

Super-Resolution Community Detection for Layer-Aggregated Multilayer Networks

Dane Taylor,^{1,2,*} Rajmonda S. Caceres,³ and Peter J. Mucha¹

¹*Carolina Center for Interdisciplinary Applied Mathematics, Department of Mathematics,
University of North Carolina, Chapel Hill, North Carolina 27599, USA*

²*Department of Mathematics, University at Buffalo, State University of New York, Buffalo,
New York 14260, USA*

³*Lincoln Laboratory, Massachusetts Institute of Technology, Lexington, Massachusetts 02420, USA*
(Received 14 September 2016; revised manuscript received 13 July 2017; published 26 September 2017)

Applied network science often involves preprocessing network data before applying a network-analysis method, and there is typically a theoretical disconnect between these steps. For example, it is common to aggregate time-varying network data into windows prior to analysis, and the trade-offs of this preprocessing are not well understood. Focusing on the problem of detecting small communities in multilayer networks, we study the effects of layer aggregation by developing random-matrix theory for modularity matrices associated with layer-aggregated networks with N nodes and L layers, which are drawn from an ensemble of Erdős–Rényi networks with communities planted in subsets of layers. We study phase transitions in which eigenvectors localize onto communities (allowing their detection) and which occur for a given community provided its size surpasses a detectability limit K^* . When layers are aggregated via a summation, we obtain $K^* \propto \mathcal{O}(\sqrt{NL}/T)$, where T is the number of layers across which the community persists. Interestingly, if T is allowed to vary with L , then summation-based layer aggregation enhances small-community detection even if the community persists across a vanishing fraction of layers, provided that T/L decays more slowly than $\mathcal{O}(L^{-1/2})$. Moreover, we find that thresholding the summation can, in some cases, cause K^* to decay exponentially, decreasing by orders of magnitude in a phenomenon we call super-resolution community detection. In other words, layer aggregation with thresholding is a nonlinear data filter enabling detection of communities that are otherwise too small to detect. Importantly, different thresholds generally enhance the detectability of communities having different properties, illustrating that community detection can be obscured if one analyzes network data using a single threshold.

DOI: [10.1103/PhysRevX.7.031056](https://doi.org/10.1103/PhysRevX.7.031056)

Subject Areas: Complex Systems,
Interdisciplinary Physics,
Statistical Physics

I. INTRODUCTION

Network-based modeling provides a powerful framework for analyzing high-dimensional data sets and complex systems [1]. Often, a network is best represented by a set of network layers that encode different types of interactions, such as categorical social ties [2] or a network at different instances in time [3], and an important pursuit involves extending network theory to the multilayer setting [4,5]. Sometimes, however, a multilayer framework can require too much computational overhead or can represent an over-modeling (e.g., when the layers are correlated, either in terms of the edge overlap [6] or other properties [7–9]), and it can be beneficial to aggregate layers [9–11]. In particular, aggregation provides a crucial step for analyzing temporal

network data, which is often binned into time windows [12,13] (see Fig. 1). Layer aggregation and other types of network preprocessing (e.g., sparsification [14], network inference [15], and denoising [16,17]) can greatly influence the resulting network structure, which in turn influences the outcomes of network analyses and their many applications. In general, there remains a significant need for improved theoretical understanding for how such network preprocessing influences network-analysis methodology.

We study the effects of layer aggregation on community detection, one of the widely used methods for studying social, biological, and physical networks [18–21]. Communities are typically studied as dense subgraphs and can represent, for example, coordinating neurons in the brain [13] or a social clique [22] in a social network. (Hereafter, we restrict our usage of the term “clique” to the graph-theoretical meaning of a subgraph with all-to-all coupling.) Of particular interest is the detection of small-scale communities, which is a paradigmatic pursuit for anomaly detection within the fields of signal processing and cybersecurity [23–28]. In this context, small communities can represent anomalous events such as attacks [23], intrusions [24], and fraud [25].

*dane.r.taylor@gmail.com

Published by the American Physical Society under the terms of the *Creative Commons Attribution 3.0 License*. Further distribution of this work must maintain attribution to the author(s) and the published article’s title, journal citation, and DOI.

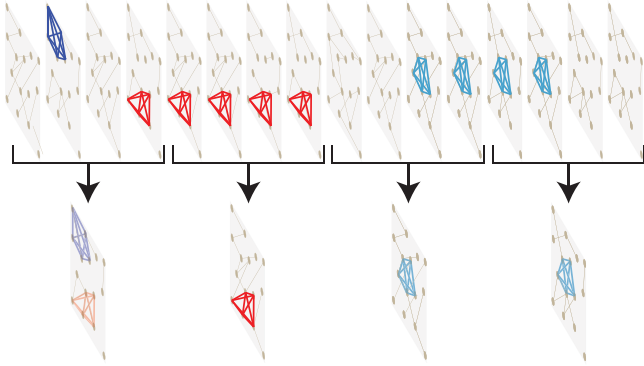


FIG. 1. Preprocessing networks (including multilayer representations of temporal networks) often involves aggregating network data into bins (or time windows). We study how many layers must contain a community in order for aggregation to enhance its detection and introduce layer aggregation with thresholding as a filter enabling super-resolution community detection.

Given these and many other applications, there is great interest in understanding fundamental limitations on community detection [11,26–36]. We highlight recent detectability results for multilayer [10,11,37] and temporal networks [29]. It is worth noting that much of the detectability research has focused on large-scale communities whose sizes are $\mathcal{O}(N)$, where N is the number of nodes in the network [29–35], and the phase transitions are typically driven by varying the prevalence (e.g., edge density) of the communities. In contrast, detectability phase transitions for small communities can also be onset by varying their size K [11,26–28] and are thus a type of resolution limit [36]. We note that the literatures on detectability and resolution limits have developed independently, and there is need for a better understanding of the relationship between these topics. In particular, a planted clique in a single-layer Erdős-Rényi (ER) network is detectable via a spectral analysis only if its size K surpasses a detectability limit $K^* \propto \mathcal{O}(\sqrt{N})$ [26], in which case, a dominant eigenvector (in this case, that corresponding to the second-largest eigenvalue of the adjacency matrix) localizes onto the clique. Extending previous research for the detectability of a clique planted in single-layer networks [26–28] and a clique that persists across all layers of a multilayer network [11], herein we study the detectability of small communities (including, but not limit to, cliques) planted in a subset of layers in a multilayer network.

