C_C(V'_{core})}{C_C(V')} \tag{13}$$

where $\mathcal{G}$ is a set of networks with the same degree sequence as the observed network. This "null model" is also called the *configuration model* [90].

There are several note-worthy aspects of this approach. First, the definition of the core is intuitive and computationally efficient. Next, the first term in (13) is "normalized" by the closeness centrality of the entire network. This means that if the entire network is well-connected, then the closeness of the core will not be significantly greater than the closeness of the network, making this term smaller. Thus, the core must be be significantly more well-connected than the rest of the network. This also ensures that the core is small since otherwise this term would be close to one. Lastly, Holme's metric is compared against a "baseline" value. By averaging over an ensemble of networks with the same degree sequence as the original network, c_{cp} gives a proxy for the significance of this structure in the network. A positive value of c_{cp}, therefore, signifies that there is a CP structure greater than what would be expected by a random network with the same degree sequence. Conversely, a negative value indicates the CP structure may simply be the result of noise in the data. There are, however, no statistical guarantees for this procedure and it is unclear whether the degree-preserving random network is a reasonable null model. We return to the topic of null models in Section 3. A drawback to this approach, as well as all ensuing transport methods, is that it can be difficult to develop statistical theory. For example, finding the moments of c_{cp} is highly non-trivial as it depends on k-cores and closeness centrality, both of which are non-linear functions of the adjacency matrix. Indeed, any metric that depends on graph distances is a complicated function of the adjacency matrix.

Another transport-based approach is from da Silva et al. [30]. The authors first introduce the idea of the *capacity* of a network to measure its overall con-

nectedness. Let d_{ij} be the distance between nodes i and j. Then the capacity of the network, $\mathcal{C}$, is

$$\mathcal{C} = \sum_{i<j} \frac{1}{d_{ij}}, \tag{14}$$

such that many short paths paths lead to a larger capacity. da Silva et al. argue that a core is well-connected to the rest of the graph in the sense that the removal of a core node substantially reduces the network's capacity, $\mathcal{C}$. Based on empirical findings, the authors define the core-coefficient, cc, as $cc = N/n$ where N satisfies

$$\sum_{i=0}^{N} \mathcal{C}_i = 0.9 \sum_{j=0}^{n} \mathcal{C}_j, \tag{15}$$

and where $\mathcal{C}_i$ is the capacity of the network after removing i nodes in decreasing order of closeness centrality. If the capacity substantially decreases with the removal of certain nodes, then these nodes proved to connect many other nodes in the network. In an airport network, for example, if CLT airport (Charlotte), the operational hub for American Airlines, became non-operational due to a hurricane, then many more connecting flights would be disrupted as compared to if RDU (Raleigh-Durham) airport became non-operational. Thus, CLT is a core node and RDU is not. The authors also suggest a threshold of $cc > 0.5$ to signify that that the network exhibits a CP structure that is more pronounced than what might otherwise be expected by chance alone. This, along with the 90% threshold in (15), however, were arbitrarily selected based on a small empirical study and not justified theoretically. This approach is similar to that of Holme [55] in that it does not require any complicated optimization to detect the core, but rather chooses the nodes with largest closeness centrality. Additionally, da Silva et al. implicitly control for the size of the core. For example, if a network had a large, highly-connected core, then the entire network would necessarily be highly-connected as well. But then removing a single node would not reduce the capacity by a significant amount and thus the node would not be deemed a part of the core.

2.3. Other metrics

Lastly, we look at two methods which do not neatly fall into either the block model or transport paradigm. In Rossa et al. [103], the authors approach the CP identification problem from a random walk perspective. Consider a connected network and a random walk from node i to node j with probability $m_{ij} = A_{ij}/d_i$, i.e., jumps with equal probability from node i to one of its neighbors. Let $\pi_i > 0$ be the asymptotic probability of being at node i. Here, $\pi_i = d_i/2m$. Now, let $S \subseteq \{1, \ldots, n\}$ denote the nodes in a sub-network. Then the *persistence probability* α_S is probability that the random walk, currently in S, stays in S after its next jump. We find that

$$\alpha_S = \frac{\sum_{i,j \in S} \pi_i m_{ij}}{\sum_{i \in S} \pi_i} = \frac{\sum_{i,j \in S} A_{ij}}{\sum_{i \in S} d_i}. \tag{16}$$

The authors argue that under an ideal CP structure, we would have $\alpha_S = 0$ when S is the periphery group, since there would be no periphery-periphery edges. Thus, they seek to find the α-*periphery* which is the largest sub-network S such that $\alpha_S \leq \alpha$ for some $0 < \alpha < 1$. In other words, if the random walk is in an α-periphery, then it will jump out of the sub-network at the ensuing step with probability $1 - \alpha$. To do this, they consider an increasing sequence of sets $S_1, \ldots, S_n$ where S_1 is the weakest connected node and each subsequent set adds the node which leads to the smallest increase in persistence probability, ending with S_n being the entire network. This leads to a *core-periphery profile*, $0 = \alpha_1 \leq \alpha_2 \leq \cdots \leq \alpha_n = 1$ and the core is taken to be the largest set S_k such that $\alpha_{S_k} \leq \alpha$.

This approach can also be used to obtain a quantification of the CP structure in the network called the *core-periphery centralization*, C. First, plot the core-periphery profile against the proportion of nodes in the core, yielding a curve. Then take the integral of this curve and compare it to an "ideal" CP network (star graph) which has a value of 0 at α_{S_k} for $k = 1, \ldots, n-1$ and 1 for α_{S_n}. Mathematically,

$$C = 1 - \frac{2}{n-1} \sum_{k=1}^{n-1} \alpha_{S_k}. \tag{17}$$

The larger the value of C, the closer the network is to the star graph and thus the more pronounced the CP structure is in the network.

Rossa et al. define the core as the nodes where the random walk "spends the most time" or is most unlikely to leave, which is fundamentally different from other core descriptions. Another key difference with previous methods is that it finds the periphery and then takes the complement of that set to be the core, as opposed to finding the core and then setting the remaining nodes to the periphery. Similar to Holme and da Silva et al., however, it is difficult to compute standard statistical measures of (17) like its mean and variance.

Finally, we consider Cucuringu et al. [29]. Here, each node is assigned a measure of "coreness" based on the ideas of the *betweenness centrality*. Specifically, the *Path-Core* of node i is defined as

$$Path\text{-}Core(i) = \sum_{j < k, \neq i, A_{jk}=1} \frac{\sigma_{jk}(i)|_{A_{jk}=0}}{\sigma_{jk}|_{A_{jk}=0}} \tag{18}$$

where $\sigma_{jk}|_{A_{jk}=0}$ is the number of shortest paths from node j to k when $A_{jk} = 0$, i.e., the edge from i to j is removed; and $\sigma_{jk}(i)|_{A_{jk}=0}$ is the number of shortest paths from node j to k which go through node i, when $A_{jk} = 0$. Note that the sum is only over node pairs with edges, but the paths are computed between these nodes with their edge removed, thus looking at how close two connected nodes would be if the edge connecting them was removed. Then (18) is interpreted as the probability that the node is in the core. The intuition is that if many of the shortest paths in a network go through a particular node, then this node is likely part of the core.

While (18) yields a measure of coreness of each node, in order to discriminate nodes as core or periphery, Cucuringu et al. suggest to find the labels

which maximize the core-core and core-periphery edge density while minimizing the periphery-periphery edge density with a penalty for imbalanced core size. Mathematically, this maximizes

$$\xi(A, \boldsymbol{c}) = \binom{K}{2}^{-1} M_{cc} + \frac{1}{K(n-K)} M_{cp} - \binom{n-K}{2}^{-1} M_{pp} - \gamma \left| \frac{K}{n} - \beta \right| \quad (19)$$

where M_{cc}, M_{cp}, M_{pp} are the number of core-core, core-periphery and periphery-periphery edges, respectively. Here β is the core size that is being shrunk towards and γ is a tuning parameter. Instead of maximizing over all possible CP labels, the authors consider a much smaller space of solutions. Starting with the node with largest Path-Core score as the only core node, the algorithm adds nodes to the core in decreasing order of Path-Core score, yielding $n-2$ values of the objective function. Then the labels which correspond to the largest value of the objective function are kept as the core labels.

