

## NETWORK SCIENCE

## A clarified typology of core-periphery structure in networks

Ryan J. Gallagher<sup>1\*</sup>, Jean-Gabriel Young<sup>2,3</sup>, Brooke Foucault Welles<sup>1,4</sup>

Core-periphery structure, the arrangement of a network into a dense core and sparse periphery, is a versatile descriptor of various social, biological, and technological networks. In practice, different core-periphery algorithms are often applied interchangeably despite the fact that they can yield inconsistent descriptions of core-periphery structure. For example, two of the most widely used algorithms, the  $k$ -cores decomposition and the classic two-block model of Borgatti and Everett, extract fundamentally different structures: The latter partitions a network into a binary hub-and-spoke layout, while the former divides it into a layered hierarchy. We introduce a core-periphery typology to clarify these differences, along with Bayesian stochastic block modeling techniques to classify networks in accordance with this typology. Empirically, we find a rich diversity of core-periphery structure among networks. Through a detailed case study, we demonstrate the importance of acknowledging this diversity and situating networks within the core-periphery typology when conducting domain-specific analyses.

## INTRODUCTION

Core-periphery structure is a fundamental network pattern, referring to the presence of two qualitatively distinct components: a dense “core” of tightly connected nodes and a sparse “periphery” of nodes loosely connected to the core and among each other. This pattern has helped explain a broad range of networked phenomena, including online amplification (1), cognitive learning processes (2), technological infrastructure organization (3, 4), and critical disease-spreading conduits (5). It applies so seamlessly across domains because it provides a succinct mesoscale description of a network’s organization around its core. By decomposing a network into core and peripheral nodes, core-periphery structure separates central processes from those on the margin, allowing us to more precisely classify the functional and dynamical roles of nodes with respect to their structural position. The analytic generality of this approach, together with the relative ubiquity of core-periphery structure among networks, makes core-periphery structure an indispensable methodological concept in the network science inventory.

Several methods and algorithms exist for extracting core-periphery structure from networks (6). They take on a variety of mathematical and algorithmic forms, ranging from statistical inference (7–9), spectral decomposition (10, 11), and diffusion mapping (12) to motif counting (13), geodesic tracing (10), and model averaging (14). These algorithms exhibit a creative diversity of approaches for extracting core-periphery structure, and each is motivated by imagery of how core and peripheral nodes connect to one another. However, underneath the different high-level descriptions of each model, there are varying and often inconsistent assumptions about how the core and periphery are mutually connected and how core-periphery structure is reflected in a network. As a result, despite the importance of core-periphery decomposition in answering substantive domain questions outside of network science, practitioners looking to apply these methods to their own fields are left without a warning that

each algorithm has a different vision of what “core-periphery structure” actually means. This threatens the ability of researchers to draw valid conclusions about the structure and dynamics of numerous networks. By introducing a core-periphery typology that distinguishes between two qualitatively and quantitatively distinct structures and by providing statistical techniques for determining where networks fall within that typology, we intend to make the distinction between various core-periphery structures clear and enable reliable network inferences by scholars and practitioners.

The two types of core-periphery characterizations in our typology are well exemplified by two of the most popular approaches for identifying core-periphery structure in networks. The first, which we refer to as the “two-block model,” is rooted in a definition originally proposed by Borgatti and Everett (15). Their mathematical formulation of core-periphery structure proposes that nodes are arranged into two groups, the core and the periphery, such that “core nodes are adjacent to other core nodes, core nodes are adjacent to some periphery nodes, and periphery nodes do not connect with other periphery nodes” (15). This paints a hub-and-spoke picture of core-periphery structure: There is a central hub of interwoven nodes and a periphery that radiates out from that hub. The hub-and-spoke core-periphery formulation is at the backbone of network science methodology because it underlies many of the more sophisticated models that have been developed since Borgatti and Everett’s foundational work (7, 9, 14). Furthermore, because the two-block formulation was originally proposed in the language of block models (16), it is often the de facto statistical definition of core-periphery structure for many network scientists.

The second core-periphery characterization is reflected in the widely used  $k$ -cores decomposition. The  $k$ -core of a network is the largest subset of nodes in the network such that every node has at least  $k$  connections to other nodes in that subset (17). The  $k$ -cores define a hierarchy of  $k$ -shells, each of which consists of all the nodes in the  $k$ -core but not the  $(k + 1)$ -core. A decomposition in terms of the  $k$ -cores highlights a network’s core-periphery structure and is typically obtained by iteratively removing the  $k$ -shells (18), starting with peripheral low-degree nodes in the outer shells and working toward embedded high-degree nodes in the inner cores. This algorithmic pruning process is accompanied by a suite of evocative

Copyright © 2021  
The Authors, some  
rights reserved;  
exclusive licensee  
American Association  
for the Advancement  
of Science. No claim to  
original U.S. Government  
Works. Distributed  
under a Creative  
Commons Attribution  
NonCommercial  
License 4.0 (CC BY-NC).

Downloaded from https://www.science.org at Seoul National University Library on August 18, 2024

<sup>1</sup>Network Science Institute, Northeastern University, Boston, MA 02115, USA. <sup>2</sup>Center for the Study of Complex Systems, University of Michigan, Ann Arbor, MI 48109, USA.

<sup>3</sup>Department of Computer Science, University of Vermont, Burlington, VT 05405, USA.

<sup>4</sup>Department of Communication Studies, Northeastern University, Boston, MA 02115, USA.

\*Corresponding author. Email: gallagher.r@northeastern.edu

language: The periphery is described as a series of “shells” (17), “onion layers” (19), or “leaves” (20), while the core is referred to as the “epicenter” (1), “corona” (21), or “nucleus” (4). The language of the  $k$ -cores decomposition conjures up an image of a layered core-periphery structure composed of a nested sequence of layers that funnel toward a core. The scalable algorithm of the  $k$ -cores decomposition (20) has made it a practical tool for studying networks of all sizes, meaning that a number of applied network analyses implicitly assume a layered network arrangement.

By these accounts, it is clear that the layered and hub-and-spoke characterizations provide distinct descriptions of core-periphery structure. The differences between the hub-and-spoke and layered core-periphery characterizations are more than a linguistic sleight of hand though; they can have repercussive consequences for substantive network analyses. In what follows, we first show that the structures extracted by the two most widely used algorithms, the two-block model and the  $k$ -cores decomposition, diverge quantitatively across many empirical networks. To establish a statistically principled way of comparing these two classes of models, we then formulate both the hub-and-spoke and layered core-periphery structures as stochastic block models (22), which allow us to encode the qualitative differences between the two characterizations and formulate an information-theoretic criterion of model fit (8, 23). With these tools, we analyze a suite of empirical networks and find a rich diversity of core-periphery structure that spans across the core-periphery typology. We finish with a case study of hashtag activism amplification and emphasize how the choice of core-periphery model critically affects the interpretation of substantive results. Our typology clarifies the distinct core-periphery structures that can emerge in networks and provides a methodologically sound approach for disentangling those structures in practice.

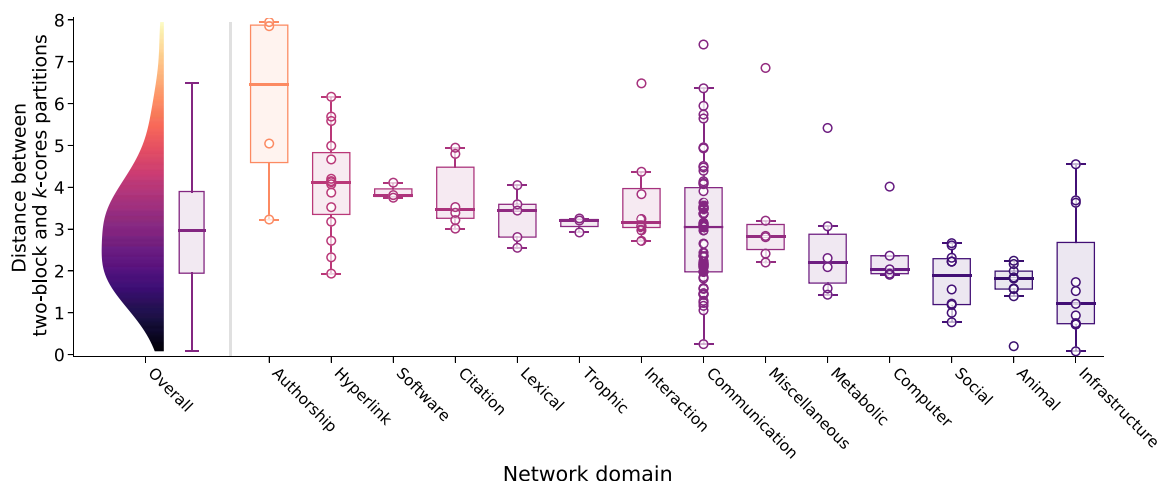
## RESULTS

### Inconsistent core-periphery partitions

We start by showing that the hub-and-spoke structure explicitly extracted by the two-block model (15) and the layered structure

implicitly suggested by the  $k$ -cores decomposition (17) provide fundamentally different descriptions of core-periphery structure for the same networks. To this end, we draw upon the Koblenz Network Collection (KONECT) (24), a diverse network repository that spans a number of social, biological, and technological domains. For each KONECT network, we extract the binary partition of nodes according to the two-block model (25) and the nested partition of nodes according to the  $k$ -cores decomposition (20). We then measure the distance between these partitions via the variation of information (VI) (26, 27) and present the pairwise comparisons in Fig. 1. The VI is an information-theoretic measure and is therefore expressed in bits per nodes. Intuitively, it can be thought of as the sum of information not shared by the two partitions. Hence, the more distant or dissimilar two partitions are, the larger the VI.