With the application of detecting small communities in mind, we study the effects of layer aggregation as a network preprocessing step. We first ask a foundational question: Across how many layers must a community persist in order for layer aggregation to benefit detection. To this end, we study a multilayer network model in which small communities are hidden in network layers generated as ER networks with N nodes and L layers with (possibly)

heterogeneous edge probabilities. We study detectability phase transitions wherein eigenvectors localize onto communities, which we analyze by developing random matrix theory for the eigenvectors of modularity matrices associated with an aggregation of the layers. When the aggregation is given by summation of the adjacency matrices, the detectability phase transition occurs when a community’s size $K \ll N$ surpasses a critical value $K^* \propto \sqrt{NL}/T$, where T is the number of layers across which a community persists. Note that if T depends on L , then summation-based layer aggregation benefits small-community detection even if the fraction T/L of layers containing the community vanishes, provided that the fraction decays more slowly than $\mathcal{O}(L^{-1/2})$.

We additionally study network preprocessing via thresholding—that is, we threshold a summation of layers’ adjacency matrices at some value \tilde{L} so that there exists an unweighted edge between two nodes in the aggregated network if and only if there exists at least \tilde{L} edges between them across the L layers. While it is well known that thresholding can be used to simultaneously sparsify and dichotomize a network, here we introduce thresholding as a nonlinear data filter [38] for enhancing small-community detection. Specifically, we find that thresholding can, in some cases, reduce K^* by orders of magnitude, revealing communities that are otherwise too small to detect. We call this phenomenon super-resolution community detection and show, for clique detection in sparse networks, that K^* decays exponentially with \sqrt{L}/T for threshold $\tilde{L} = T$. Importantly, we find that different thresholds enhance the detection of communities with different properties (e.g., size and edge density), illustrating how community structure can be obscured if one uses a single threshold, which is an important insight for network preprocessing in general.

The remainder of this paper is organized as follows. In Sec. II, we further specify our model. In Sec. III, we study the effects of layer aggregation on detectability phase transitions characterized by eigenvector localization. In Sec. IV, we highlight implications of our findings with a numerical experiment involving small-community detection in a temporal network. We provide a discussion in Sec. V

II. MODEL

A. Multilayer networks with planted small communities

We generate L network layers with N nodes so that each layer $l \in \{1, \dots, L\}$ is an ER random graph with edge probability $p_l \in (0, 1)$, which is allowed to vary across the layers. We plant R communities via the following process. For $r \in \{1, \dots, R\}$, uniformly at random, we select a set $\mathcal{T}_r \subset \{1, \dots, L\}$ of layers and a set $\mathcal{K}_r \subset \mathcal{V} = \{1, \dots, N\}$ of nodes, and we define an edge probability ρ_r . The variable $K_r = |\mathcal{K}_r| \ll N$ denotes the size of community r , and we refer to $T_r = |\mathcal{T}_r|$ as its persistence across network layers. Then, for each r , we construct a dense subgraph between

nodes \mathcal{K}_r in layers \mathcal{T}_r by first removing edges between them occurring under the ER model and creating new edges with probability ρ_r . To ensure that the communities are denser than the remaining network, we assume $\rho_r > \langle p_l \rangle$, where $\langle \cdot \rangle$ denotes the mean value across all layers. We allow self-edges in both the ER model and the planted communities. We note that the layers are not required to have a particular ordering, and the community is not restricted only to consecutive layers. Moreover, we restrict our study to nonoverlapping communities by assuming that the communities involve different nodes so that $\mathcal{K}_r \cap \mathcal{K}_s = 0$ for any $r \neq s$. We leave open the study of eigenvector localization in the case of overlapping communities. Finally, we assume $\sum_r K_r \ll N$ so that only a small fraction of nodes are involved in communities, making them anomalous structures.

B. Layer-aggregation methods

We find that layer aggregation is a preprocessing step for multilayer networks that can be used to reduce data size and/or as a data filter to benefit network-analysis outcomes such as community detection. Following the approach in Ref. [10], we study two methods for aggregating layers of a multilayer network:

- (i) The summation network corresponds to the weighted adjacency matrix $\bar{\mathbf{A}} = \sum_l \mathbf{A}^{(l)}$, where $\mathbf{A}^{(l)}$ denotes the symmetric adjacency matrix encoding each network layer $l \in \{1, \dots, L\}$.
- (ii) The family of thresholded networks represented by unweighted adjacency matrices $\{\hat{\mathbf{A}}^{(\tilde{L})}\}$ are obtained by applying a threshold $\tilde{L} \in \{1, \dots, L\}$ to the entries $\{\bar{A}_{ij}\}$ of matrix $\bar{\mathbf{A}}$,

$$\hat{A}_{ij}^{(\tilde{L})} = \begin{cases} 1 & \text{if } \bar{A}_{ij} \geq \tilde{L} \\ 0 & \text{otherwise.} \end{cases} \quad (1)$$

Note that thresholding dichotomizes the network, and one can vary \tilde{L} to tunably sparsify the network.

III. DETECTABILITY OF SMALL COMMUNITIES WITH EIGENVECTOR LOCALIZATION

We now develop random matrix theory to analyze how layer aggregation affects small-community detection. In Sec. III A, we present results for aggregation by summation, studying the fraction of layers that must contain a community in order for layer aggregation to enhance detection. In Sec. III B, we present results for layer aggregation with thresholding, highlighting that certain threshold values can yield super-resolution community detection.