Curcuringu et al. utilize both block model and transport ideas in their approach. A centrality measure is used to compute the Path-Core score while the objective function assumes a block model CP structure. The search space of the optimization problem is also greatly reduced from exponential in n to linear. Another attractive aspect is that the maximum of $\xi(A, \boldsymbol{c})$ over $\boldsymbol{c}$ estimates an intuitive statistical quantity, namely $\bar{p}_{cc} + \bar{p}_{cp} - \bar{p}_{pp}$, where $\bar{p}_{cc}, \bar{p}_{cp}, \bar{p}_{pp}$ are the average core-core, core-periphery and periphery-periphery edge probabilities, respectively. Lastly, the core size is explicitly controlled for by a penalization term.

3. Statistical inference

Having studied some of the existing CP quantification methods, we now turn our attention to statistical inference for CP structures. This section focuses on four main aspects of a statistical inference framework: generative models, estimation, hypothesis testing, and Bayesian approaches.

3.1. Generative models

The most common model to generate a CP structure is the block model. Consider a two block SBM where p_{11}, p_{00}, and p_{01} are the block-block edge probabilities, i.e.,

$$P(A_{ij} = 1) = p_{c_i, c_j} \quad (20)$$

where $c_i = 1$ if node i is in block 1 and 0 otherwise. If $p_{11} > p_{01} > p_{00}$, then the resulting network has a CP structure with community 1 being the core and community 0 being the periphery. This model is intuitive and simple, and thus is used in many works [e.g. 137]. However, this model might be considered simplistic due to two reasons. First, the sharp boundary between the core and periphery is unrealistic. Second, the model requires perfect homogeneity within the core and periphery groups, which is unlikely to occur in real-world networks.

Fox and Caron [22] propose another generative model. Each node is given a *sociability parameter* $\theta_i > 0$ and the probability of an edge between nodes $i \neq j$ is

$$P(A_{ij} = 1) = 1 - e^{-2\theta_i \theta_j}. \tag{21}$$

Clearly, the larger θ_i and θ_j, the greater the probability of an edge. To generate a network with a CP structure, we endow nodes in the core with a larger value of θ_i than those in the periphery. This naturally results in a dense core and a sparse periphery with the core-periphery edge density being in-between. Thus, this model attains a CP structure while also allowing for heterogeneous edge probabilities and a smoother transition between the core and periphery as compared to the block model. Jia and Benson [59] describe a similar model where the edge probability is

$$P(A_{ij} = 1) = \frac{e^{\theta_i + \theta_j}}{1 + e^{\theta_i + \theta_j}}. \tag{22}$$

Notice that, as a special case, if each node in the core has parameter θ_1 and each periphery node has parameter θ_2, then this model is equivalent to the SBM. The authors propose a further extension of this model which can account for spatial information of each node. Let K_{ij} be an arbitrary (spatial) kernel function, e.g., $K_{ij} = ||\mathbf{x}_i - \mathbf{x}_j||_2$, where $\mathbf{x}$ are spatial positions and $|| \cdot ||_2$ is the Euclidean distance. Then

$$P(A_{ij} = 1) = \frac{e^{\theta_i + \theta_j}}{K_{ij} + e^{\theta_i + \theta_j}}. \tag{23}$$

This allows for node covariate information, if available, to be used to generate (or model) networks.

The main difference between these approaches is that the block model enforces a hard cutoff between the core and periphery nodes whereas the other models allow a more flexible transition. The choice of model therefore depends on the user's definition of the boundary between the core and the periphery, and on the context of the network being studied. Under both modeling frameworks, higher degree nodes are core nodes and lower degree nodes are periphery nodes. This raises a question on the interplay between node degrees and CP structure: at what point does the network simply have degree heterogeneity and at what point does it become a CP structure? This question is further investigated in Section 3.3.

3.2. Estimation/consistency

An important question when estimating underlying or hidden labels in a block model is: did we recover the *correct* labels? That is, do the labels returned by the algorithm correspond to the ground-truth? This question relates to the statistical *consistency* of a method. Since this topic is well-studied for block models, we use the 2-block SBM to discuss this further. Each node has the probability π_k of being in community k for $k \in \{0, 1\}$ and $\pi_0 + \pi_1 = 1$. Let Ω_{uv}

be the probability of an edge between nodes in group u and v for $0 \le u \le v \le 1$. Consider a adjacency matrix A sampled from this model where $\boldsymbol{c}^*$ are the true-labels and $\hat{\boldsymbol{c}}$ are estimated labels, perhaps found using a procedure in Section 2. We say that $\hat{\boldsymbol{c}}$ *is strongly consistent* [11] for $\boldsymbol{c}^*$ if

$$\lim_{n \to \infty} P(\hat{\boldsymbol{c}} = \boldsymbol{c}^*) \to 1 \tag{24}$$

where $n \to \infty$ means that the size (number of nodes) of the network is increasing.

Despite this being a foundational question, there is little work on it for the CP structure. Though not explicitly in the context of this structure, in [116] general conditions are provided under which a procedure is consistent. One such approach is the following: let $d_{(1)}, \dots, d_{(n)}$ be the (ordered) degrees of the nodes and let I correspond to the index for which $d_{(I+1)} - d_{(I)}$ is maximal. Then assign $c_i = 0$ if $i \le I$ and $c_i = 1$ if $i > I$. If $n_1/n \to \gamma_1$ where n_1 is the number of nodes in block 1 and $\gamma_0 p_{01} + (1-\gamma_0)p_{00} < \gamma_0 p_{11} + (1-\gamma_0)p_{01}$, then (24) holds (see [116] for details). In words, find the largest gap in consecutive, sorted degrees and then assign the nodes with smaller degrees to block 0 and those with larger degrees to block 1. This yields a simple procedure to consistently estimate the ground-truth labels. Further investigation is needed, however, to ensure that this result applies to the CP structure. Additionally, this a very specific procedure so it is unclear whether consistency holds for other algorithms. There is rich literature on consistency for community detection algorithms [138, 4, 101, 11, 87], so perhaps these ideas can be (carefully) extended to CP structures. More work is needed to study the statistical consistency of the various CP metrics from Section 2.

3.3. Hypothesis testing

Consider the following question: given a network, does it exhibit a statistically meaningful CP structure? This is the question of hypothesis testing for a CP structure. For any statistical hypothesis test, there are four key ingredients: (1) model parameter of interest, (2) a test statistic based on an estimator of this model parameter, (3) null model to establish 'non-significance' and (4) rejection threshold to decide when to reject the null hypothesis.

The biggest hurdles in networks hypothesis testing are determining the model parameter of interest (1) and deciding on a sensible null model (3). While (1) is not strictly necessary to carry out a test, it is crucial in order to root the test in the standard statistical hypothesis testing framework. None of the methods presented, save Zhang et al. [137], mention the model parameter being estimated. This is not a merely a problem for the CP structure, but also in community detection hypothesis tests [e.g., 12]. These methods simply present a test statistic without a model parameter. Additionally, a careful treatment of the null model is also lacking in the literature, although other authors have highlighted its importance [28].