Across network domains, we find that the core-periphery partitions identified by the two-block model and  $k$ -cores decomposition are quite dissimilar, with an overall median VI of 2.9 bits per nodes. A normalized version of the VI, which can only be consistently interpreted for individual networks and not across datasets (26), yields a median of 35% the maximal value across domains. In other words, for each individual network, the partitions are about a third as distant as possible. Furthermore, the differences in outcomes are not exhibited uniformly across domains. Some classes of networks (e.g., social, animal, and infrastructure networks) see relatively more agreement between the two-block and  $k$ -cores partitions. For other network types (e.g., authorship, hyperlink, and software networks), however, the two core-periphery algorithms almost always extract distant structures. Even within domains, there can be a wide heterogeneity in the similarities: For example, the range of distances is 7.2 bits for communication networks, 4.9 bits for infrastructure networks, and 4.2 bits for hyperlink networks. We find even less agreement between algorithms when we use other measures to compare partitions, such as the adjusted (28) or reduced (29) mutual information. We are able to somewhat improve the agreement by matching the partitions on sizes, to correct for the discrepancy between the number of  $k$ -cores and the number of groups in the two-block structure, but nonetheless, the partitions disagree in general (see the Supplementary Materials).



**Fig. 1. Distributions of distances between partitions extracted by the two-block model (15) and the  $k$ -cores decomposition (18) across network domains.** The two-block and  $k$ -cores partitions are dissimilar across a variety of networks and network types, indicating that they generally extract different core-periphery structures. Distance is measured by the VI (26), where higher values indicate more dissimilar partitions. Thick lines in each domain's box plot indicate the median difference and are also represented as the color of each box. The distributions furthest to the left indicate the kernel density estimate and box plot of the distance distribution across all networks. Detailed results are reported in the Supplementary Materials.

The results of Fig. 1 do not imply that one algorithm or the other is intrinsically flawed but, rather, that the algorithms do not agree in general. If one's goal is only to describe a network, then this disagreement is not an issue because each algorithm simply provides its own description of the network, whatever that may be (30). However, there is a strict statistical sense in which the algorithms cannot both equally well-characterize a given network: Each core-periphery partition corresponds to a statistical description of the network, one of which will necessarily be more concise and precise than the other (31, 32). So, if we want to make network inferences based on core-periphery structure, then we need to call on methods that can identify models that give better statistical descriptions of a network's structure than others. Selecting models, however, requires that we have statistical models in the first place, and the notions of core-periphery structure that we have applied so far have only been defined through algorithms. We therefore turn to Bayesian stochastic block models (8) to establish the missing connection between the two.

### Core-periphery stochastic block models

The stochastic block model is a general statistical model of a network's mesoscale structure (22). At its core, it assumes that nodes belong to different groups, or "blocks," such as the core and the periphery. These blocks then specify the probability that any two nodes are connected. More formally, suppose that we have the adjacency matrix  $\mathbf{A}$  of an unweighted, undirected, simple network with  $N$  nodes. We assume that there are a fixed number of  $B$  blocks, and let the block assignments of all the nodes be recorded in  $\theta$ , a vector of length  $N$  where  $\theta_i = r$  indicates that node  $i$  belongs to block  $r$ . The probability that any two nodes in the network are connected is given by  $\mathbf{p}$ , a  $B \times B$  matrix where  $p_{rs}$  is the probability that a node in block  $r$  connects to a node in block  $s$ . This is the defining characteristic of the stochastic block model: The block assignments completely determine the probability of connection between any two nodes.

In practice, we do not know the block assignments of the nodes  $\theta$  or the probability of connection between blocks  $\mathbf{p}$ . We are interested, then, in the distribution  $P(\theta, \mathbf{p} | \mathbf{A})$ , the probability that we have a particular arrangement of nodes into blocks and connections between them, given our observed network data  $\mathbf{A}$ . Applying Bayes' rule, we have

$$P(\theta, \mathbf{p} | \mathbf{A}) \propto P(\mathbf{A} | \theta, \mathbf{p}) P(\theta) P(\mathbf{p}) \quad (1)$$

The posterior distribution  $P(\theta, \mathbf{p} | \mathbf{A})$  is proportional to three components: the likelihood  $P(\mathbf{A} | \theta, \mathbf{p})$  of the network  $\mathbf{A}$ , the prior on the block assignments  $P(\theta)$ , and the prior on the block

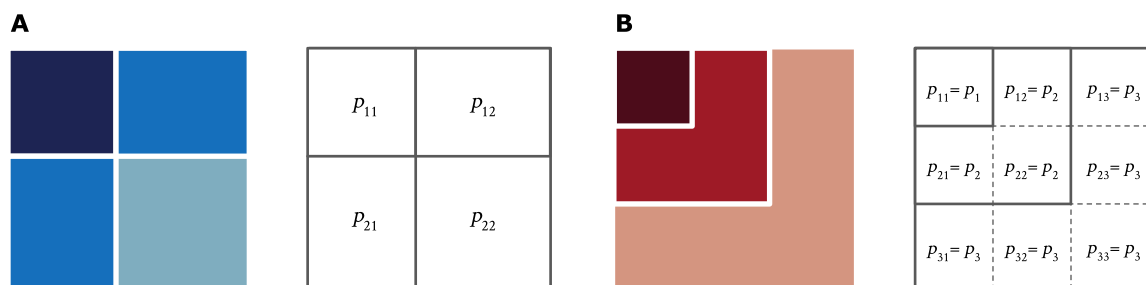
connectivity matrix  $P(\mathbf{p})$ . We outline the standard setup of the likelihood and block assignment prior in Materials and Methods. For constructing core-periphery stochastic block models, our main concern is with the prior on the block connectivity matrix. When we know that we want to model core-periphery structure specifically, we only want to consider particular arrangements of connection probabilities  $\mathbf{p}$  and that prior knowledge should be reflected in  $P(\mathbf{p})$ . This is a different view than is usually taken for block models: Rather than assuming nothing about a network's structure and applying a general, unconstrained block model, we intentionally seek and encode core-periphery structure in our models.

We propose a core-periphery typology that contains two structures: the hub-and-spoke structure and the layered structure. Both characterizations can be phrased in the language of block models by arranging the block connectivities of  $\mathbf{p}$  in different ways, as depicted in Fig. 2. Through the Bayesian approach to the stochastic block modeling, we can alter the prior  $P(\mathbf{p})$  to encode these different arrangements and constrain the model to adhere to those structures (30). The constrained models allow us to only consider networks with respect to the core-periphery typology and classify them appropriately according to the structure that they exhibit. From here onward, we use "hub-and-spoke" to refer to either the theoretical typology characterization or the stochastic block model implementation, depending on context. However, we only use "two-block model" or "two-block algorithm" when specifically referencing Borgatti and Everett's implementation (15). Similarly, we use "layered" to refer to the typology characterization or the block model but only "k-cores" to refer to the heuristic algorithm.

The hub-and-spoke characterization specifies two blocks, one for the core and one for the periphery. If we let the core be denoted by the first block and the periphery by the second, then the original two-block model presented by Borgatti and Everett (15) can be recovered by setting  $p_{11} = 1$ ,  $p_{12} = 1$ , and  $p_{22} = 0$ . We consider a relaxation of this structure (7), which allows for flexibility in the connections by only requiring  $p_{11} > p_{12} > p_{22}$ . This configuration (shown in Fig. 2A) conveys the intuition of the hub-and-spoke structure: There is a densely connected core moderately connected with a periphery that is only loosely connected among itself. Statistically, we enforce this constraint through a uniform prior over all block matrices that satisfy  $0 < p_{22} < p_{12} < p_{11} < 1$ . In notation, we write

$$P(\mathbf{p}) \propto \mathbb{1}_{\{0 < p_{22} < p_{12} < p_{11} < 1\}} \quad (2)$$

where  $\mathbb{1}$  is the indicator function that takes on the value 1 if the constraint is satisfied and 0 otherwise.