A. Layer aggregation via summation

1. Random matrix theory for modularity matrices

We first describe the statistical properties of matrix entries $\{\bar{A}_{ij}\}$. For edges $(i, j) \in \cup_r \{\mathcal{K}_r \times \mathcal{K}_r\}$, $\{\bar{A}_{ij}\}$ are

independent and identically distributed (i.i.d.) random variables following a Poisson binomial distribution, $P(\bar{A}_{ij} = a) = f_{\text{PB}}(a; L, \{p_l\})$, where

$$f_{\text{PB}}(a; L, \{p_l\}) = \sum_{\mathcal{S} \in \mathcal{S}_a} \prod_{l \in \mathcal{S}} p_l \prod_{m \in \{1, \dots, L\} \setminus \mathcal{S}} (1 - p_m), \quad (2)$$

and \mathcal{S}_a denotes the set of $\binom{L}{a}$ different subsets of layers $\{1, \dots, L\}$ that have cardinality a (i.e., $\mathcal{S}_1 = \{\{1\}, \{2\}, \dots\}$, $\mathcal{S}_2 = \{\{1, 2\}, \{1, 3\}, \dots\}$, and so on). We note that $f_{\text{PB}}(a; L, \{p_l\})$ has mean $L\langle p_l \rangle$ and variance $L\langle p_l(1 - p_l) \rangle$. When the edge probability is identical across the layers (i.e., $p_l = p$), then Eq. (2) simplifies to the binomial distribution,

$$f(a; L, p) = \binom{L}{a} p^a (1 - p)^{L-a}, \quad (3)$$

with mean Lp and variance $Lp(1 - p)$.

For within-community edges $(i, j) \in \{\mathcal{K}_r \times \mathcal{K}_r\}$ associated with community r , the entries $\{\bar{A}_{ij}\}$ are i.i.d. random variables following $f_{\text{PB}}(a; L, \{q_l^{(r)}\})$, where $q_l^{(r)} = \rho_r$ for $l \in \mathcal{T}_r$ and otherwise $q_l^{(r)} = p_l$. It follows that the entries have mean $T_r \rho_r + \sum_{l \in \{1, \dots, L\} \setminus \mathcal{T}_r} p_l$ and variance $T_r \rho_r (1 - \rho_r) + \sum_{l \in \{1, \dots, L\} \setminus \mathcal{T}_r} p_l (1 - p_l)$. Because the layers \mathcal{T}_r are selected uniformly at random, the expected mean and variance across all possible choices for \mathcal{T}_r are given by $T_r \rho_r + (L - T_r) \langle p_l \rangle$ and $T_r \rho_r (1 - \rho_r) + (L - T_r) \langle p_l (1 - p_l) \rangle$, respectively.

We now study the spectra of the modularity matrix [39],

$$\bar{\mathbf{B}} = \bar{\mathbf{A}} - L \langle p_i \rangle \mathbf{1} \mathbf{1}^T, \quad (4)$$

based on an ER null model in which each edge has expected weight $L \langle p_i \rangle$. Importantly, this null model does not use knowledge that edges (i, j) between nodes $i, j \in \mathcal{K}_r$ have different expected edge probability [i.e., $T_r \rho_r + (L - T_r) \langle p_i \rangle$ vs $L \langle p_i \rangle$], which respects our assumption that it is unknown which nodes are in the hidden community. We note that one could also define the ER null model with the observed mean edge probability $L \langle p_i \rangle + \sum_r [(K_r^2 T_r) / N^2 L] (\rho_r - \langle p_i \rangle)$ to account for the slight increase in overall edge probability due to the presence of small communities. However, this change does not affect the position of the dominant eigenvalues relative to the bulk, which is the relevant issue for community detectability, as we will see below. In particular, since $[(K_r^2 T_r) / N^2 L] \ll 1$ for each r , even the shift of the single associated eigenvalue within the bulk is negligible; therefore, we focus on the null model with expected edge weight $L \langle p_i \rangle$.

We develop random matrix theory based on the analysis in Refs. [27,40]. To this end, we note that $\bar{\mathbf{B}}$ can be written in the form

$$\bar{\mathbf{B}} = \langle \bar{\mathbf{B}} \rangle + \mathbf{X}, \quad (5)$$

where

$$\langle \bar{\mathbf{B}} \rangle = \sum_r \theta_r \mathbf{u}^{(r)} (\mathbf{u}^{(r)})^T \quad (6)$$

is a rank- R matrix with eigenvalues given by

$$\theta_r = T_r K_r (\rho_r - \langle p_l \rangle), \quad (7)$$

and $\{\mathbf{u}^{(r)}\}$ are normalized indicator vectors for the R communities that have entries

$$u_i^{(r)} = \begin{cases} \sqrt{1/K_r} & i \in \mathcal{K}_r \\ 0 & \text{otherwise.} \end{cases} \quad (8)$$

The random matrix \mathbf{X} has zero-mean entries X_{ij} with variance $T\rho_r(1-\rho_r) + (L-T_r)\langle p_l(1-p_l) \rangle$ if $(i,j) \in \mathcal{K}_r \times \mathcal{K}_r$, and $L\langle p_l(1-p_l) \rangle$ otherwise. In the $N \rightarrow \infty$ limit, and assuming the sizes $\{K_r\}$ grow more slowly than N , then the $\sum_r K_r^2 \ll N^2$ matrix entries corresponding to communities become negligible and \mathbf{X} limits to a Wigner matrix [41]. This allows us to use known results for the limiting dominant eigenvector of low-rank perturbations of Wigner matrices with variance $1/N$. Specifically, we define $\gamma = 1/\sqrt{NL\langle p_l(1-p_l) \rangle}$ so that the matrix $\gamma\mathbf{X}$ has entries with variance $1/N$ in the limit. We similarly define

$$\bar{\theta}_r = \gamma\theta_r = \frac{T_r K_r}{\sqrt{NL}} \frac{\rho_r - \langle p_l \rangle}{\sqrt{\langle p_l(1-p_l) \rangle}} \quad (9)$$

so that $\gamma\bar{\mathbf{B}} = \sum_r \bar{\theta}_r \mathbf{u}^{(r)} (\mathbf{u}^{(r)})^T + \gamma\mathbf{X}$. It follows that the limiting $N \rightarrow \infty$ dominant eigenvectors $\{\mathbf{v}^{(r)}\}$ of $\gamma\bar{\mathbf{B}}$ (and of $\bar{\mathbf{B}}$ since scalar multiplication does not affect eigenvectors) satisfy [40,42]

$$|\langle \mathbf{v}^{(r)}, \mathbf{u}^{(r)} \rangle|^2 = \begin{cases} 1 - 1/\bar{\theta}^2 & \bar{\theta} > 1 \\ 0 & \text{otherwise.} \end{cases} \quad (10)$$

Note we assume that the dominant eigenvectors have been suitably enumerated so that $\mathbf{v}^{(r)}$ corresponds to the eigenvector localizing on community r . The value $\bar{\theta}_r = 1$ identifies critical points at which there is a phase transition in eigenvector localization and detectability for community r , and this gives the critical community size

$$K_r^* = \sqrt{T_r^{-2} NL} \frac{\sqrt{\langle p_l(1-p_l) \rangle}}{\rho_r - \langle p_l \rangle}. \quad (11)$$

In other words, a small community can be detected using a dominant vector $\mathbf{v}^{(r)}$ of $\bar{\mathbf{B}}$ only when $K_r > K_r^*$. We note that setting $L = T_r = 1$, $\rho_r = 1$, and $p_l = p$ in Eq. (11)

recovers $K_r^* = \sqrt{Np/(1-p)}$, which describes the detectability transition for a single planted clique in a single-layer network [26].