Developing rigorous statistical methods to determine the presence of a CP structure is crucial to network science. Otherwise we would likely claim that ev-
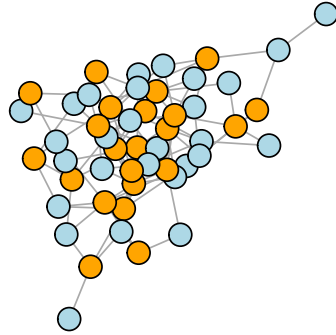
FIG 3. *Erdös-Rényi network generated with $n = 50$ and $p = 0.10$. The method from [15] was used to find core-periphery labels where orange indicates a core node and blue indicates a periphery node.*

ery network has a CP structure, which could be a false inference with significant practical implications. For example, consider a world trade network where nodes are countries and edges represent trade between the countries and assume that a recession occurs in one country. If the network has a CP structure, then this recession would impact the entire world economy if it is a core node but only a minimal impact if it is a periphery node. As another example, consider a social network where nodes are people and edges are interactions, and assume that a highly contagious disease breaks out in a small set of people. If these people are part of a core, then the disease could easily spread to the entire network, but not if they are part of the periphery. As a further motivating example, consider the network in Figure 3. This network was generated from an ER model with $n = 50$ and $p = 0.10$. The Borgatti and Everett [15] algorithm returned these CP labels where orange indicates a core node and blue indicates a periphery node. Since an ER model generated this network, these labels do not correspond to a meaningful CP structure. Thus, any investigation into similarities or characteristics of these "core" nodes could be potentially misleading.

The first hypothesis test is from Boyd et al. [17]. The authors use ρ in (1) from Borgatti and Everett [15] to run a simple permutation test to yield a p-value. Specifically, the method computes the value of ρ from the given network and then generates bootstrap samples from the network by randomly re-wiring edges in such a way that the total number of edges, m, is preserved but the degree distributions are not. Then the p-value is the proportion of bootstrapped networks which yield a larger value of ρ than that of the original network. This routine is simple, easy to implement and relies on the most well-studied CP metric.

The Boyd et al. test, however, fails to address key hypothesis testing ingredients. It does not specify the underlying model parameter being estimated. Additionally, the total number of edges is kept fixed during the bootstrap step, but this quantity is not an intrinsic feature of the network and thus may not be a reasonable feature of the network to use in the null. In other words, drawing

another network from the same data generating process is unlikely to yield the same value of m. In [68] and [39] the authors use a parametric bootstrap to generate ER networks with $p = \hat{p} = \sum_{i<j} A_{ij}/\binom{n}{2}$ computed from the original network and then find the p-value empirically. This bootstrap step preserves the *expected* number of edges so the null model is similar to that of Boyd et al. The parametric bootstrap is a promising approach for hypothesis testing but the null model must be chosen with care. It is unlikely that any real-world network resembles an ER model so using it as the null model can lead to a high Type I error when a network is not ER but also does not have a CP structure *a la* the results shown in [130].

In Rossa et al. [103] the authors present an informal test of significance. After computing the core-periphery centralization C from (17), they generate 100 new networks with the same degree sequence as the original network. Then they compute C_i^* using (17) for each randomized network $i = 1, \ldots, 100$ and find the z-score,

$$z = \frac{C - \text{mean}(C_i^*)}{\text{stdev}(C_i^*)}. \tag{25}$$

A large z-score implies that the network has a significant CP structure. While this is a natural direction to follow, there is no theory establishing the validity of the normal approximation, and formal guarantees for the Type-I error or power of the test do not exist. Furthermore, similar to Holme [55], this method uses the configuration model as the null. This approach, however, has known drawbacks. As many previous works highlight [68, 69, 39, 70], during the bootstrap step the high degree nodes are again assigned to the core which may lead to low power for the test. Taking this to the extreme, consider a network with an ideal CP structure ($A_{ij} = 1$ if node i or j is in the core and 0 otherwise). The only network which preserves this degree sequence is the original network itself. Thus, there can be no re-wiring so any hypothesis test would return a p-value of 1, i.e., we would wrongly infer that there is no CP structure. In [69] the authors propose adding a third block to the network to deal with this problem.

A fundamentally different approach to determining the significance of a CP structure (or more generally any mesoscale feature) comes from the *surprise* formulation [2, 32, 82]. This paradigm assumes that there is a population of $\binom{n}{2}$ node pairs of which m (the number of observed edges) have been drawn. Of these m realized edges, M_c are "successes" meaning that they are core-core or core-periphery edges. Then the p-value is the probability of observing as many or more "successful" edges out of m draws using the hyper-geometric distribution. These p-value computations are exact and do not require a bootstrap step.

3.4. Bayesian inference

Bayesian inference on networks has seen great progress in areas such as estimation [53, 52], exponential random graph models [21, 121, 119] and community detection [98, 86, 124]. Bayesian methods for the CP structure, however, are far less developed. We devote this subsection to the existing work in this area.

Snijders and Nowicki [116] were perhaps the first to use Bayesian techniques with block models. This paper uses a two-block SBM likelihood simlar to (4) from Zhang et al. [137]. By requiring $p_{11} > p_{00}$, the parameters are identifiable. Then the authors propose the following Gibbs sampling procedure, starting at step s of the sampler:

1. Draw $(\gamma_1^{(s+1)}, p_{11}^{(s+1)}, p_{01}^{(s+1)}, p_{00}^{(s+1)})$ from the posterior of $(\gamma_1, p_{11}, p_{01}, p_{00})|$ $(A, \boldsymbol{c}^{(s)})$
2. Draw $c_1^{(s+1)}$ from $\boldsymbol{c}_1|(\gamma_1^{(s+1)}, p_{11}^{(s+1)}, p_{01}^{(s+1)}, p_{00}^{(s+1)}, A, c_2^{(s)}, \dots, c_n^{(s)})$
3. For $i = 2, \dots, n-2$, draw $c_i^{(s+1)}$ from $c_i|(\gamma_1^{(s+1)}, p_{11}^{(s+1)}, p_{01}^{(s+1)}, p_{00}^{(s+1)},$ $A, c_2^{(s+1)}, \dots, c_{i-1}^{(s+1)}, c_{i+1}^{(s)}, \dots, c_n^{(s)})$
4. Draw $c_n^{(s+1)}$ from $c_n|(\gamma_1^{(s+1)}, p_{11}^{(s+1)}, p_{01}^{(s+1)}, p_{00}^{(s+1)}, A, c_2^{(s+1)}, \dots, c_{n-1}^{(s+1)})$

Steps 2–4 leverage the fact that

$$\frac{P(c_i = 1|A, c_j, j \neq i)}{P(c_i = 0|A, c_j, j \neq i)}$$
$$= \frac{\gamma_1}{1 - \gamma_1}(1 - p_{11})^{-n_1}(1 - p_{01})^{2n_1 - n}(1 - p_{00})^{n - n_1} \times$$
$$\left(\frac{p_{11}}{1 - p_{11}}\right)^{-M_1} \left(\frac{p_{01}}{1 - p_{01}}\right)^{M_1 - M_0} \left(\frac{p_{00}}{1 - p_{00}}\right)^{M_0} \quad (26)$$

where $M_k = \sum_j A_{ij}\mathbb{I}(c_j = k)$ and n_k is the number of nodes in group $k \in \{0, 1\}$. The authors also show that if the block labels are known and vague priors are used for the edge probabilities, then the posterior distributions are:

$$\gamma_1 \sim \mathsf{Beta}(n_1 + 1, n_0 + 1), \ p_{11} \sim \mathsf{Beta}(M_{11} + 1, \tfrac{1}{2}n_1(n_1 - 1) - M_{11} + 1)$$
$$p_{01} \sim \mathsf{Beta}(M_{01}+1, n_1 n_0 - M_{01}+1), \ p_{00} \sim \mathsf{Beta}(M_{00}+1, \tfrac{1}{2}n_0(n_0-2)-M_{00}+1). \quad (27)$$

Although never expressly mentioned in their paper, this method clearly applies to estimating a CP structure. A major advantage of the Bayesian approach is that the posterior of $\boldsymbol{c}$ gives the *probability* of each node being in the core as opposed to the *binary* assignment from a frequentist paradigm. If binary CP labels are required, then all nodes with a posterior mean of $c_i > 0.5$ could be assigned to the core, for example.