**Fig. 2. The core-periphery typology formalized through block model representations of the hub-and-spoke and layered structures.** Each figure depicts the block connectivity matrix  $\mathbf{p}$ , where darker colors indicate higher densities of links. (A) The hub-and-spoke model is defined according to two blocks, where  $p_{11} > p_{12} > p_{22}$ . (B) The layered core-periphery model is defined according to  $\ell$  layers, which are ordered as  $p_1 > p_2 > \dots > p_\ell$ .

We can similarly formulate the layered block model. For convenience, we let  $p_{rr} = p_r$  and assume that there are  $\ell$  layers, equal to the number of blocks  $B$ . To configure the layered structure shown in Fig. 2B, we first specify

$$p_{rs} = p_{\max(r,s)} \quad (3)$$

which binds the matrix  $\mathbf{p}$  into layers. Similar to the hub-and-spoke model, we then order the layers through a uniform prior over all  $\mathbf{p}$  that satisfy  $0 < p_\ell < p_{\ell-1} < \dots < p_1 < 1$

$$P(\mathbf{p}) \propto \mathbb{1}_{\{0 < p_\ell < p_{\ell-1} < \dots < p_1 < 1\}} \quad (4)$$

The layered model can be seen as a special case of the hub-and-spoke model when  $\ell = 2$  and  $p_{12} \approx p_{22}$ . In this block structure, peripheral nodes in the second layer connect to any other node, whether in the core or periphery, with the same probability. For  $\ell \geq 3$ , any node is agnostic to the specific role of nodes that are in cores more central than itself, connecting to them all with the same probability. In this sense, the layered model can be viewed as a sequence of nested and increasingly dense subgraphs.

Recall that we are introducing these priors to determine the type of core-periphery structure that best describes a particular network. In practice, we find these structures by computing the most likely core-periphery position of each node in each model (see Materials and Methods for details). Doing so for all nodes yields both hub-and-spoke and layered partitions. While the hub-and-spoke and layered formulations may seem innocuous modifications of existing stochastic block models, they complicate the computational tractability substantially (8, 16). To infer the distributions of  $\theta$  and  $\mathbf{p}$  for the two models, we therefore have to resort to a Gibbs sampling procedure, detailed in the Supplementary Materials.

We compare the inferred core-periphery structures to select the most statistically appropriate among the two. Formally, for a partition  $\theta_{\mathcal{H}}$  inferred through the hub-and-spoke model  $\mathcal{H}$  and a partition  $\theta_{\mathcal{L}}$  inferred through the layered model  $\mathcal{L}$ , we want to identify which model and its assignment of block labels to nodes is a better fit of the network data  $\mathbf{A}$ . The answer is given by the posterior odds ratio (23)

$$\Lambda = \frac{P(\theta_{\mathcal{H}}, \mathcal{H} | \mathbf{A})}{P(\theta_{\mathcal{L}}, \mathcal{L} | \mathbf{A})} \quad (5)$$

If the posterior odds ratio  $\Lambda > 1$ , then the hub-and-spoke model better characterizes the core-periphery structure of the network, while  $\Lambda < 1$  implies that the layered model is a better descriptor. Assuming that we are agnostic about the models a priori, and so  $P(\mathcal{H}) = P(\mathcal{L}) = 1/2$ , we can equivalently consider

$$-\log \Lambda = \sum_{\mathcal{H}} - \sum_{\mathcal{L}} = -\log P(\mathbf{A}, \theta_{\mathcal{H}} | \mathcal{H}) + \log P(\mathbf{A}, \theta_{\mathcal{L}} | \mathcal{L}) \quad (6)$$

the difference between description lengths of the hub-and-spoke and layered models. The description length,  $\sum_{\mathcal{M}} = -\log P(\mathbf{A}, \theta_{\mathcal{M}} | \mathcal{M})$ , of a model  $\mathcal{M}$  describes how well that model can compress the information expressed by a network's structure (23, 33). A model that is able to efficiently describe a network with a smaller number of parameters is a better descriptor of the network and will have a minimal description length. So, if the hub-and-spoke model  $\mathcal{H}$  is a better descriptor of a network's core-periphery structure than the

layered model  $\mathcal{L}$ , then we will have a posterior odds ratio where  $\Lambda > 1$  and, equivalently, a negative difference in description lengths,  $\sum_{\mathcal{H}} - \sum_{\mathcal{L}} < 0$ . We use the description length to quantify model fit, by considering either the pairwise difference in description lengths between two models or the minimum description length (MDL) across many models (see the Supplementary Materials for numerical details). This measure allows us to distinguish which block model most aptly describes a particular network and properly situate it within the core-periphery typology.

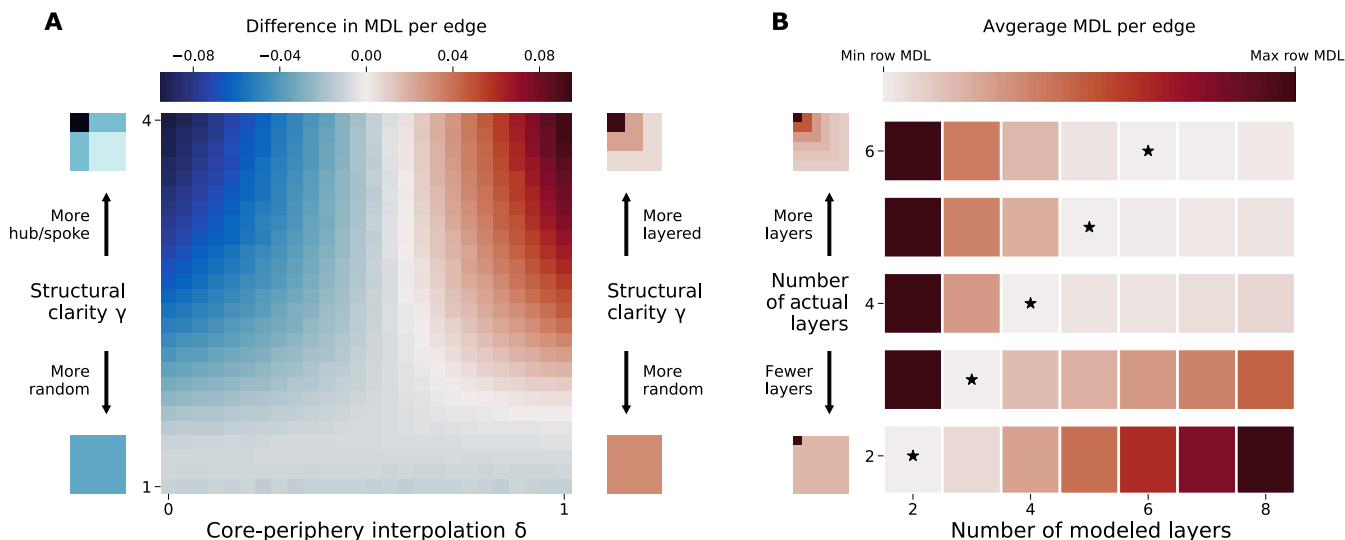
We briefly note connections of our core-periphery block models to prior work. With respect to the hub-and-spoke model, Zhang *et al.* (7) identified the ordering  $p_{11} > p_{12} > p_{22}$  as a relaxed version of the hub-and-spoke block structure introduced by Borgatti and Everett. However, they did not formally encode this constraint in their model and, instead, relied on the susceptibility of stochastic block models to heterogeneous degree distributions (16) to retrieve core-periphery structure. We examine the suitability of this assumption in the Supplementary Materials. With respect to the layered model, Borgatti and Everett (15) presented a special case of our model where  $\ell = 2$ ,  $p_1 = 1$ , and  $p_2 = 0$ . However, given that those binary layer densities imply a network that consists of a connected core component surrounded by a cloud of isolate periphery nodes, they only briefly remarked on the model's limited conceptual utility. Regardless, it is important to emphasize that while the more general layered block model introduced here shares the intuition of nested, increasingly dense layers with the  $k$ -cores decomposition, it is distinct from that algorithm. The layered block model is a fuzzier interpretation of how core-periphery structure can be reflected through layers; it allows for low degree nodes to be placed in core layers if they are embedded among other core nodes, while, by definition, nodes can only be in higher  $k$ -shells if they have a higher degree. In general, then, the block assignments from the layered model will not exactly align with the partition produced by the  $k$ -cores decomposition.