We highlight an important consequence of Eq. (11). First, if the community persists across some fixed fraction of the layers, $T(L) = cL$, then $K_r^* \propto \sqrt{N/L}$; therefore, if N , p , and T_r/L are held fixed and L increases, then K_r^* vanishes with scaling $\mathcal{O}(L^{-1/2})$. This square-root scaling behavior is similar to that obtained for detection in layer aggregation of large-scale communities that persist across all layers [10]. Second, for fixed N and p , a community of fixed size K_r and persistence T_r will become impossible to detect as L increases because K_r^* increases with scaling $\mathcal{O}(L^{1/2})$. This result highlights the importance of knowing which layers potentially contain the community since the aggregation of layers lacking the community can severely inhibit its detection.

Digging further, one can let T_r vary with L and then ask how K_r^* depends on the scaling behavior for T_r . For $T_r \propto L^\beta$, Eq. (11) implies $K_r^* \propto L^{1/2-\beta}$ so that as $L \rightarrow \infty$,

$$K_r^* \rightarrow \begin{cases} 0 & \beta > 1/2 \\ \infty & \beta < 1/2. \end{cases} \quad (12)$$

In other words, T_r , the number of layers containing the community, must increase with L at least as $\mathcal{O}(L^{1/2})$; otherwise, summation-based layer aggregating will inhibit (rather than promote) small-community detection. Note that $T \propto L^{-1/2}$ is a critical case in which K_r^* is independent of L . We highlight that Eq. (12) is somewhat surprising since summation-based aggregation benefits detection even if the fraction T_r/L of layers containing the community vanishes with L , provided that it decays more slowly than $\mathcal{O}(L^{-1/2})$.

2. Numerical validation and scaling behavior

We support Eqs. (10) and (11) in Fig. 2, using numerical experiments with $N = 10^4$ nodes and edge probabilities $\{p_l\}$ drawn from a Gaussian distribution with mean $p = 0.01$ and standard deviation $\sigma_p = 0.001$. We focus on the case of clique detection (i.e., $\rho = 1$), hiding the clique in $T = 2$ of the $L = 16$ layers. In Fig. 2(a), we plot the entries $\{v_i^{(r)}\}$ (symbols) of the dominant eigenvector of the modularity matrix for the summation network as well as the entries $\{u_i^{(r)}\}$ for the indicator vector, which are nonzero only for nodes $i \in \mathcal{K}$ involved in the clique. We show results for community sizes $K_r \in \{6, 26, 86\}$, which respectively place the system below, just above, and well above the phase transition. The illustration highlights that as K increases, vector $\mathbf{v}^{(r)}$ aligns with $\mathbf{u}^{(r)}$. We quantify this localization phenomenon by plotting in Fig. 2(b) observed (symbols) and predicted values of $|\langle v, u \rangle|^2$ given by Eq. (10) (curve). Note that the values of $|\langle v^{(r)}, u^{(r)} \rangle|^2$

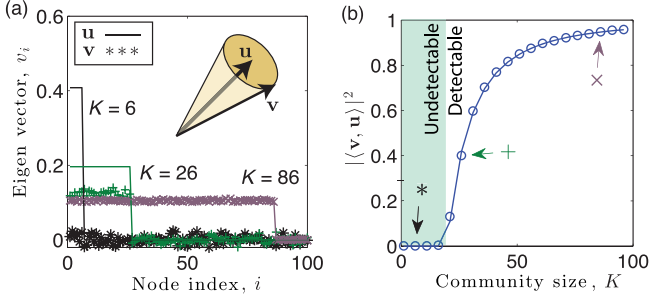


FIG. 2. Eigenvector localization yields detectability phase transition. (a) Entries $v_i^{(r)}$ (symbols) of a dominant eigenvector of the modularity matrix for the summation network of a multilayer network with a hidden community of size K_r . Parameters include $T_r = 2$, $L = 16$, $N = 10^4$, $\rho = 1$, and the edge probabilities $\{p_l\}$ of layers are Gaussian distributed with mean $\langle p_l \rangle = 0.01$ and standard deviation $\sigma_p = 0.001$. To allow visualization, we assume nodes $i \in \{1, \dots, K\}$ are in the community, and we only visualize $v_i^{(r)}$ for nodes $i \in \{1, 100\}$. As shown by the illustration, as K_r increases, $\mathbf{v}^{(r)}$ aligns with the indicator vector $\mathbf{u}^{(r)}$, which is nonzero only for the $K_r \ll N$ entries $u_i^{(r)}$ that correspond to nodes in the community, \mathcal{K}_r . (b) Observed (symbols) and predicted (curves) values of $|\langle \mathbf{v}^{(r)}, \mathbf{u}^{(r)} \rangle|^2$ given by Eq. (10) quantify this localization phenomenon. Arrows indicate the values of K used for panel (a). The critical size K_r^* such that $|\langle \mathbf{v}^{(r)}, \mathbf{u}^{(r)} \rangle|^2 = 0$ for $K_r \leq K_r^*$, whereas $|\langle \mathbf{v}^{(r)}, \mathbf{u}^{(r)} \rangle|^2 > 0$ for $K_r > K_r^*$ marks a phase transition—that is, both in terms of eigenvector localization and detectability of the community.

depict a phase transition that occurs at a critical subgraph size K_r^* given by Eq. (11): $|\langle v^{(r)}, u^{(r)} \rangle|^2 > 0$ when $K_r > K_r^*$, whereas $|\langle v, u \rangle|^2 = 0$ when $K_r \leq K_r^*$. This phase transition in eigenvector localization drives a phase transition for community detection based on $\mathbf{v}^{(r)}$. Arrows indicate the values of K_r used in panel (a).