Another paper that applies Bayesian methodology to study the CP structure is Gallagher et al. [43]. The author's likelihood is the same as that of Snijders and Nowicki [116]. This paper differs from previous work by focusing on the prior distribution for the block probabilities and explicitly enforcing the CP condition through this prior. In particular,

$$P(p_{11}, p_{01}, p_{00}) \propto \mathbb{I}(0 \leq p_{00} \leq p_{01} \leq p_{11} \leq 1). \quad (28)$$

The authors also propose a generalization of this prior to allowed for a "layered" CP structure. Assuming there are K blocks, or layers, the prior becomes

$$P(p_1, \dots, p_K) \propto \mathbb{I}(0 \leq p_K \leq p_{K-1} \leq p_1 \leq 1) \quad (29)$$

where $p_{lm} = p_{\max\{l,m\}}$ is the probability of an edge between blocks l and m. Lastly, a posterior odds ratio can be used to determine which type of CP structure is a better fit to the data. Mathematically,

$$\Lambda = \frac{P(\boldsymbol{c}_b|A)}{P(\boldsymbol{c}_\ell|A)} \tag{30}$$

where $\boldsymbol{c}_b$ are the labels for the block model and $\boldsymbol{c}_\ell$ are the labels from the layered model. If $\Lambda > 1$, then the block model gives a better fit and vice-versa if $\Lambda < 1$.

The main contribution of Gallagher et al. [43] is the data-driven comparison of two different formulations of the CP structure. This does not enforce a particular type of CP structure, rather, it allows the data to dictate the more likely structure. Conversely, the CP structure is assumed or "hard-coded" into the model, meaning it does not yield a sense of the presence or absence of a CP structure, but only which type of CP structure is more probable. Still, this approach seems promising to apply to the hypothesis testing problem as well.

4. Computation

Due to improved data collection mechanisms, the modern statistician now has access to massive amounts of data. For example, online social networks can have thousands [105], millions [7] or even billions [133] of nodes and edges. Since the total possible number of core-periphery assignments increases exponentially with n, the number of nodes, if CP methods are to be relevant in our data science age, they must be computationally scalable. This is an important problem because even the authors of the well known UCINET package [16] (which includes the Borgatti and Everett [15] method) state that the computational routines are only fast enough for networks with 5000 nodes or less. This section discusses different computational techniques, including those for big data.

4.1. Objective function maximization

Most CP methods require maximizing an objective function over the set of candidate assignments whose size increases exponentially with n. Since an exhaustive search for the maximum is computationally infeasible, heuristics and approximations must be employed.

A well-known strategy for objective function maximization is a *greedy algorithm* [e.g., 62], where the locally optimal choice is selected at each iteration. This algorithm switches node labels between core and periphery and keeps the label which results in a larger value of the objective function. Many existing methods use variants of this approach [17, 68, 69, 114]. This paradigm is simple and fast but can return local optima. One way to mitigate this risk is to consider several different starting values.

Another optimization paradigm is *genetic algorithms* [6, 108]. Taking their cue from micro-evolutionary concepts, these algorithms consider a "population"

of possible solutions where the "fittest" solutions "mate" to form new solutions. Here, a solution is a CP membership vector c, the fitness is the value of the objective function for that label, and mating is the merging of two labels. Borgatti and Everett [15, 16] implement this algorithm in their work. These methods are better than greedy algorithms at reaching the global maxima, but suffer from increased computational complexity. A large number of tuning parameters and slow speeds can also plague these algorithms. Other popular combinatorial optimization routines include simulated annealing [65, 125, 37] and tabu search [47, 48]. To the extent of our knowledge, there has not been a systematic comparison of these different algorithms on different CP quantifiers.

The previous procedures can, in theory, search the entire space of solutions. A fundamentally different approach which greatly restricts the search space, but increases speed, is what we coin as *node-ordering methods*. Here, some variable like degree or centrality is used to order the nodes. Then, starting with the first (largest value) node in the core, nodes are added to the core one-by-one in decreasing order and the objective function is computed at each iteration. After all nodes have been added to the core, the labels which correspond to the largest objective function are kept. Several authors propose this algorithm with ordering based on degree [Lip, 76], centrality [da Silva et al., 30] or Core-Score [Cucuringu et al., 29]. This approach is significantly faster because it only considers $O(n)$ possible solutions as opposed to $O(2^n)$, but this smaller search spaces also means that the global maximum may not be achieved.

To see why this idea is appealing for large networks, consider Lip [76] which leverages this framework using the objective function (3) proposed by Brusco [20]. The author shows that, for a fixed core size K, (3) can be re-expressed as

$$Z(A, c) = \left(\frac{1}{2} \sum_i d_i + \frac{K(K-1)}{2} \right) - \sum_{i,c_i=1} d_i. \tag{31}$$

The first term does not depend on the core so it follows that the core should be the K nodes with largest degree. This generalizes to the case when K is unknown by adding nodes to the core in decreasing order of degree and keeping the core which corresponds to the largest value of the objective function. This algorithm runs in $O(n^2)$ time, and finds the core of a network with $n = 50,000$ in under one second. While the particular form of the objective function in (31) lends itself to a node-ordering algorithm, in general, it is unknown when the node-ordering paradigm adequately approximates the global maximum.

Lastly, likelihood maximization allows for an even broader set of computational procedures. Since the CP labels can be thought of as missing or hidden data, the Expectation-Maximization (EM) algorithm [34] is a natural choice for likelihood maximization [e.g., 116, 137]. Bayesian methods like Markov Chain Monte Carlo [e.g., 44] or Variational Bayes [e.g., 63] are other popular choices when working with likelihoods. See Section 3.4 for a further discussion on Bayesian methods. As a case study, consider Shen et al. [114] which uses the likelihood-based approach to allow for multiple cores and multiple edges between a pair of nodes. Here we present a simplified version for a single core and

binary edges. The authors consider a DCSBM to model the CP structure. If the probability of a core-core or core-periphery edge is p_c and periphery-periphery edge is p_p, then the likelihood is

$$P(A|\boldsymbol{c}, p_c, p_p) = \prod_{i<j, c_i=1 \text{ or } c_j=1} (\theta_i \theta_j p_c)^{A_{ij}} e^{-\theta_i \theta_j p_c} \prod_{i<j, c_i=c_j=0} (\theta_i \theta_j p_p)^{A_{ij}} e^{-\theta_i \theta_j p_p}$$

(32)

Then the authors propose a greedy algorithm to maximize (32) and apply it to networks with up to $80,000$ nodes.

4.2. Transport

Transport-based methods admit an entirely new suite of computational tools. For example, Holme [55] uses k-cores. We briefly describe a routine based on [90]. First, any node with degree less than k cannot possibly be in a k-core so these nodes (and corresponding edges) are removed from the network. But this means that some of the remaining nodes will now have a smaller degree, so now again we remove any nodes with degree less than k. This process continues until all remaining nodes have degree k and what remains is, by definition, a k-core. Even though this core definition is slightly restrictive, calculating a sequence of k-cores is computationally cheap ($O(m)$), which makes it appealing for large networks. Indeed, Holme applies this approach to a network with hundreds of thousands of nodes and over a million edges in [55]. Other transport methods rely on measures like distance and betweenness centrality, which take $O(m+n)$ and $O(n(m+n))$ time, respectively. We omit discussions of these algorithms but refer the interested reader to [90]. Several other existing authors specifically address computational feasibility in large networks. Most approaches, however, are designed specifically for one particular CP descriptor. Therefore, an open problem is to develop techniques which work for an arbitrary metric as there is in the community detection literature, e.g. divide-and-conquer [99, 136, 87].

5. Real-world applications

Core-periphery structures have been observed in many empirical studies across disciplines. One of the most common applications where CP structures arise is in infrastructure networks. For example,[3] every traveler from a small city in the United States knows that airport networks have a CP structure since, to travel to most places, the first flight is to a large airport and then the second flight is the destination of interest. Alternatively, those who live near large airports can easily fly to other large airports as well as to small airports. In this case, large or "hub" airports make up the core and all other airports comprise the periphery. This structure has been formally studied in [68], [77], [78] and [89]. One of the interesting findings from [89] is that hubs for Southwest Airlines (MDW and

---

[3]This example derives from a transport-based view of CP structures as it is based on shortest paths between nodes.