## Synthetic network experiments

As an essential validation step, we experimentally verify that our two models of core-periphery properly recover hub-and-spoke and layered structures when we know that they exist within a network. Our first experiment measures the capacity for the block models to discern between hub-and-spoke and layered structures. We generate synthetic networks according to the stochastic block model and design  $\mathbf{p}$  to have a known, ground truth core-periphery block arrangement. We configure  $\mathbf{p}$  according to a two-parameter model (see Materials and Methods for details). The first parameter  $\delta$  interpolates between hub-and-spoke and layered core-periphery structure: When  $\delta = 0$ , the network has a known hub-and-spoke core-periphery structure, and when  $\delta = 1$ , the network has a known layered structure consisting of three layers. The second parameter  $\gamma$  defines the structural clarity: When  $\gamma = 1$ , the network is random and neither model should be able to infer structure, and for large  $\gamma$ , the networks have a well-defined core-periphery structure.

The results, shown in Fig. 3A, demonstrate that the block models effectively discern the two types of planted core-periphery structures. Within each regime of the interpolation parameter  $\delta$ , the MDL appropriately identifies the correct model as a better description of the network structure, and it is more confident as the parameter reaches the boundaries of its range, at which point only that structure definitively exists in the network. For low values of structural clarity  $\gamma$ , for which the networks are purely random, the models have approximately





**Fig. 3. Synthetic network experiments validating the core-periphery block models and the use of MDL as a measure of model fit.** (A) The difference in MDL per edge between the layered model and hub-and-spoke model on networks constructed to have varying degrees of each structure. Negative values (blue) indicate that the hub-and-spoke model is a better model fit, while positive values (red) indicate that the layered model is a better fit. The MDL accurately discerns which model is the best descriptor of the ground truth network structure, as indicated by the gradual transition from blue to red as the true structure varies from hub-and-spoke to layered. (B) The average MDL per edge of layered models on networks with planted layered structure. For each fixed number of actual layers in the synthetic networks (vertical axis), we run the inference with a varying number of modeled layers  $\ell$  (horizontal axis). Stars (\*) indicate the number of modeled layers  $\ell$  that yields the lowest MDL for a fixed number of planted layers. The alignment of the stars on the diagonal shows that the layered model properly discerns the number of layers used in the generative model.

equal MDLs, and neither is strongly designated better in terms of model fit.

Our second experiment tests the layered model's ability to identify the appropriate number of layers in a synthetic network with a known layered structure. We start with a synthetic network of six equally sized layers and progressively reduce the effective number of layers by merging layers until there are only two layers (see Materials and Methods for details). For each fixed number of layers in the synthetic network, we run multiple layered core-periphery models for different choices of the parameter  $\ell$ , which designates the number of layers to infer. The results, given in Fig. 3B, indicate that the average MDL accurately identifies the number of layers that exist in each synthetic network. This demonstrates that we can not only use the MDL for model selection between the hub-and-spoke and layered models, per the results of the first experiment, but also use it for choosing the number of layers.

### Core-periphery diversity of empirical networks

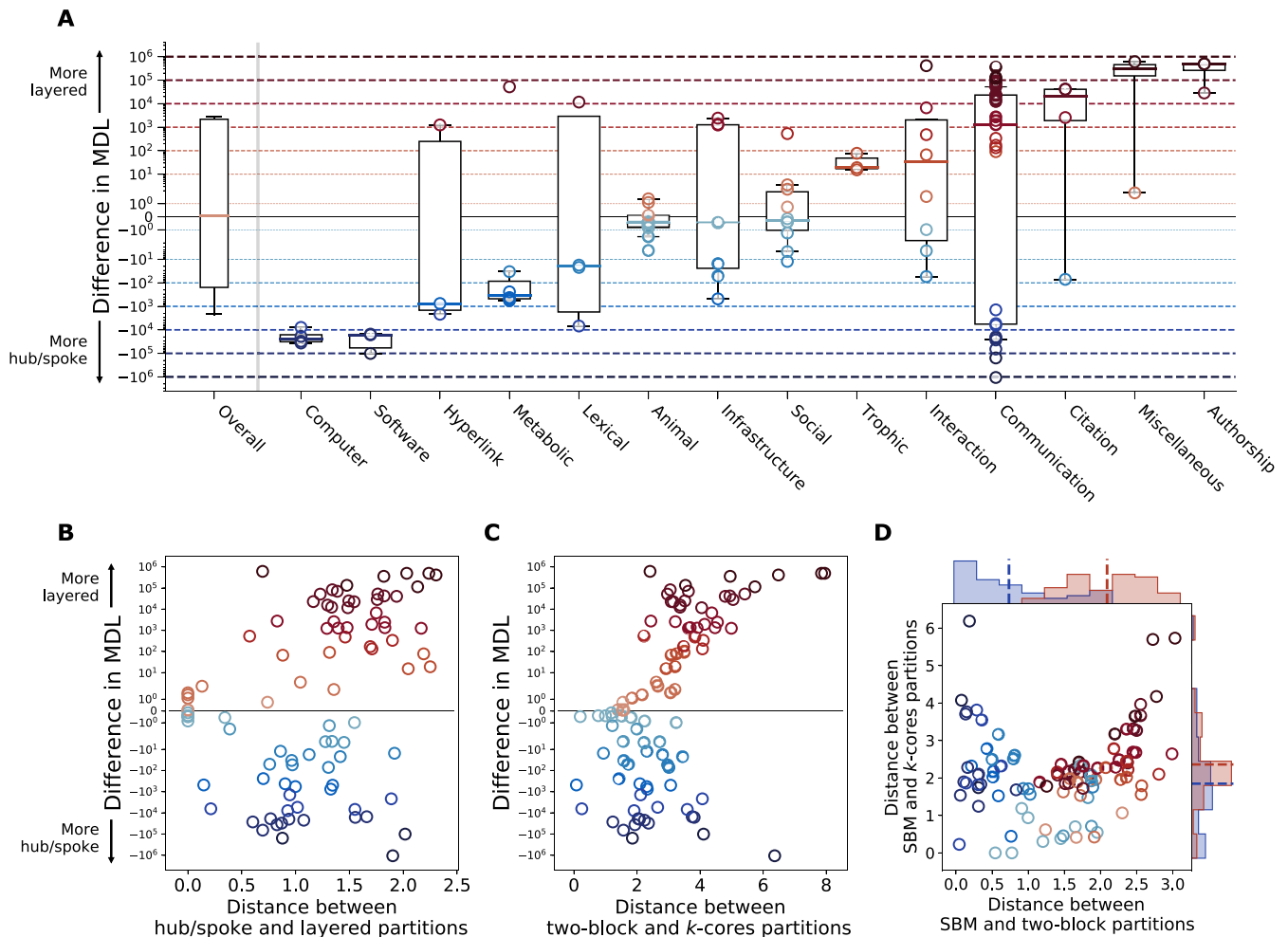
Having validated the core-periphery block models and the use of the MDL as a measure of model fit, we establish the diversity of core-periphery structure expressed by empirical networks. For all networks with up to 200,000 nodes in the KONECT dataset (24), we infer partitions according to each of the hub-and-spoke and layered core-periphery models (see Materials and Methods for details). Recall that we use the terminology hub-and-spoke to refer to the stochastic block model (SBM) implementation and two-block to refer specifically to Borgatti and Everett's original implementation (15). Similarly, layered refers to the SBM implementation, and  $k$ -cores refers to the heuristic algorithm.

In Fig. 4A, we show the breadth of network structure exhibited across the core-periphery typology. As we clearly see, both the hub-and-spoke and layered core-periphery structures are expressed to a

wide degree of intensity across all types of networks; neither model is a universal, best descriptor of core-periphery structure. Some classes of networks seem to be generally well described by either just the hub-and-spoke characterization or just the layered characterization, but many more show a range of structure across the core-periphery spectrum. Communication networks in KONECT, for example, exhibit the full range of core-periphery prevalence across both characterizations. The diversity of core-periphery structure in these empirical networks demonstrates the danger in assuming a core-periphery type a priori and the need to situate a network within the core-periphery typology to mitigate later downstream network mischaracterizations.

We also observe that a smaller portion of networks do not strongly exhibit either a hub-and-spoke or layered structure. In Fig. 4B, we show that these networks are often the ones for which both core-periphery block models extract partitions that are similar or identical. As discussed earlier when deriving the block models, the layered model with  $\ell = 2$  layers is a special case of the hub-and-spoke model. A number of smaller networks in particular, including all of the animal networks, are best modeled with two layers, and so, both models extract similar partitions and have similar MDLs. On the other hand and as expected, we also find that the partitions become less similar as one core-periphery model or the other is preferred according to the MDL.