In Fig. 3(a), we compare observed (symbols) and predicted values of $|\langle v, u \rangle|^2$ given by Eq. (10) (curves) for varying K_r with $T_r \in \{1, 2, 4, 8\}$. Open symbols indicate the parameters used in Fig. 2, whereas filled symbols indicate the mean value of $|\langle \mathbf{v}, \mathbf{u} \rangle|^2$ for 10 trials in which the layers' edge probabilities $\{p_l\}$ are drawn uniformly from $[0, 0.02]$. Note that as T_r increases, the curves shift to the left, illustrating that as the community persists across more layers, the localization phenomenon is stronger and the hidden community is easier to detect. In Fig. 3(b), we study the dependence of K_r^* on the number of layers, L , and we compare the effect of keeping T_r fixed vs allowing T_r to grow with L . Specifically, we set either $T_r(L) = 20$ or $T_r(L) = L$, and we plot the value of K_r^* given by Eq. (11). Note that if the community persists across a fraction of the layers—that is, $T_r(L) = cL$ for some constant c —then K_r^* vanishes with scaling $\mathcal{O}(L^{-1/2})$. However, if T_r is held fixed, then K_r^* increases with scaling $\mathcal{O}(L^{1/2})$.

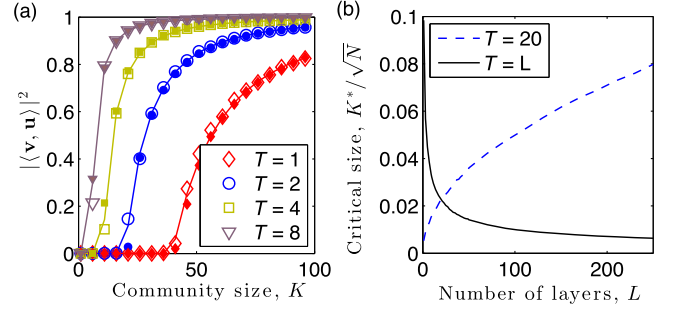


FIG. 3. Influence of community persistence T_r on eigenvector localization for summation-based layer aggregation. (a) Observed (symbols) and predicted values of $|\langle \mathbf{v}^{(r)}, \mathbf{u}^{(r)} \rangle|^2$ given by Eq. (10) (curves) vs K_r for $T_r \in \{1, 2, 4, 8\}$. Open symbols indicate the parameters used in Fig. 2, whereas filled symbols indicate when the layers' edge probabilities $\{p_l\}$ are drawn uniformly from $[0, 0.02]$; we plot the mean value of $|\langle \mathbf{v}^{(r)}, \mathbf{u}^{(r)} \rangle|^2$ across 10 choices for the sets \mathcal{K}_r and T_r . (b) Critical size K_r^* given by Eq. (11) vs L for fixed T_r (dashed line) and $T_r = L$ (solid line). As indicated by Eq. (12), layer aggregation by summation can enhance or inhibit detection depending on whether or not the scaling for $T_r(L)$ exceeds $\mathcal{O}(L^{1/2})$.

In summary, these experiments illustrate how layer aggregation through summation can enhance small-community detection if the community persists across sufficiently many layers, but it can obscure detection if the community is present in too few layers. We will see in the next section that thresholding the summation can help overcome this problem, potentially reducing the detectability limit by orders of magnitude to yield super-resolution community detection.

B. Thresholding as a nonlinear data filter

1. Random matrix theory for modularity matrices

We now study layer aggregation with thresholding as a filter that enhances small-community detection. We begin by solving for *effective* edge probabilities for the thresholding process [10]. Thresholding the summation $\sum_l \mathbf{A}^{(l)}$ at \tilde{L} yields a binary adjacency matrix $\hat{\mathbf{A}}^{(\tilde{L})}$ with entries $\hat{A}_{ij}^{(\tilde{L})} \in \{0, 1\}$ indicating whether or not $\bar{A}_{ij} \geq \tilde{L}$. For edges $(i, j) \in \cup_r \{\mathcal{K}_r \times \mathcal{K}_r\}$, \bar{A}_{ij} follows a Poisson binomial distribution $f_{\text{PB}}(a; L, \{p_l\})$ given by Eq. (2), and the inequality is satisfied with probability

$$\hat{p}^{(\tilde{L})} = P[\bar{A}_{ij} \geq \tilde{L}] = 1 - F_{\text{PB}}(\tilde{L} - 1, L, \{p_l\}), \quad (13)$$

where $F_{\text{PB}}(a, L, \{p_l\})$ is the associated cumulative distribution function (CDF). For edges $(i, j) \in \{\mathcal{K}_r \times \mathcal{K}_r\}$, \bar{A}_{ij} follows a Poisson binomial distribution $f_{\text{PB}}(a; L, \{q_l^{(r)}\})$ given by Eq. (2), and the inequality is satisfied with probability

$$\hat{\rho}_r^{(\tilde{L})} = P[\bar{A}_{ij} \geq \tilde{L}] = 1 - F_{\text{PB}}(\tilde{L} - 1, L, \{q_l^{(r)}\}), \quad (14)$$

where $q_l^{(r)} = \rho_r$ for $l \in \mathcal{T}_r$ and otherwise $q_l^{(r)} = p_l$. In the case of a clique (i.e., $\rho_r = 1$), Eq. (14) can be written as

$$\hat{\rho}_r^{(\tilde{L})} = 1 - F_{\text{PB}}(\tilde{L} - T_r - 1, L - T_r, \{p_l\}_{l \notin \mathcal{T}_r}). \quad (15)$$