DAL) do not appear in the core of their analysis which means that Southwest may not use the same route structures as other airlines, i.e., funnelling all flights to their respective hubs.

Another field where CP structures are well-studied is economics. Krugman [71] famously studied how different economies arrange themselves in a CP structure and [81] found similar economic structure specifically in the European Union (EU). These authors note how this naturally leads to unequal power dynamics, making it difficult for periphery countries to affect policy and financial decisions. Not surprisingly, then, the countries hardest hit by the Eurozone crisis were those in the periphery, e.g., Greece, Spain, Portugal, Italy. In [122], the authors highlight a well-known fact in international relations that the vast majority of core countries are in the northern hemisphere and periphery countries are in the southern hemisphere, a different kind of imbalance associated with a CP structure. Lastly, the authors in [42] explore the CP structure of overnight money lending markets. With Italian banks as nodes and interbank positions as edges, the authors found that not only does this network exhibit a CP structure, but that the core shows little change over time. This phenomenon has also been observed in German banks [26].

Many social networks also exhibit a CP structure. In [23], the authors explore creativity in the film industry, with nodes representing crew members from the seven major Hollywood film studios (Universal, Paramount, Warner Bros, Columbia-Tristar, Disney, 20th Century Fox and MGM) and edges representing collaboration on the same movie. They conclude that people in-between the core and periphery have the greatest creativity since they are close enough to the core to have sufficient influence but far away enough that they still maintain their unique, creative skills. In Yang et. al [134], the authors explore the relationship between the community and CP structure in Twitter data. For this network, a node is a Twitter user in Dubai and there is an edge between users if there is a "retweet" or "mention" of another user over a thirty day period. One of the main findings in this work is that, in this network with many communities, each individual community exhibits a CP structure.

Other applications that have observed CP structures include: academic citations [88, 36, 135, 111, 129], open source software development [5], online discussion groups [9], grieving widow's family structure [85], and political blogs [1, 137]. In the political blogs example (Zhang et. al [137]), this structure is unsurprising as all blogs tend to cite famous blogs (core) whereas relatively unknown blogs (periphery) are less likely to be referenced. This structure, however, may make it more difficult for a small blog to gain prominence as the majority of the focus is on established blogs.

While many statistics papers have included real-world applications, most, if not all, of the applied papers studying CP structures have come from other disciplines and are published in non-statistics journals. Thus, an open avenue for research is for statisticians to investigate interesting, complex, real-world networks for CP structures. For the interested reader, we provide the following public network data repositories as a starting point: [38, 74, 104, 109, 27].

6. Open problems

We conclude with some important open problems in statistical inference for CP structures.

1. *How can the current CP structure methods be unified, specifically regarding a model parameter intrinsic to this structure?*
   Each method presented in Section 2 can be interpreted as an estimator that implicitly defines the CP structure in a unique way. The majority of the existing techniques make no mention of an underlying model parameter. Not only does this lack statistical rigor but it also makes it difficult to compare different methods. A major breakthrough would be identifying and estimating a model parameter that describes the CP structure in a network. Ideally, this parameter would not be tied to a specific model structure (like SBM) to allow for fair comparison across different models. Additionally, it should be interpretable in order to yield a succinct summary of the strength of the CP structure.

2. *What is an appropriate null model to represent "no CP structure"?*
   Without a satisfactory answer to this question, it is impossible to delineate between a "true" CP structure and simply noise or some other feature in the network. To this end, many authors have suggested different null models, e.g., Erdös-Rényi (ER), re-wiring preserving total number of edges and re-wiring preserving degrees (configuration model). There has, however, been little justification for the choice as well as no systematic comparison of the different nulls. Furthermore, these simplistic null models can lead to ubiquity, meaning that CP structures seemingly appear in *all* real-world networks and thus making it a useless measure in practice.[4] We argue, however, that the CP structure is *not* ubiquitous but is mistakenly attributed to many networks because of simplistic benchmarks. For example, the ER model is a popular CP benchmark but it is almost tautological that real-world networks deviate from this model. It is this deviation that leads to the "universality" of CP structures in real-world networks. Thus, having a nuanced and rigorously justified null model is paramount to advancing hypothesis testing for CP structures.

3. *What is the connection and/or trade-off between the CP and other meso-scale structures like communities?*
   It is unlikely that any real-world network *exclusively* exhibits a CP or community structure. Rather, it is more likely that networks are a mixture of different features, similar to the DCSBM which incorporates degree heterogeneity and block community structure. The PABM proposed by [113] generalizes the DCSBM and allows for flexible core-periphery structure within communities. In this direction, [132] argues that cores arise from the intersection of many overlapping communities. In [123] the authors provide a unified formulation that allows for a hybrid of a community and CP structure. In particular, the edge probability, p_{ij} of a node in group i

---

[4]For a discussion of this phenomena in other meso-scale features, see [8, 35].

and j is modeled as

$$p_{ij} = a\delta_{c_i,c_j}(C_i + C_j - C_iC_j) + b \tag{33}$$

where δ_{xy} is the Kronecker delta taking value 1 if $x = y$ and 0 otherwise, c_i is the group of node i and C_i is some measure of "coreness" for node i. Thus, the δ_{c_i,c_j} accounts for the community structure and $C_i + C_j - C_iC_j$ accounts for the CP structure. In [134] the authors found that Twitter networks are composed of multiple communities with a CP structure within each community. Lastly, in [69], the authors argue that a third block (like a community) is needed for a CP structure.

4. *Multiple core-periphery structures*

   Related to the previous point, another important research direction is identifying multiple CP structures in a network. Many of the existing approaches (e.g., Borgatti and Everett [15]) are defined to identify a single core in the network. Real-world networks, however, have been shown to exhibit multiple cores and peripheries, as in Zhang et. al [137] and Yang et. al [134], where CP structures are observed within communities. There has been some methodological work on this problem by Kojaku and Masuda [68] and Shen et al. [114]. Kojaku and Masuda extend the Borgatti and Everett [15] metric in (1) to allow for multiple cores. The proposed algorithm also automatically estimates the number of CP structures in the network and yields the statistical significance of this structure by considering an ER null model.

5. *Core-periphery identification in large networks*

   While a plethora of methods exist for community detection in large networks [e.g., 13, 33, 106], much less work exists for CP identification. In fact, the largest networks analyzed for a CP structure appear in Holme [55] which used a restrictive definition. The method in Borgatti and Everett [15], for example, is typically limited to networks with no greater than $n = 5000$ nodes. Creating new algorithms, extending existing algorithms to larger networks, borrowing ideas from other network literature [e.g., divide and conquer, 117, 87] or some combination of these is necessary for CP identification to be scalable to huge networks.

6. *More complicated networks*

   Most of the work done so far has been on relatively simple network structures. It would be interesting to investigate more complicated scenarios like temporal [56, 57], multilayer [66, 14] or heterogenous [115, 127] networks as well as hypergraphs [45, 96].

Acknowledgments

References

[1] ADAMIC, L. A. and GLANCE, N. (2005). The political blogosphere and the 2004 US election: divided they blog. In *Proceedings of the 3rd international workshop on Link discovery* 36–43.

[2] ALDECOA, R. and MARÍN, I. (2013). Surprise maximization reveals the community structure of complex networks. *Scientific Reports* **3** 1–9.

[3] ALM, E. and ARKIN, A. P. (2003). Biological networks. *Current Opinion in Structural Biology* **13** 193–202.

[4] AMINI, A. A., CHEN, A., BICKEL, P. J. and LEVINA, E. (2013). Pseudo-likelihood methods for community detection in large sparse networks. *Ann. Statist.* **41** 2097–2122. MR3127859

[5] AMRIT, C. and VAN HILLEGERSBERG, J. (2010). Exploring the impact of soclo-technlcal core-periphery structures in open source software development. *Journal of Information Technology* **25** 216–229.

[6] ANDERSON, E. J. and FERRIS, M. C. (1994). Genetic algorithms for combinatorial optimization: the assemble line balancing problem. *ORSA Journal on Computing* **6** 161–173.