Last, we connect the core-periphery stochastic block models back to the motivating algorithms, the two-block model and the  $k$ -cores decomposition. In Fig. 4C, we show that networks that are equally well modeled by either the layered or hub-and-spoke block model are also those for which the two-block model and  $k$ -cores decomposition extract similar partitions. Furthermore, the networks that have the most distant two-block and  $k$ -cores core-periphery partitions are also those that mostly strongly exhibit a hub-and-spoke or layered



**Fig. 4. Structural diversity across the core-periphery typology.** (A) Distributions of differences in MDL between the best-fit hub-and-spoke and layered models, by network domain. Different networks within and across domains are better modeled by hub-and-spoke or layered structure, indicating that there is no one universal best descriptor of core-periphery structure. (B) Difference in MDL plotted against the distance between the hub-and-spoke and layered partitions for each KONECT network. Networks that have similar MDLs are generally those where both core-periphery models extracted similar partitions. (C) Difference in MDL plotted against the distance between the two-block and  $k$ -cores partitions for each KONECT network. The stronger the disagreement between heuristics, the clearer the fit for either of the SBM-based core-periphery. (D) Distance between the best-fit block model (either hub-and-spoke, indicated by blue, or layered, indicated by red) and the two-block and  $k$ -cores partitions. Histograms show the marginal distributions of distances, where dashed lines indicate the mean distance. Partition distance in all subplots is measured in bits according to the VI.

structure according to the description length. Figure 4D complements these findings by comparing the distances between the inferred partitions according to the stochastic block models and the partitions of the two-block model and  $k$ -cores decomposition. The hub-and-spoke partitions found with the SBM are consistently closer to the two-block partitions than the  $k$ -core decomposition. The relationship between the layered partitions and the  $k$ -cores partitions is less sharp, with layered and hub-and-spoke partitions being about equally distant from the  $k$ -cores partitions on average. These results provide evidence that the two-block model is representative of the hub-and-spoke characterization. The layered characterization, however, finds partitions that are not quite the same as the  $k$ -cores algorithm. This is due, in part, to the fact that it aggregates nodes in fewer layers. However, as mentioned earlier, the layered block model, along with the hub-and-spoke model, is more flexible than the  $k$ -cores decomposition, which requires core nodes to be high-degree nodes. Instead

and critically, both block models allow for low-degree nodes to be core nodes if they are embedded among other nodes. So, the similar distances of the hub-and-spoke and layered partitions from the  $k$ -cores partitions are also partly explained by how both block models allow for a more fluid interpretation of what it means for a node to be in the core or the periphery.

#### Case study: Hashtag activism amplification

To emphasize the importance of distinguishing between hub-and-spoke and layered core-periphery structures, we briefly conclude with a case study of hashtag activism amplification. Social media are notable for creating spaces where historically disenfranchised individuals can come together and share their stories at an unprecedented scale (34–36). Hashtag activism, in particular, has been a critical vehicle for driving those marginalized voices into the mainstream public sphere (35, 36), as exemplified by hashtags such as #BlackLivesMatter

and #MeToo (35, 37). The amplification of those voices is a fundamentally networked process; the core consists of those who are most visible exactly because many peripheral amplifiers share the core's posts through emergent crowdsourcing (1). Core-periphery structure is a natural network model for such amplification processes.

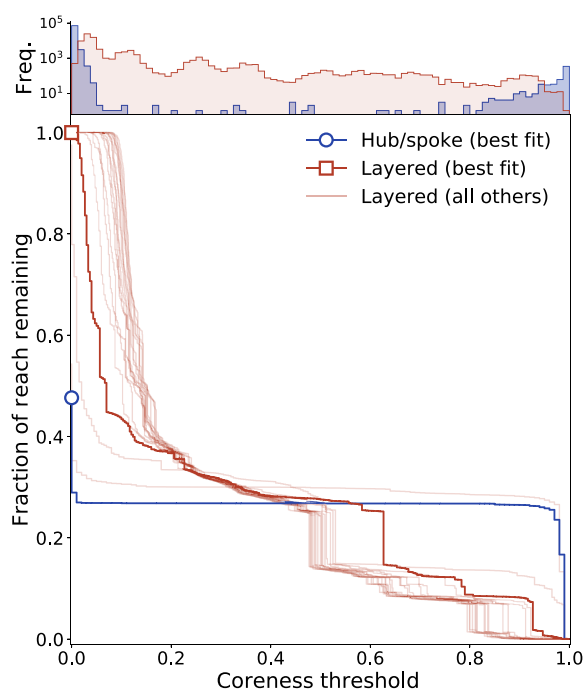
Although those at the periphery of hashtag activism events are sometimes derided as "slacktivists," Barberá *et al.* (1) demonstrated that the periphery contributes significantly to the amplification of core protest voices. We perform a similar analysis on the retweet network of the hashtag #MeToo, a hashtag that highlighted the pervasiveness of sexual violence against women by creating a space for them to publicly disclose their experiences (35, 37). We fit hub-and-spoke and layered core-periphery models to the #MeToo network and calculate the coreness of each individual according to each model. The coreness, which varies from 0 to 1, indicates whether an individual is more likely to be situated in the periphery or core, respectively (see Materials and Methods for details). In line with prior work (1), we operationalize amplification by measuring the total reach as the sum of the number of followers that each individual in the network has.

To quantify the relationship between amplification and the core-periphery structure, we iteratively remove individuals from the retweet network according to their coreness, decomposing the network from the periphery to the core, and measure how the hashtag reach varies (see Fig. 5). We observe that the cumulative reach, the total number of possible followers exposed to the hashtag, declines sharply for both the hub-and-spoke and layered models as the abundance of peripheral amplifiers is removed. This is consistent with the findings of Barberá *et al.* (1). However, the reach drops more rapidly for the hub-and-spoke model (MDL = 7.6 bits per edge, the best-fit model overall) than any of the layered models and, in particular, the best-fit layered model with  $\ell = 4$  layers (MDL = 9.9 bits per edge). Comparatively, the layered models markedly underestimate the contribution of the periphery to the early reach of #MeToo; taking the reach of the hub-and-spoke model as the expected value, the estimates of reach by the best-fit layered model have a percent error of 63% at a coreness threshold of 0.1 and a percent error of 36% at a threshold of 0.2. Hence, we see that using a network model that does not properly describe the core-periphery structure of a hashtag activism network notably misestimates how amplification varies across core and peripheral participants.

This example illustrates why it is critical to account for the core-periphery typology to make sound network inferences. Qualitatively, the Bayesian block models give us a succinct description of the #MeToo retweet network, informing us that it is best described as a hub-and-spoke structure that serves to broadcast a small set of core voices, rather than a layered structure with many connections among those disclosing at the periphery. Quantitatively, using the MDL to select the hub-and-spoke model as the best fit to our network data allows us to confidently estimate the periphery's contribution to the hashtag's reach. This measure can be used to compare across instances of hashtag activism and assess the effectiveness of peripheral amplification or to develop interventions to counteract amplification manipulation tactics, such as those deployed by social bots and coordinated information operations.

## DISCUSSION

We have presented a typology of core-periphery structure that raises the important distinction between two characterizations:



**Fig. 5. Core-periphery amplification of the hashtag #MeToo during its first 12 hours of use in October 2017.** Reach is measured as the cumulative number of followers among those in the network. Curves show how the fraction of total reach decomposes as the coreness threshold for inclusion into the retweet network is increased. The solid blue curve indicates the best-fit hub-and-spoke curve (and best fit overall); the solid red line indicates the best-fit layered curve ( $\ell = 4$  layers), and lighter red lines indicate other layered models with 2 to 20 layers. Markers on the vertical axis indicate the reach after removing nodes with coreness of exactly 0. The histogram above the plot shows the distribution of coreness among nodes in the network for each best-fit model.

hub-and-spoke and layered. These structures, which are reflected in two of the most widely used core-periphery algorithms (15, 18), often yield starkly different descriptions of a network's core-periphery layout. To elucidate the typology, we have formulated two Bayesian stochastic block models that statistically encode the hub-and-spoke and layered structures. By applying description length as an information-theoretic measure of model fit across a large network database, we have shown empirically that networks express a rich variety of core-periphery structure. Through a case study of online amplification of hashtag activism, we have demonstrated that the choice of core-periphery model used to describe a network affects the substantive interpretation of the network's structure and function, indicating the need to distinguish between hub-and-spoke and layered structures.

While a number of algorithms exist for extracting core-periphery structure, they generally take a vaguely intuitive view of core-periphery structure: We have core-periphery structure when we have a core of densely connected nodes and a sparse periphery around that core. Our work challenges the ambiguity of this definition and demonstrates that there are at least two distinct ways that we can conceptualize core-periphery structure, neither of which is a universal descriptor across all networks. Although there is no universal way of describing core-periphery structure, our typology classifies networks into distinct categories based on their specific connectivity patterns between the core and periphery. Within these categories,

there is the potential to unveil larger structural patterns that cut across networks and domains; networks may systematically exhibit layered or hub-and-spoke structures as a result of a number of phenomena, including domain-specific generative processes, network-specific data collection traditions, and context-specific network constraints. Likely, all of these play a role in explaining why hub-and-spoke and layered structures emerge in networks. However, there is significant work to be done across a variety of social, biological, and technological domains to disentangle exactly how and under what conditions those factors affect core-periphery structure. Our work shows that there are core-periphery commonalities across networks and domains that are still unexplained and that a universal notion of core-periphery structure cannot be taken as a given.