Given the effective edge probabilities for the network and a community (i.e., $\hat{p}^{(\tilde{L})}$ and $\hat{\rho}_r^{(\tilde{L})}$, respectively), it is straightforward to study the detectability limits of a community for thresholded networks using Eqs. (10) and (11). In particular, we substitute $L = T_r = 1$ to obtain

$$|\langle \hat{\mathbf{v}}^{(r)}, \mathbf{u}^{(r)} \rangle|^2 = \begin{cases} 1 - 1/\hat{\theta}_r^2 & \hat{\theta}_r > 1 \\ 0 & \text{otherwise,} \end{cases} \quad (16)$$

where $\hat{\mathbf{v}}^{(r)}$ is a dominant eigenvector of modularity matrix

$$\hat{\mathbf{B}} = \hat{\mathbf{A}}^{(\tilde{L})} - \hat{p}^{(\tilde{L})} \mathbf{1}\mathbf{1}^T \quad (17)$$

and $\hat{\theta}_r = K(\hat{\rho}_r^{(\tilde{L})} - \hat{p}^{(\tilde{L})})/\sqrt{N\hat{p}^{(\tilde{L})}(1 - \hat{p}^{(\tilde{L})})}$. Setting $\hat{\theta}_r = 1$ gives a detectability limit for each community r in terms of the effective edge probabilities $\hat{p}^{(\tilde{L})}$ and $\hat{\rho}_r^{(\tilde{L})}$,

$$\hat{K}_r^* = \frac{\sqrt{N\hat{p}^{(\tilde{L})}(1 - \hat{p}^{(\tilde{L})})}}{\hat{\rho}_r^{(\tilde{L})} - \hat{p}^{(\tilde{L})}}. \quad (18)$$

Equations (16)–(18) illustrate that the detectability limits for thresholded networks depend only on the effective edge probabilities; however, these depend sensitively on the choice of threshold \tilde{L} .

Importantly, \hat{K}_r^* given by Eq. (18) can potentially be orders of magnitude smaller than K_r^* given by Eq. (11), a phenomenon we call super-resolution detection. In addition to numerical experiments that will follow below, we further study this phenomenon by comparing \hat{K}_r^* and K_r^* for network parameters wherein we can obtain deeper insight. We consider clique detection (i.e., $\rho_r = 1$) in a sparse network (i.e., $p_l \ll 1$) and focus on the threshold value $\tilde{L} = T_r$ to obtain

$$\hat{K}_r^* \approx \sqrt{N} \sqrt{\hat{p}^{(T_r)}}. \quad (19)$$

Using these assumptions also in Eqs. (13) and (15), we find the effective edge probabilities $\hat{p}^{(T_r)} = 1 - F_{\text{PB}}(T_r - 1, L, \{p_l\})$ and $\hat{\rho}_r^{(T_r)} = 1$. Furthermore, we apply Hoeffding's inequality [43] to obtain $\hat{p}^{(T_r)} \leq e^{-2L(\langle p_l \rangle - T_r/L)^2}$. Noting $0 < \langle p_l \rangle \ll T_r/L$, we find the $\langle p_l \rangle \rightarrow 0$ limiting bound

$$\hat{p}^{(T_r)} \leq e^{-2T_r^2/L}, \quad (20)$$

illustrating that $\hat{p}^{(T_r)}$ and \hat{K}_r^* decay exponentially with T_r^2/L . On the other hand, we use the sparsity assumption in Eq. (11) to obtain

$$K_r^* \approx \frac{\sqrt{NL\langle p_l \rangle}}{\sqrt{T_r^2}}. \quad (21)$$

Thus, in this case, K_r^* decays as $\mathcal{O}(1/\sqrt{T_r^2/L})$, whereas \hat{K}_r^* decays exponentially (i.e., considerably faster) with T_r^2/L .

2. Numerical validation and super-resolution detection

We now support Eqs. (13)–(18) with numerical experiments and illustrate that certain thresholds lead to super-resolution community detection. We consider the detection of a dense subgraph that is hidden in both (a) a dense network with $\langle p_l \rangle = 0.5$ and (b) a sparse network with $\langle p_l \rangle = 0.01$. Both networks were constructed with $N = 10^4$, $\sigma_p = 0.001$, $\rho_r = 1$, $L = 16$, and $T_r = 5$.

In Fig. 4, we compare observed (symbols) and predicted values (curves) of the effective edge probabilities $\hat{p}^{(\tilde{L})}$ given by Eq. (13) and $\hat{\rho}_r^{(\tilde{L})}$ given by Eq. (14) as a function of the threshold \tilde{L} . Note in both panels that the effective edge probability $\hat{p}^{(\tilde{L})}$ of the background network always decays with increasing \tilde{L} . In contrast, the effective edge probability between nodes in the community depends on whether or not $\tilde{L} > T_r$: $\hat{\rho}_r^{(\tilde{L})} = 1$ when $\tilde{L} \leq T_r$ since $\rho = 1$, whereas $\hat{\rho}_r^{(\tilde{L})}$ decays with increasing \tilde{L} for $\tilde{L} > T_r$. Importantly, the rate of decay depends on the network's mean edge density $\langle p_l \rangle$: $\hat{\rho}_r^{(\tilde{L})}$ slowly decreases for the dense network, whereas it abruptly drops for the sparse network.

In Fig. 5, we plot observed (symbols) and predicted values (curves) for $|\langle \hat{\mathbf{v}}^{(r)}, \mathbf{u}^{(r)} \rangle|^2$ given by Eq. (16) vs K for different choices of \tilde{L} . The parameters used are identical to those of Fig. 4, and panels (a) and (b) again depict results for $\langle p_l \rangle = 0.5$ and $\langle p_l \rangle = 0.01$, respectively. We highlight several important observations. First, note in both panels that $\tilde{L} = T_r = 5$ yields better detectability than $\tilde{L} = 1$. However, when $\tilde{L} > T_r$, we find contrasting results for sparse and dense networks. For the sparse network shown in Fig. 5(b), the hidden community becomes harder to detect when $\tilde{L} > T_r$ (see curve for $\tilde{L} = 16$), which intuitively occurs because $\hat{\rho}_r^{(\tilde{L})}$ rapidly decays and the thresholded networks will no longer contain a dense subgraph. On the other hand, for the dense network depicted in Fig. 5(a), increasing \tilde{L} can improve detectability when $\tilde{L} > T_r$ (see curve for $\tilde{L} = 10$).