[7] BACKSTROM, L., HUTTENLOCHER, D., KLEINBERG, J. and LAN, X. (2006). Group formation in large social networks: membership, growth, and evolution. In *Proceedings of the 12th ACM SIGKDD international conference on Knowledge discovery and data mining* 44–54.

[8] BASSETT, D. S. and BULLMORE, E. T. (2017). Small-world brain networks revisited. *The Neuroscientist* **23** 499–516.

[9] BECK, R. J., FITZGERALD, W. J. and PAUKSZTAT, B. (2003). Individual behaviors and social structure in the development of communication networks of self-organizing online discussion groups. In *Designing for Change in Networked Learning Environments* 313–322. Springer.

[10] BHADRA, S., CHAKRABORTY, K., SENGUPTA, S. and LAHIRI, S. (2019). A Bootstrap-based Inference Framework for Testing Similarity of Paired Networks. *arXiv preprint arXiv:1911.06869*.

[11] BICKEL, P. J. and CHEN, A. (2009). A nonparametric view of network models and Newman–Girvan and other modularities. *Proceedings of the National Academy of Sciences* **106** 21068–21073.

[12] BICKEL, P. J. and SARKAR, P. (2016). Hypothesis testing for automated community detection in networks. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **78** 253–273. MR3453655

[13] BLONDEL, V. D., GUILLAUME, J.-L., LAMBIOTTE, R. and LEFEBVRE, E. (2008). Fast unfolding of communities in large networks. *Journal of Statistical Mechanics: Theory and Experiment* **P10008**.

[14] BOCCALETTI, S., BIANCONI, G., CRIADO, R., DEL GENIO, C. I., GÓMEZ-GARDENES, J., ROMANCE, M., SENDINA-NADAL, I., WANG, Z. and ZANIN, M. (2014). The structure and dynamics of multilayer networks. *Physics reports* **544** 1–122. MR3270140

[15] BORGATTI, S. P. and EVERETT, M. G. (2000). Models of core/periphery structures. *Social Networks* **21** 375–395.

[16] BORGATTI, S. P., EVERETT, M. G. and FREEMAN, L. C. (2002). Ucinet for Windows: Software for social network analysis. *Harvard, MA: Analytic Technologies* **6**.

[17] BOYD, J. P., FITZGERALD, W. J. and BECK, R. J. (2006). Computing core/periphery structures and permutation tests for social relations data. *Social Networks* **28** 165–178.

[18] BRADLEY, J., DEVARAKONDA, S., DAVEY, A., KOROBSKIY, D., LIU, S., LAKHDAR-HAMINA, D., WARNOW, T. and CHACKO, G. (2020). Co-citations in context: Disciplinary heterogeneity is relevant. *Quantitative Science Studies* **1** 264–276.

[19] BRANDES, U., ROBINS, G., MCCRANIE, A. and WASSERMAN, S. (2013). What is network science? *Network Science* **1** 1–15.

[20] BRUSCO, M. (2011). An exact algorithm for a core/periphery bipartitioning problem. *Social Networks* **33** 12–19.

[21] CAIMO, A. and FRIEL, N. (2011). Bayesian inference for exponential random graph models. *Social Networks* **33** 41–55.

[22] CARON, F. and FOX, E. B. (2017). Sparse graphs using exchangeable random measures. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **79** 1295–1366. MR3731666

[23] CATTANI, G. and FERRIANI, S. (2008). A core/periphery perspective on individual creative performance: Social networks and cinematic achievements in the Hollywood film industry. *Organization Science* **19** 824–844.

[24] CHANDRASEKHARAN, S., ZAKA, M., GALLO, S., ZHAO, W., KOROBSKIY, D., WARNOW, T. and CHACKO, G. (2021). Finding scientific communities in citation graphs: Articles and authors. *Quantitative Science Studies* **2** 184–203.

[25] CHUNG, F. and LU, L. (2002). The average distances in random graphs with given expected degrees. *Proceedings of the National Academy of Sciences* **99** 15879–15882. MR1944974

[26] CRAIG, B. and VON PETER, G. (2014). Interbank tiering and money center banks. *Journal of Financial Intermediation* **23** 322–347.

[27] CSARDI, G. (2021). igraphdata R package version 1.0.1.

[28] CSERMELY, P., LONDON, A., WU, L.-Y. and UZZI, B. (2013). Structure and dynamics of core/periphery networks. *Journal of Complex Networks* **1** 93–123.

[29] CUCURINGU, M., ROMBACH, P., LEE, S. H. and PORTER, M. A. (2016). Detection of core–periphery structure in networks using spectral methods and geodesic paths. *European Journal of Applied Mathematics* **27** 846–887. MR3570008

[30] DA SILVA, M. R., MA, H. and ZENG, A.-P. (2008). Centrality, network capacity, and modularity as parameters to analyze the core-periphery structure in metabolic networks. *Proceedings of the IEEE* **96** 1411–1420.

[31] DASGUPTA, A. and SENGUPTA, S. (2022). Scalable estimation of epidemic thresholds via node sampling. *Sankhya A* **84** 321–344. MR4402754

[32] DE JEUDE, J. V. L., CALDARELLI, G. and SQUARTINI, T. (2019). Detecting core-periphery structures by surprise. *EPL (Europhysics Letters)* **125** 68001.

[33] DE MEO, P., FERRARA, E., FIUMARA, G. and PROVETTI, A. (2011). Generalized Louvain method for community detection in large networks. In *2011 11th International Conference on Intelligent Systems Design and Applications* 88–93.

[34] DEMPSTER, A. P., LAIRD, N. M. and RUBIN, D. B. (1977). Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **39** 1–22. MR0501537

[35] DONG, G., FAN, J., SHEKHTMAN, L. M., SHAI, S., DU, R., TIAN, L., CHEN, X., STANLEY, H. E. and HAVLIN, S. (2018). Resilience of networks with community structure behaves as if under an external field. *Proceedings of the National Academy of Sciences* **115** 6911–6915.

[36] DOREIAN, P. (1985). Structural equivalence in a psychology journal network. *Journal of the American Society for Information Science* **36** 411–417.

[37] DOWSLAND, K. A. and THOMPSON, J. (2012). Simulated annealing. *Handbook of Natural Computing* 1623–1655.

[38] DUBOIS, C. L. (2008). UCI Network Data Repository.

[39] ELLIOTT, A., CHIU, A., BAZZI, M., REINERT, G. and CUCURINGU, M. (2020). Core–periphery structure in directed networks. *Proceedings of the Royal Society A* **476** 20190783. MR4172561

[40] ERDÖS, P. and RENYI, A. (1959). On random graphs. *Publicationes Mathematicae Debrecen* 260–297. MR0120167

[41] FORTUNATO, S. (2010). Community detection in graphs. *Physics Reports* **486** 75–174. MR2580414

[42] FRICKE, D. and LUX, T. (2015). Core–periphery structure in the overnight money market: evidence from the e-mid trading platform. *Computational Economics* **45** 359–395.

[43] GALLAGHER, R. J., YOUNG, J.-G. and WELLES, B. F. (2021). A clarified typology of core-periphery structure in networks. *Science Advances* **7** eabc9800.

[44] GEYER, C. J. (1991). Markov chain Monte Carlo maximum likelihood. Interface Foundation of North America.

[45] GHOSHAL, G., ZLATIĆ, V., CALDARELLI, G. and NEWMAN, M. E. (2009). Random hypergraphs and their applications. *Physical Review E* **79** 066118. MR2551286

[46] GIRVAN, M. and NEWMAN, M. E. J. (2002). Community structure in social and biological networks. *Proceedings of the National Academy of Sciences* **99** 7821–7826. MR1908073

[47] GLOVER, F. (1986). Future paths for integer programming and links to artificial intelligence. *Computers & Operations Research* **13** 533–549. MR0868908

[48] GLOVER, F. and LAGUNA, M. (1998). Tabu search. In *Handbook of Combinatorial Optimization* 2093–2229. Springer. MR1665424

[49] GRIFFIN, J. E. and LEISEN, F. (2017). Compound random measures and their use in Bayesian non-parametrics. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **79** 525–545. MR3611758

[50] GUIMERA, R. and AMARAL, L. A. N. (2004). Modeling the world-wide airport network. *The European Physical Journal B* **38** 381–385.