In light of this lack of universal organization, we must adjust our practical approach to measuring core-periphery structure going forward. Researchers and network practitioners need to be more explicit about their theories of core-periphery structure and more deliberate in what method they choose to infer that structure. In cases where it is not theoretically clear what kind of core-periphery structure should be reflected in a network—hub-and-spoke, layered, or otherwise—researchers should use the tools that we have introduced here to make informed decisions about what best describes a given network at hand. The core-periphery stochastic block models that we have developed require that we explicitly restrict the search space of the statistical inference to a smaller subset of network structures. These restrictions reduce the expressivity of the model but, at the same time, allow substantive domain experts to guide models as they apply core-periphery models to network datasets, which we argue is critical (30). In the context of our hashtag activism case study, using constrained block models allowed us to focus our inquiry on core-periphery structure, the theoretical network model that describes the structure of online amplification (1). If we were to use a general stochastic block model, which could return community structure, disassortativity, or any other mesoscale pattern, then we would not be able to properly align our domain knowledge with the network analysis. Constrained core-periphery block models force researchers to be more intentional about how they describe core-periphery structure and, in turn, help researchers theorize about core-periphery network effects in a more precise and statistically sound manner.

Both the conceptual typology and the statistical methods that we have presented are only the first step in a broader line of work that interrogates how core-periphery structure is reflected in networks. There are natural methodological extensions for network scientists to develop, which would expand the range of the core-periphery typology. Our models focus on identifying a single core-periphery structure in a network, but there could be multiple or even interacting sets of cores and peripheries (9). Expanding the scope of the core-periphery typology to include this multiplicity would allow for detailed network descriptions that encode the interactions between cores, peripheries, and communities. The same modeling approach could be used to incorporate edge weights, to extend the typology in terms of core-periphery cohesiveness, and directionality, to extend the typology in terms of in-cores and out-cores (38). When we express these extensions and formulations of core-periphery structure (39) in the language of Bayesian block models, we have a statistically consistent framework for adjudicating between them and determining which best describes any given network. By presenting a core-periphery typology with accompanying statistical models that are readily extensible and generalizable, we have

provided the foundation for unifying the notion of core-periphery structure both methodologically and theoretically.

Network scientists increasingly recognize that there is no “ground truth” structure of networks (40); there are only models that do and do not help us address particular questions. Our constrained block models, typology, and measure of model fit make it possible to more acutely answer questions about core and peripheral dynamics that were not previously possible (39). As researchers and practitioners use our methods to be more deliberate about the kind of core-periphery structure that they want to describe, they will undoubtedly raise questions that cannot be answered with the current network models at hand. This presents an opportunity for network scientists to fill those methodological gaps and present models that, themselves, may open doors to new theories and questions. Our core-periphery typology and models clarify the ways in which core-periphery algorithms can be applied to networks and provide an example of how we, as both domain experts and network scientists, can begin to better align our structural methodology with our substantive questions.

## MATERIALS AND METHODS

### KONECT network data

As of the time of data collection, the KONECT (24) consists of 261 networks and represents a variety of network domains. Networks in the collection may be undirected, directed, or bipartite, and they can contain multiedges and self-loops. The edges themselves can be unweighted, weighted, signed, or temporal. We take the following preprocessing steps: (i) Weighted edges are treated as unweighted, and all multiedges are collapsed to a single edge. (ii) Self-loops are disregarded. (iii) Directed edges are treated as undirected. (iv) Only the largest weakly connected component is considered.

We exclude all temporal and dynamic networks in KONECT to avoid the ambiguity in choosing a time scale to define static networks. We also exclude all networks that were marked as “incomplete” in KONECT (24). Last, we exclude bipartite networks because they should be modeled with stochastic block models that can account for their special structure and high local density when projected (41, 42). After these preprocessing and inclusion criteria, we are left with 142 networks, listed in the Supplementary Materials.

We note that the simplifications made during preprocessing likely affect the core-periphery modeling of the KONECT networks. A node's strength, the sum of its edge weights, often correlates with its degree (43), such that the weight-agnostic models may underestimate the cohesiveness of core and how tightly peripheral nodes connect to it. Converting directed edges into undirected ones also forces symmetry upon the adjacency matrix, which can obscure other prominent patterns particular to directed networks (39) that may more comprehensively describe the core-periphery organization.

We infer block models for all KONECT networks with up to 200,000 nodes, a total of 95 of the 142 networks. The networks with more than 200,000 nodes are concentrated on a small set of network domains: 25% are social networks, 25% are hyperlink networks, and 23% are communication networks. Fitting larger networks is possible but requires considerable computation.

### Stochastic block model formulation

Recall that we let  $A$  be the adjacency matrix of an unweighted, undirected, simple network with  $N$  nodes. Nodes are assigned to a



fixed number of  $B$  blocks, represented by  $\theta$ , a vector of length  $N$  where  $\theta_i = r$  indicates that node  $i$  belongs to block  $r$ . Connections between blocks are specified by the  $B \times B$  matrix  $\mathbf{p}$ , where  $p_{rs}$  is the probability that a node in block  $r$  is connected to a node in block  $s$ . The posterior distribution of the parameters  $\theta$  and  $\mathbf{p}$  given our network data  $\mathbf{A}$  is written as

$$P(\theta, \mathbf{p} | \mathbf{A}) \propto P(\mathbf{A} | \theta, \mathbf{p}) P(\theta) P(\mathbf{p}) \quad (7)$$

Earlier, we constructed the prior  $P(\mathbf{p})$  on the block connectivity matrix, which is the primary alteration needed for the core-periphery block models

$$P_{\mathcal{H}}(\mathbf{p}) = 3! \cdot \mathbb{1}_{\{0 < p_{22} < p_{12} < p_{11} < 1\}} \quad (8)$$

and

$$P_{\mathcal{L}}(\mathbf{p}) = \ell! \cdot \mathbb{1}_{\{0 < p_{\ell\ell} < p_{\ell-1\ell} < \dots < p_{11} < 1\}} \quad (9)$$

where the leading numerical factors ensure normalization. Here, we think of blocks as layers, so we let  $B = \ell$ , where  $\ell$  is the number of layers.

We otherwise use a standard formulation of the likelihood  $P(\mathbf{A} | \theta, \mathbf{p})$  and the block assignments prior  $P(\theta)$  (8, 16). The network likelihood rests upon the cornerstone assumption of the stochastic block model: Connections are independently generated on the basis of only the block assignments of nodes. Let  $m_{rs}$  be the number of edges that exist between blocks  $r$  and  $s$ , and let  $M_{rs}$  be the maximum number of edges that could potentially exist between the two blocks. This number equals  $n_r n_s$  for two different blocks of size  $n_r$  and  $n_s$  and  $n_r(n_r - 1)/2$  when considering the internal edges of block  $r$ . The likelihood can be then calculated as the product of independent Bernoulli processes across edges and then aggregated at the block level to yield

$$P(\mathbf{A} | \theta, \mathbf{p}) = \prod_{r \leq s} p_{rs}^{m_{rs}} (1 - p_{rs})^{M_{rs} - m_{rs}} \quad (10)$$

The constraints on  $\mathbf{p}$  yield a more compact form for this likelihood (see the Supplementary Materials). The last missing piece of the model is the layer assignment prior  $P(\theta)$ . The prior on  $\theta$  can then be expressed in three parts (8). First, we consider the probability  $P(\ell)$  of choosing a particular number of layers  $\ell$ , which is always  $\ell = 2$  for the hub-and-spoke model and a free parameter for the layered model. Next, given the number of layers, we consider the probability  $P(\mathbf{n} | \ell)$  of drawing a particular sequence of layer sizes  $\mathbf{n} = \{n_1, n_2, \dots, n_\ell\}$ . Last, given the layer sizes, we determine the probability  $P(\theta | \mathbf{n})$  of seeing a particular allotment of nodes to layers. All together in notation, the prior on the block assignments  $\theta$  is expressed as

$$P(\theta) = P(\theta | \mathbf{n}) P(\mathbf{n} | \ell) P(\ell) = \frac{\prod_r n_r!}{N!} \binom{N-1}{\ell-1}^{-1} N^{-1} \quad (11)$$

With these three parts of the model specified, we can calculate the posterior probability of the model. For more details on the stochastic block model formulation, see (8, 16, 44).

We fit the model with a Metropolis-within-Gibbs algorithm, detailed in the Supplementary Materials. We estimate the true layer of

a node by selecting the layer that maximizes its marginal posterior distribution over layers. In doing so, we average the random fluctuations found in real systems and avoid overfitting to a particular partition when there are many similar optima (45, 46).