We now present an experiment highlighting the occurrence of super-resolution community detection for certain threshold values. In Fig. 6, we study the dependence of

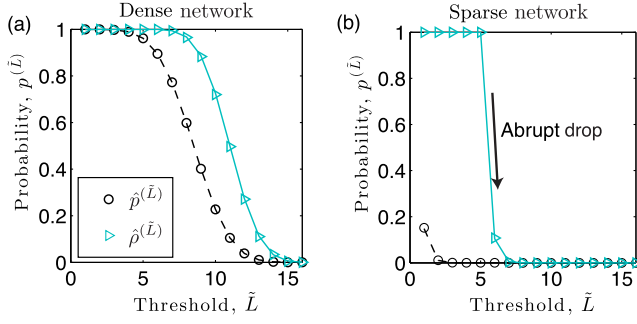


FIG. 4. Effective edge probabilities for threshold-based layer aggregation. Observed (symbols) and predicted values given by Eqs. (13) and (15) (curves) for the effective edge probability of the background network, $\hat{p}(\tilde{L})$, and for a community, $\hat{\rho}(\tilde{L})$, as a function of \tilde{L} . Network parameters include $N = 10^4$, $L = 16$, $T = 5$, and $\sigma_p = 0.001$ and either (a) $\langle p_l \rangle = 0.5$ or (b) $\langle p_l \rangle = 0.01$. Note that for the sparse network in panel (b), $\hat{\rho}(\tilde{L})$ undergoes an abrupt drop when \tilde{L} surpasses $T_r = 5$.

the critical community size K_r^* on the threshold \tilde{L} . We plot \hat{K}_r^* given by Eq. (18) as a function of \tilde{L} for $p \in \{0.01, 0.05, 0.2, 0.5\}$, $N = 10^4$, $\rho = 1$, $\sigma_p = 0.001$, $L = 16$, and either (a) $T_r = 5$ or (b) $T_r = 10$. Note that for the sparsest network, i.e., $p = 0.01$, the minimum value of K_r^* occurs when $\tilde{L} = T_r$ (vertical dashed line). Interestingly, as the mean edge density $p = \langle p_l \rangle$ increases, the threshold \tilde{L} at which \hat{K}_r^* attains its minimum value shifts from $\tilde{L} = T_r$ towards $\tilde{L} = L$. The horizontal lines on the right edge of the panels indicate K_r^* given by Eq. (11) for the summation network.

Importantly, note that for a wide range of parameters, \hat{K}_r^* for the thresholded networks is significantly smaller than K_r^* for the corresponding summation networks. In particular, one can observe for $p = 0.1$ and $\tilde{L}/L = T_r/L$ in Fig. 6(b) that \hat{K}_r^* is many orders of magnitude smaller than K_r^* [$\mathcal{O}(10^{-6})$ times here]. In other words, thresholding

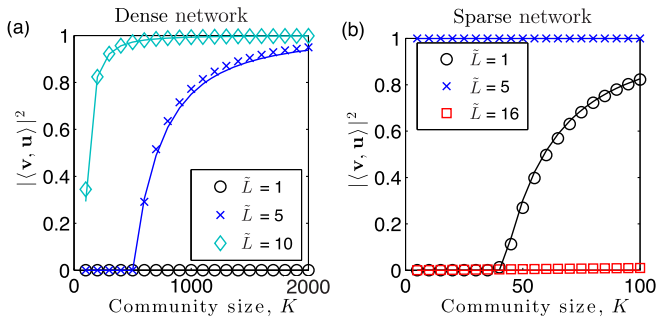


FIG. 5. Detectability phase transitions for threshold-based layer aggregation. We plot $|\langle \mathbf{v}^{(r)}, \mathbf{u}^{(r)} \rangle|^2$ vs community size K_r with identical parameters to those used to produce Fig. 4 except with selected choices for the threshold \tilde{L} .

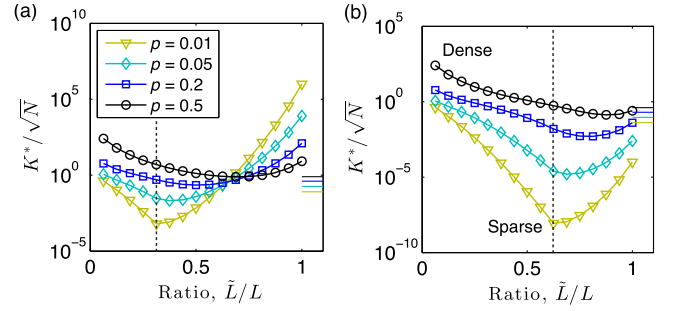


FIG. 6. Super-resolution community detection for threshold-based layer aggregation. We plot \hat{K}_r^* given by Eq. (18) as a function of \tilde{L} for $p \in \{0.01, 0.05, 0.2, 0.5\}$, $N = 10^4$, $\rho = 1$, $\sigma_p = 0.001$, $L = 16$, and either (a) $T_r = 5$ or (b) $T_r = 10$. Note that the \tilde{L} value yielding the minimum \hat{K}_r^* occurs at $\tilde{L} = T_r$ (vertical dotted lines) for sparse networks, whereas it increases with increasing p [e.g., compare $p = 0.01$ and $p = 0.5$ in panel (b)]. The horizontal lines on the right edge of the panels indicate K_r^* given by Eq. (11) for summation networks. Importantly, thresholding can potentially decrease \hat{K}_r^* by many orders of magnitude as compared to K_r^* .

the summation can dramatically improve detectability as compared to summation without thresholding. This surprising result contrasts our previous findings for the detectability of large communities that persist across all layers [10], where it was found that thresholding always inhibited detection (although optimal thresholds were found to minimize inhibition).

IV. SMALL-COMMUNITY DETECTION IN TIME-VARYING NETWORKS

We now present an experiment involving small-community detection in time-varying networks to highlight several practical insights following from our theoretical results. Note that unlike Sec. III, where there were no restrictions on which layers a community persists, we now assume that each community persists across consecutive layers. We conducted experiments for a synthetic temporal network with $N = 10^4$ nodes and $L = 32$ time layers, each of which is drawn from an ER network with edge probability p_l , which we drew from a Gaussian distribution with mean $p = 0.01$ and standard deviation $\sigma_p = 0.001$. We then planted $R = 4$ communities, each involving $K_r = K = 8$ nodes, in the following sets of layers: $\mathcal{T}_1 = \{3, 4, 5\}$ for community 1, $\mathcal{T}_2 = \{7, \dots, 15\}$ for community 2, $\mathcal{T}_3 = \{18, \dots, 22\}$ for community 3, and $\mathcal{T}_4 = \{24, \dots, 30\}$ for community 4. In Fig. 7(a), we provide a representative illustration of the temporal network, where we indicate in which layers the communities are present. We also illustrate by the shaded region an example time window, or bin, $\mathcal{W}_w(t) = \{t - (w - 1)/2, \dots, t + (w - 1)/2\}$ for $t \in \{(w - 1)/2, L - (w - 1)/2\}$, that contains layers to be aggregated.