[51] GUO, Z., CHO, J.-H., CHEN, R., SENGUPTA, S., HONG, M. and MITRA, T. (2020). Online social deception and its countermeasures: A survey. *IEEE Access* **9** 1770–1806.

[52] HANDCOCK, M. S., RAFTERY, A. E. and TANTRUM, J. M. (2007). Model-based clustering for social networks. *Journal of the Royal Statistical Society: Series A (Statistics in Society)* **170** 301–354. MR2364300

[53] HOFF, P. D., RAFTERY, A. E. and HANDCOCK, M. S. (2002). Latent space approaches to social network analysis. *Journal of the American Statistical Association* **97** 1090–1098. MR1951262

[54] HOLLAND, P. W., LASKEY, K. B. and LEINHARDT, S. (1983). Stochastic Block Models: First steps. *Social Networks* **5** 109–137. MR0718088

[55] HOLME, P. (2005). Core-periphery organization of complex networks. *Physical Review E* **72** 046111.

[56] HOLME, P. (2015). Modern temporal network theory: a colloquium. *The European Physical Journal B* **88** 1–30.

[57] HOLME, P. and SARAMÄKI, J. (2012). Temporal networks. *Physics Reports* **519** 97–125.

[58] JESKE, D. R., STEVENS, N. T., TARTAKOVSKY, A. G. and WILSON, J. D. (2018). Statistical methods for network surveillance. *Applied Stochastic Models in Business and Industry* **34** 425–445. MR3845890

[59] JIA, J. and BENSON, A. R. (2019). Random Spatial Network Models for Core-Periphery Structure. *WSDM'19* 366–374. Association for Computing Machinery, New York, NY, USA.

[60] KANE, G. C., ALAVI, M., LABIANCA, G. and BORGATTI, S. P. (2014). What's different about social media networks? A framework and research agenda. *MIS Quarterly* **38** 275–304.

[61] KARRER, B. and NEWMAN, M. E. J. (2011). Stochastic blockmodels and community structure in networks. *Physical Review E* **83** 016107. MR2788206

[62] KERNIGHAN, B. W. and LIN, S. (1970). An efficient heuristic procedure for partitioning graphs. *The Bell System Technical Journal* **49** 291–307.

[63] KINGMA, D. P. and WELLING, M. (2013). Auto-encoding variational bayes. *arXiv preprint arXiv:1312.6114*.

[64] KINGMAN, J. (1967). Completely random measures. *Pacific Journal of Mathematics* **21** 59–78. MR0210185

[65] KIRKPATRICK, S., GELATT, C. D. and VECCHI, M. P. (1983). Optimization by simulated annealing. *Science* **220** 671–680. MR0702485

[66] KIVELÄ, M., ARENAS, A., BARTHELEMY, M., GLEESON, J. P., MORENO, Y. and PORTER, M. A. (2014). Multilayer networks. *Journal*

of complex networks **2** 203–271.

[67] Kodali, L., Sengupta, S., House, L. and Woodall, W. H. (2020). The Value of Summary Statistics for Anomaly Detection in Temporally-Evolving Networks: A Performance Evaluation Study. *Applied Stochastic Models in Business and Industry* **2020** 1–34. MR4201128

[68] Kojaku, S. and Masuda, N. (2017). Finding multiple core-periphery pairs in networks. *Physical Review E* **96** 052313.

[69] Kojaku, S. and Masuda, N. (2018). Core-periphery structure requires something else in the network. *New Journal of Physics* **20** 043012.

[70] Kostoska, O., Mitikj, S., Jovanovski, P. and Kocarev, L. (2020). Core-periphery structure in sectoral international trade networks: A new approach to an old theory. *PloS One* **15** e0229547.

[71] Krugman, P. (1996). *The Self Organizing Economy.* John Wiley & Sons.

[72] Lehmann, S., Lautrup, B. and Jackson, A. D. (2003). Citation networks in high energy physics. *Physical Review E* **68** 026113. MR2561122

[73] Lei, J. and Rinaldo, A. (2015). Consistency of Spectral Clustering in Sparse Stochastic Block Models. *Annals of Statistics* **43** 215–237. MR3285605

[74] Leskovec, J. and Krevl, A. (2020). SNAP Datasets: Stanford Large Network Dataset Collection. http://snap.stanford.edu/data.

[75] Li, W. and Cai, X. (2004). Statistical analysis of airport network of China. *Physical Review E* **69** 046106.

[76] Lip, S. Z. (2011). A fast algorithm for the discrete core/periphery bipartitioning problem. *arXiv preprint arXiv:1102.5511.*

[77] Lordan, O. and Sallan, J. M. (2017). Analyzing the multilevel structure of the European airport network. *Chinese Journal of Aeronautics* **30** 554–560.

[78] Lordan, O. and Sallan, J. M. (2019). Core and critical cities of global region airport networks. *Physica A: Statistical Mechanics and its Applications* **513** 724–733.

[79] Lovekar, K., Sengupta, S. and Paul, S. (2021). Testing for the Network Small-World Property. *arXiv preprint arXiv:2103.08035.*

[80] Ma, H.-W. and Zeng, A.-P. (2003). The connectivity structure, giant strong component and centrality of metabolic networks. *Bioinformatics* **19** 1423–1430.

[81] Magone, J. M., Laffan, B. and Schweiger, C. (2016). *Core-periphery relations in the European Union: Power and conflict in a dualist political economy.* Routledge.

[82] Marchese, E., Caldarelli, G. and Squartini, T. (2022). Detecting mesoscale structures by surprise. *Communications Physics* **5** 132.

[83] McPherson, M., Smith-Lovin, L. and Cook, J. M. (2001). Birds of a feather: Homophily in social networks. *Annual Review of Sociology* **27** 415–444.

[84] Michailidis, G. (2012). Statistical challenges in biological networks. *Journal of Computational and Graphical Statistics* **21** 840–855. MR3005800

[85] MORGAN, D. L., NEAL, M. B. and CARDER, P. (1997). The stability of core and peripheral networks over time. *Social Networks* **19** 9–25.

[86] MØRUP, M. and SCHMIDT, M. N. (2012). Bayesian community detection. *Neural Computation* **24** 2434–2456. MR2986776

[87] MUKHERJEE, S. S., SARKAR, P. and BICKEL, P. J. (2021). Two provably consistent divide-and-conquer clustering algorithms for large networks. *Proceedings of the National Academy of Sciences* **118** e2100482118. MR4390029

[88] MULLINS, N. C., HARGENS, L. L., HECHT, P. K. and KICK, E. L. (1977). The group structure of cocitation clusters: A comparative study. *American Sociological Review* 552–562.

[89] NAIK, C., CARON, F. and ROUSSEAU, J. (2021). Sparse networks with core-periphery structure. *Electronic Journal of Statistics* **15** 1814–1868. MR4255305

[90] NEWMAN, M. (2018). *Networks.* Oxford university press. MR3838417

[91] NEWMAN, M. E. J. (2006). Finding community structure in networks using the eigenvectors of matrices. *Physical Review* **74**. MR2282139

[92] NEWMAN, M. E. J. and GIRVAN, M. (2004). Finding and evaluating community structure in networks. *Physical Review E* **69**. MR2282139

[93] NG, A. Y., JORDAN, M. I. and WEISS, Y. (2002). On spectral clustering: Analysis and an algorithm. In *Advances in neural information processing systems* 849–856.

[94] NOROOZI, M. and PENSKY, M. (2020). Statistical inference in heterogeneous block model. *stat* **1050** 7.

[95] NOROOZI, M. and PENSKY, M. (2021). The Hierarchy of Block Models. *Sankhya A* **84** 64–107. MR4402747

[96] OUVRARD, X. (2020). Hypergraphs: an introduction and review. *arXiv preprint arXiv:2002.05014*.