## Synthetic networks

### Discernment experiment

The first synthetic network experiment tests the ability of the core-periphery models to discern between hub-and-spoke and layered structures (see Fig. 3A). We generate networks through the stochastic block model according to block matrices given by

$$\begin{bmatrix} p\gamma & p & p(1-\delta) + \frac{p}{\gamma}\delta \\ p & p\delta + \frac{p}{\gamma}(1-\delta) & \frac{p}{\gamma} \\ p(1-\delta) + \frac{p}{\gamma}\delta & \frac{p}{\gamma} & \frac{p}{\gamma} \end{bmatrix} \quad (12)$$

where  $p > 0$  is the baseline density of the network,  $\gamma \in [1, 1/p]$  is the structural clarity parameter, and  $\delta \in [0, 1]$  is the interpolation parameter. The structural clarity parameter  $\gamma$  determines the prevalence of core-periphery structure in the network. When  $\gamma = 1$ , the network as a whole is simply an Erdős-Rényi random network with density  $p$ . When  $\gamma \gg 1$ , the core-periphery structure is well defined. The interpolation parameter  $\delta$  specifies whether a layered or hub-and-spoke core-periphery structure is reflected in the network. When  $\delta = 0$ , the block densities arrange in such a way that the network is effectively generated from two blocks and a hub-and-spoke structure is present. When  $\delta = 1$ , the network exhibits a three-layer structure. We note that  $\delta = 1/2$  holds no special meaning in this interpolation: The structure smoothly transitions from one type to the other as  $\delta$  is varied.

For the experiment, each synthetic network consists of  $n = 10,000$  nodes, divided equally among the three blocks. We set  $p = 0.0075$  and generate networks over the parameter ranges  $\delta \in [0, 1]$  and  $\gamma \in [1, 4]$ . See the “Block model inference and parameters” section for details on inference of the experimental network structure.

### Number of layers experiment

The second synthetic network experiment tests the ability of the layered model to identify the number of layers in synthetic layered networks (see Fig. 3B). We first generate a network  $G$  via the layered stochastic block model, where  $G$  has  $n = 10,000$  nodes evenly split among  $\ell = 6$  layers, where layers are connected according to an initial connectivity matrix  $\mathbf{p}^{(G)}$ . We then consider a new network  $G_k$  of the same number of nodes and layers but where  $p_r^{(G_k)} = p_r^{(G)}$  for  $r < k$  and  $p_r^{(G_k)} = q_k$  for  $r \geq k$ . This is a network where the inner layers have the same density as in  $G$  but where the outermost layers are effectively merged because they have the same density  $q_k$ . We set  $q_k$  such that the overall average degree  $\kappa$  of  $G$  is preserved, i.e., the average degree of  $G_k$  is  $\kappa$  for all  $k$ . The merged layers density  $q_k$  preserving  $\kappa$  is given by

$$q_k = \frac{\binom{n}{2} \sum_{r=k}^{\ell} p_r^{(G)} + n^2 \sum_{r=k}^{\ell} (r-1) p_r^{(G)}}{\binom{n}{2} (\ell - k + 1) + n^2 \sum_{r=k}^{\ell} (r-1)} \quad (13)$$

For each choice of  $k \in [2, 6]$ , we make 10 networks generated from the same block matrix. We define the block matrix  $\mathbf{p}^{(G)}$  of the original network such that  $p_1 = 0.002$ ,  $p_6 = 0.1$ , and  $p_r$  for  $1 < r < 6$  is geometrically distributed between  $p_1$  and  $p_6$ . See the Block model inference and parameters section for details on inference of the experimental network structure.

### Hashtag activism case study

For the hashtag activism case study, we consider all of the tweets containing the hashtag #MeToo that were posted 12 hours after actress Alyssa Milano's "me too" tweet, which catalyzed the hashtag campaign on 15 October 2017 [see (37) for further data details]. The use of human-generated data gathered from Twitter was reviewed and approved by the Institutional Review Board at Northeastern University. In the first 12 hours, there were 208,926 tweets. We construct a retweet network from these tweets by representing individuals as nodes and retweets as edges. For the purpose of core-periphery modeling, we treat edges as undirected and unweighted and remove self-loops from the network. We model the largest weakly connected component of the network (47), which consists of 74,214 nodes and 130,277 edges.

We measure the coreness of each individual by taking into account all the potential core-periphery descriptions identified by a model. We can consider the average block or layer  $\theta_i$  of a node as a measure of its distance to the core of the network and use that to define coreness as

$$c_i = 1 - \frac{1}{\ell} \sum_{r=1}^{\ell} r P(\theta_i = r | \mathbf{A}) \quad (14)$$

In this expression,  $\ell$  is the number of blocks and  $P(\theta_i = r | \mathbf{A})$  is the probability that node  $i$  takes on block assignment  $r$ . The latter probability is the marginal distribution of  $\theta_i$ , formally defined as

$$P(\theta_i = r | \mathbf{A}) = \sum_{\theta} P(\theta | \mathbf{A}) \mathbb{1}_{\{\theta_i = r\}} \quad (15)$$

Coreness varies between 0 and 1, where an individual positioned consistently in the core will have a higher coreness score.

### Block model inference and parameters

For the discernment experiment, we run the hub-and-spoke and layered models three times each for each  $(\gamma, \delta)$  parameter tuple. For each model, we use the best model according to the MDL. For each run of each model, we sweep over 250 Gibbs samples and let each Markov chain Monte Carlo (MCMC) simulation run for 10 times the number of nodes in the network (see the Supplementary Materials for numerical details). We use samples from the second half of the Gibbs sampling chain to infer the parameters  $\hat{\theta}$ , the most probable block labels. We use  $10^7$  samples to approximate the MDL (see the Supplementary Materials for numerical details).

For the layered experiment, we consider  $L$ , the actual number of layers in each synthetic network, and  $\ell$ , the fixed parameter in the layered model. For each  $L$ , there are  $N_L = 10$  networks. For each of those networks, we run the layered models three times and choose the best model from those three runs according to the MDL. We then average the MDL over the  $N_L$  networks to get the average MDL per  $(L, \ell)$  pair. We perform inference similar to the discernment experiment but instead use 20 steps per node for the MCMC chains. To account for more layers than the previous experiment, we use  $10^8$  samples to approximate the MDL.

For each KONECT network and the #MeToo case study network, we run both the hub-and-spoke and layered models three times each. For each model, we choose the best run, as determined by the MDL. For the KONECT networks, we run layered models for  $\ell \in [2, 10]$  and use the model with the best MDL across all choices of  $\ell$ . For the #MeToo network, we vary  $\ell$  in the range  $\ell \in [2, 20]$  and choose the best model overall (thick red line in Fig. 5) and for each individual choice of  $\ell$  (light red lines in Fig. 5) according to the MDL. We use 200 Gibbs samples for the KONECT and #MeToo models and infer partitions according to the second half of each chain. For the MCMC chains, we use 10 steps per node. We use  $10^8$  samples to estimate the MDL.

### SUPPLEMENTARY MATERIALS

Supplementary material for this article is available at <http://advances.sciencemag.org/cgi/content/full/7/12/eabc9800/DC1>