[97] PRIEBE, C. E., CONROY, J. M., MARCHETTE, D. J. and PARK, Y. (2005). Scan statistics on Enron graphs. *Computational & Mathematical Organization Theory* **11** 229–247.

[98] PSORAKIS, I., ROBERTS, S., EBDEN, M. and SHELDON, B. (2011). Overlapping community detection using bayesian non-negative matrix factorization. *Physical Review E* **83** 066114.

[99] PUJOL, J. M., ERRAMILLI, V. and RODRIGUEZ, P. (2009). Divide and conquer: Partitioning online social networks. *arXiv preprint arXiv:0905.4918*.

[100] RADICCHI, F., FORTUNATO, S. and CASTELLANO, C. (2008). Universality of citation distributions: Toward an objective measure of scientific impact. *Proceedings of the National Academy of Sciences* **105** 17268–17272.

[101] RIOLO, M. A. and NEWMAN, M. (2020). Consistency of community structure in complex networks. *Physical Review E* **101** 052306. MR4115264

[102] ROMBACH, P., PORTER, M. A., FOWLER, J. H. and MUCHA, P. J. (2017). Core-Periphery Structure in Networks (Revisited). *SIAM Review* **59** 619–646. MR3683683

[103] ROSSA, F. D., DERCOLE, F. and PICCARDI, C. (2013). Profiling core-periphery network structure by random walkers. *Scientific Reports* **3** 1467.

[104] ROSSI, R. and AHMED, N. (2015). The network data repository with interactive graph analytics and visualization. In *Proceedings of the AAAI Conference on Artificial Intelligence* **29**.

[105] ROZEMBERCZKI, B., ALLEN, C. and SARKAR, R. (2021). Multi-scale attributed node embedding. *Journal of Complex Networks* **9** 1–22. MR4256683

[106] RUAN, Y., FUHRY, D. and PARTHASARATHY, S. (2013). Efficient Community Detection in Large Networks Using Content and Links. In *Proceedings of the 22nd International Conference on World Wide Web. WWW'13* 1089–1098. Association for Computing Machinery, New York, NY, USA.

[107] SABIDUSSI, G. (1966). The centrality index of a graph. *Psychometrika* **31** 581–603. MR0205879

[108] SASTRY, K., GOLDBERG, D. and KENDALL, G. (2005). Genetic algorithms. In *Search Methodologies* 97–125. Springer.

[109] SCHOCH, D. (2021). networkdata: Repository of Network Datasets R package version 0.1.9.

[110] SCOTT, J. (1988). Social network analysis. *Sociology* **22** 109–127.

[111] SEDITA, S. R., CALOFFI, A. and LAZZERETTI, L. (2020). The invisible college of cluster research: a bibliometric core–periphery analysis of the literature. *Industry and Innovation* **27** 562–584.

[112] SENGUPTA, S. and CHEN, Y. (2015). Spectral clustering in heterogeneous networks. *Statistica Sinica* **25** 1081–1106. MR3410299

[113] SENGUPTA, S. and CHEN, Y. (2018). A block model for node popularity in networks with community structure. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **80** 365–386. MR3763696

[114] SHEN, X., HAN, Y., LI, W., WONG, K.-C. and PENG, C. (2021). Finding core-periphery structures in large networks. *Physica A: Statistical Mechanics and its Applications* **581** 126224. MR4283261

[115] SHI, C., LI, Y., ZHANG, J., SUN, Y. and PHILIP, S. Y. (2016). A survey of heterogeneous information network analysis. *IEEE Transactions on Knowledge and Data Engineering* **29** 17–37.

[116] SNIJDERS, T. A. and NOWICKI, K. (1997). Estimation and prediction for stochastic blockmodels for graphs with latent block structure. *Journal of Classification* **14** 75–100. MR1449742

[117] SONG, G., ZHOU, X., WANG, Y. and XIE, K. (2014). Influence maximization on large-scale mobile social network: a divide-and-conquer method. *IEEE Transactions on Parallel and Distributed Systems* **26** 1379–1392.

[118] SUPPER, J., SPANGENBERG, L., PLANATSCHER, H., DRÄGER, A., SCHRÖDER, A. and ZELL, A. (2009). BowTieBuilder: modeling signal transduction pathways. *BMC Systems Biology* **3** 1–13.

[119] TAN, L. S. and FRIEL, N. (2020). Bayesian variational inference for exponential random graph models. *Journal of Computational and Graphical Statistics* **29** 910–928. MR4191251

[120] Tang, W., Zhao, L., Liu, W., Liu, Y. and Yan, B. (2019). Recent advance on detecting core-periphery structure: a survey. *CCF Transactions on Pervasive Computing and Interaction* **1** 175–189.

[121] Thiemichen, S., Friel, N., Caimo, A. and Kauermann, G. (2016). Bayesian exponential random graph models with nodal random effects. *Social Networks* **46** 11–28.

[122] Tickner, A. B. (2013). Core, periphery and (neo) imperialist International Relations. *European Journal of International Relations* **19** 627–646.

[123] Tunç, B. and Verma, R. (2015). Unifying inference of meso-scale structures in networks. *PloS One* **10** e0143133.

[124] van der Pas, S. and van der Vaart, A. (2018). Bayesian community detection. *Bayesian Analysis* **13** 767–796. MR3807866

[125] Van Laarhoven, P. J. and Aarts, E. H. (1987). Simulated annealing. In *Simulated Annealing: Theory and Applications* 7–15. Springer. MR0904050

[126] Von Luxburg, U., Belkin, M. and Bousquet, O. (2008). Consistency of spectral clustering. *The Annals of Statistics* 555–586. MR2396807

[127] Wang, X., Ji, H., Shi, C., Wang, B., Ye, Y., Cui, P. and Yu, P. S. (2019). Heterogeneous graph attention network. In *The World Wide Web Conference* 2022–2032.

[128] Wasserman, S. and Faust, K. (1994). Social network analysis: Methods and applications.

[129] Wedell, E., Park, M., Korobskiy, D., Warnow, T. and Chacko, G. (2022). Center–periphery structure in research communities. *Quantitative Science Studies* **3** 289–314.

[130] Yanchenko, E. and Sengupta, S. (2021). A generalized hypothesis test for community structure and homophily in networks. *arXiv preprint arXiv: 2107.06093*.

[131] Yang, B., Huang, X., Hu, X., Cheng, W., Pei, Z. and Li, X. (2021). Optimizing Robustness of Core-Periphery Structure in Complex Networks. *IEEE Transactions on Circuits and Systems II: Express Briefs* **68** 3572–3576.

[132] Yang, J. and Leskovec, J. (2014). Structure and overlaps of ground-truth communities in networks. *ACM Transactions on Intelligent Systems and Technology (TIST)* **5** 1–35.

[133] Yang, J. and Leskovec, J. (2015). Defining and evaluating network communities based on ground-truth. *Knowledge and Information Systems* **42** 181–213.

[134] Yang, J., Zhang, M., Shen, K. N., Ju, X. and Guo, X. (2018). Structural correlation between communities and core-periphery structures in social networks: Evidence from Twitter data. *Expert Systems with Applications* **111** 91–99.

[135] Zelnio, R. (2012). Identifying the global core-periphery structure of science. *Scientometrics* **91** 601–615.

[136] Zeng, J. and Yu, H. (2015). Parallel modularity-based community detection on large-scale graphs. In *2015 IEEE International Conference on*

Cluster Computing 1–10. IEEE.

[137] Zhang, X., Martin, T. and Newman, M. E. J. (2015). Identification of core-periphery structure in networks. *Physical Review E* **91** 032803.

[138] Zhao, Y., Levina, E. and Zhu, J. (2012). Consistency of community detection in networks under degree-corrected stochastic block models. *The Annals of Statistics* **40** 2266–2292. MR3059083

[139] Zhou, S. and Mondragón, R. J. (2004). The rich-club phenomenon in the Internet topology. *IEEE Communications Letters* **8** 180–182.