### REFERENCES AND NOTES

1. P. Barberá, N. Wang, R. Bonneau, J. T. Jost, J. Nagler, J. Tucker, S. González-Bailón, The critical periphery in the growth of social protests. *PLOS ONE* **10**, e0143611 (2015).
2. D. S. Bassett, N. F. Wymbs, M. P. Rombach, M. A. Porter, P. J. Mucha, S. T. Grafton, Task-based core-periphery organization of human brain dynamics. *PLOS Comput. Biol.* **9**, e1003171 (2013).
3. J. I. Alvarez-Hamelin, L. Dall'Asta, A. Barrat, A. Vespignani, K-core decomposition of Internet graphs: Hierarchies, self-similarity and measurement biases. *Netw. Heterog. Media* **3**, 371–393 (2008).
4. S. Carmi, S. Havlin, S. Kirkpatrick, Y. Shavitt, E. Shir, A model of Internet topology using k-shell decomposition. *Proc. Natl. Acad. Sci. U.S.A.* **104**, 11150–11154 (2007).
5. M. Kitsak, L. K. Gallos, S. Havlin, F. Liljeros, L. Muchnik, H. E. Stanley, H. A. Makse, Identification of influential spreaders in complex networks. *Nat. Phys.* **6**, 888–893 (2010).
6. F. D. Malliaros, C. Giatsidis, A. N. Papadopoulos, M. Vazirgiannis, The core decomposition of networks: Theory, algorithms and applications. *Vldb J.* **29**, 61–92 (2020).
7. X. Zhang, T. Martin, M. E. J. Newman, Identification of core-periphery structure in networks. *Phys. Rev. E* **91**, 032803 (2015).
8. T. P. Peixoto, Bayesian stochastic blockmodeling, in *Advances in Network Clustering and Blockmodeling*, P. Doreian, V. Batagelj, A. Ferligoj, Eds. (Wiley, ed. 1, 2019), pp. 289–332.
9. S. Kojaku, N. Masuda, Finding multiple core-periphery pairs in networks. *Phys. Rev. E* **96**, 052313 (2017).
10. M. Cucuringu, P. Rombach, S. H. Lee, M. A. Porter, Detection of core-periphery structure in networks using spectral methods and geodesic paths. *Eur. J. Appl. Math.* **27**, 846–887 (2016).
11. F. Tudisco, D. J. Higham, A nonlinear spectral method for core-periphery detection in networks. *SIAM J. Math. Data Sci.* **1**, 269–292 (2019).
12. F. Della Rossa, F. Dercole, C. Piccardi, Profiling core-periphery network structure by random walkers. *Sci. Rep.* **3**, 1467 (2013).
13. C. Ma, B.-B. Xiang, H.-S. Chen, M. Small, H.-F. Zhang, Detection of core-periphery structure in networks based on 3-tuple motifs. *Chaos* **28**, 053121 (2018).
14. M. P. Rombach, M. A. Porter, J. H. Fowler, P. J. Mucha, Core-periphery structure in networks. *SIAM J. Appl. Math.* **74**, 167–190 (2014).
15. S. P. Borgatti, M. G. Everett, Models of core/periphery structures. *Soc. Netw.* **21**, 375–395 (2000).
16. B. Karrer, M. E. J. Newman, Stochastic blockmodels and community structure in networks. *Phys. Rev. E* **83**, 016107 (2011).
17. B. Bollobás, *Modern Graph Theory* (Springer, 1998).
18. S. B. Seidman, Network structure and minimum degree. *Soc. Netw.* **5**, 269–287 (1983).
19. L. Hébert-Dufresne, J. A. Grochow, A. Allard, Multi-scale structure and topological anomaly detection via a new network statistic: The onion decomposition. *Sci. Rep.* **6**, 31708 (2016).
20. V. Batagelj, M. Zaveršnik, An  $O(m)$  algorithm for cores decomposition of networks. arXiv:0310049 [cs.DS] (2003).
21. A. V. Goltsev, S. N. Dorogovtsev, J. F. F. Mendes, k-core (bootstrap) percolation on complex networks: Critical phenomena and nonlocal effects. *Phys. Rev. E* **73**, 056101 (2006).
22. P. W. Holland, K. B. Laskey, S. Leinhardt, Stochastic blockmodels: First steps. *Soc. Netw.* **5**, 109–137 (1983).
23. T. P. Peixoto, Nonparametric Bayesian inference of the microcanonical stochastic block model. *Phys. Rev. E* **95**, 012317 (2017).
24. J. Kunegis, KONECT: The Koblenz Network Collection, in *Proceedings of the 22nd International Conference on World Wide Web (ACM, 2013)*, pp. 1343–1350.

25. S. Z. W. Lip, A fast algorithm for the discrete core/periphery bipartitioning problem. [arxiv:1102.5511](https://arxiv.org/abs/1102.5511) [physics.soc-ph] (2011).
26. M. Meilă, Comparing clusterings—An information based distance. *J. Multivar. Anal.* **98**, 873–895 (2007).
27. A. J. Gates, Y.-Y. Ahn, CluSim: A Python package for calculating clustering similarity. *J. Open Source Softw.* **4**, 1264 (2019).
28. N. X. Vinh, J. Epps, J. Bailey, Information theoretic measures for clusterings comparison: Variants, properties, normalization and correction for chance. *J. Mach. Learn. Res.* **11**, 2837–2854 (2010).
29. M. E. J. Newman, G. T. Cantwell, J.-G. Young, Improved mutual information measure for clustering, classification, and community detection. *Phys. Rev. E* **101**, 042304 (2020).
30. J.-G. Young, G. St-Onge, P. Desrosiers, L. J. Dubé, Universality of the stochastic block model. *Phys. Rev. E* **98**, 032309 (2018).
31. S. C. Olhede, P. J. Wolfe, Network histograms and universality of blockmodel approximation. *Proc. Natl. Acad. Sci. U.S.A.* **111**, 14722–14727 (2014).
32. T. P. Peixoto, Parsimonious module inference in large networks. *Phys. Rev. Lett.* **110**, 148701 (2013).
33. D. J. C. MacKay, *Information Theory, Inference and Learning Algorithms* (Cambridge Univ. Press, ed. 1, 2003).
34. Y. Benkler, *The Wealth of Networks: How Social Production Transforms Markets and Freedom* (Yale Univ. Press, 2006).
35. S. Jackson, M. Bailey, B. Foucault Welles, *#HashtagActivism: Networks of Race and Gender Justice* (MIT Press, 2020).
36. Z. Papacharissi, *Affective Publics: Sentiment, Technology, and Politics* (Oxford Univ. Press, 2015).
37. R. J. Gallagher, E. Stowell, A. G. Parker, B. Foucault Welles, Reclaiming stigmatized narratives: The networked disclosure landscape of #MeToo. *Proc. ACM Hum. Comput. Interact.* **3**, 1–30 (2019).
38. J. P. Boyd, W. J. Fitzgerald, M. C. Mahutga, D. A. Smith, Computing continuous core/periphery structures for social relations data with MINRES/SVD. *Soc. Netw.* **32**, 125–137 (2010).
39. P. Csermely, A. London, L.-Y. Wu, B. Uzzi, Structure and dynamics of core/periphery networks. *J. Complex Netw.* **1**, 93–123 (2013).
40. L. Peel, D. B. Larremore, A. Clauset, The ground truth about metadata and community detection in networks. *Sci. Adv.* **3**, e1602548 (2017).
41. D. B. Larremore, A. Clauset, A. Z. Jacobs, Efficiently inferring community structure in bipartite networks. *Phys. Rev. E* **90**, 012805 (2014).
42. M. Gerlach, T. P. Peixoto, E. G. Altmann, A network approach to topic models. *Sci. Adv.* **4**, eaaq1360 (2018).
43. A. Barrat, M. Barthélemy, R. Pastor-Satorras, A. Vespignani, The architecture of complex weighted networks. *Proc. Natl. Acad. Sci. U.S.A.* **101**, 3747–3752 (2004).
44. T. P. Peixoto, Efficient Monte Carlo and greedy heuristic for the inference of stochastic block models. *Phys. Rev. E* **89**, 012804 (2014).
45. A. Decelle, F. Krzakala, C. Moore, L. Zdeborová, Asymptotic analysis of the stochastic block model for modular networks and its algorithmic applications. *Phys. Rev. E* **84**, 066106 (2011).
46. P. Zhang, C. Moore, Scalable detection of statistically significant communities and hierarchies, using message passing for modularity. *Proc. Natl. Acad. Sci. U.S.A.* **111**, 18144–18149 (2014).
47. M. Newman, *Networks* (Oxford Univ. Press, 2018).

**Acknowledgments:** We thank A. Clauset and D. Larremore for the initial inspiration and depiction of the layered block model presented in this work. We also thank N. Beauchamp, A. Clauset, L. Torres, and J. Davis for conversations early in the project and B. Klein for assistance and advice on the data visualization. **Funding:** This work was supported, in part, by equipment and computing resources from NVIDIA Corporation and Northeastern University's Discovery computing cluster. J.-G.Y. was supported by a James S. McDonnell Foundation Postdoctoral Fellowship Award. **Author contributions:** R.J.G. and B.F.W. conceptualized the project. R.J.G. and J.-G.Y. developed the methods, designed all experiments, and validated all results. B.F.W. and J.-G.Y. supervised the project. R.J.G. implemented and validated all computer code, curated all data, and wrote the initial draft. All authors reviewed and edited the final manuscript. **Competing interests:** The authors declare that they have no competing interests. **Data and materials availability:** All data needed to evaluate the conclusions in the paper are present in the paper and/or the Supplementary Materials. The Python code for inferring the hub-and-spoke and layered core-periphery models and evaluating their model fit is freely available online at [https://github.com/ryanjgallagher/core\\_periphery\\_sbm](https://github.com/ryanjgallagher/core_periphery_sbm). The KONECT dataset used in this work is freely available at <http://konect.cc/>. The Twitter data underlying the #MeToo case study is available at the University of Michigan's Inter-University Consortium for Political and Social Research upon submission and acceptance of a Restricted Data Use Agreement. Additional information related to this paper may be requested from the authors.

Submitted 26 May 2020

Accepted 29 January 2021

Published 17 March 2021

10.1126/sciadv.abc9800

**Citation:** R. J. Gallagher, J.-G. Young, B. F. Welles, A clarified typology of core-periphery structure in networks. *Sci. Adv.* **7**, eabc9800 (2021).