We first consider aggregation by summation. In Fig. 7(b), we illustrate by color the values $|\langle \mathbf{v}^{(r)}, \mathbf{u}^{(r)} \rangle|^2$ for the aggregation of layers across bins $\mathcal{W}_w(t)$. In particular, we show Eq. (10) under the variable substitutions $T_r(\mathcal{W}_w(t)) \mapsto T$ and $w \mapsto L$, where $T_r(\mathcal{W}_w(t)) = |\mathcal{W}_w(t) \cap \mathcal{T}_r|$ is the number of layers in which community r is present in bin $\mathcal{W}_w(t)$. We show results for several bin widths $w \in \{1, 3, 5, 7, 9\}$. The green arrows indicate, for each r , the bin location and w value at which $|\langle \mathbf{v}^{(r)}, \mathbf{u}^{(r)} \rangle|^2$ obtains its maximum. As expected, $|\langle \mathbf{v}^{(r)}, \mathbf{u}^{(r)} \rangle|^2$ obtains its maximum for each community r when the bin $\mathcal{W}_w(t)$ is exactly the set of layers in which community r is present, $\mathcal{W}_w(t) = \mathcal{T}_r$ (i.e., when $T_r = w$).

Before studying aggregation by summation and thresholding, we first make several important observations using Fig. 7. First, note that for $w = 1$ in panel (b), no communities are detectable. In other words, all communities are undetectable if the layers are studied in isolation. However, they can be detected if the layers are binned into time windows. Second, because the optimal bin size w is unique to every community (i.e., because they have different

persistence $T_r \in [3, 9]$), there is no bin size that is best for all communities. In fact, detectability requires $K_r > K_r^*$ given by Eq. (11), which requires that, for each community, w is not too large or too small. For example, community 1 is only detectable when $w = 3$, and community 3 is only detectable when $w \in [3, 7]$.

One final important observation for Fig. 7(b) is that even when communities are detectable, the values $|\langle \mathbf{v}^{(r)}, \mathbf{u}^{(r)} \rangle|^2$ are not very large—specifically, $|\langle \mathbf{v}^{(r)}, \mathbf{u}^{(r)} \rangle|^2 \leq 0.7$ in all cases. This can be problematic since detection error rates increase as $|\langle \mathbf{v}^{(r)}, \mathbf{u}^{(r)} \rangle|^2$ decreases, approaching 100% error as $|\langle \mathbf{v}^{(r)}, \mathbf{u}^{(r)} \rangle|^2 \rightarrow 0$. (See Ref. [27] for an analysis of error rates based on a hypothesis-testing framework for clique detection in single-layer networks.) Because $|\langle \mathbf{v}^{(r)}, \mathbf{u}^{(r)} \rangle|^2$ remains small for community 1 for all choices of w , it effectively remains undetectable by summation-based layer aggregation.

We now illustrate layer aggregation with thresholding as a filter that can allow greatly improved small-community detection for the temporal network shown in Fig. 7(a), including the accurate recovery of community 1. In Fig. 8,

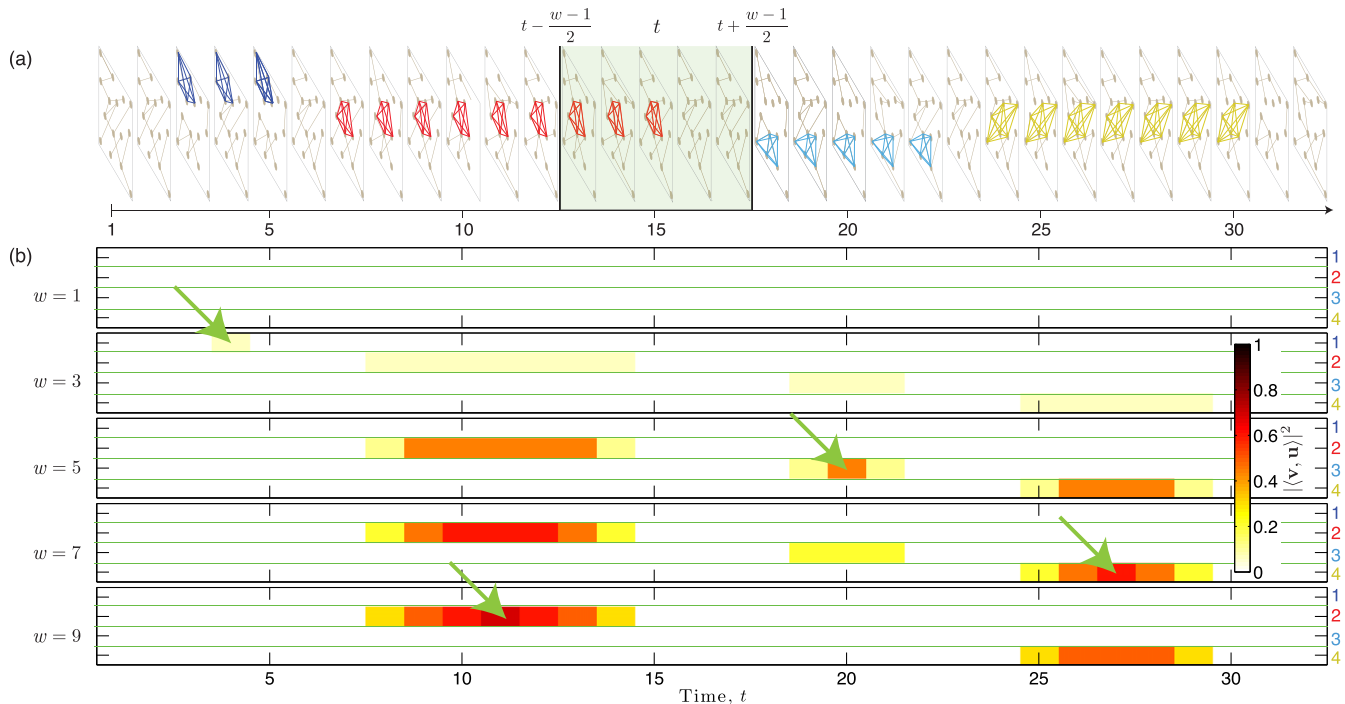


FIG. 7. Detectability of small communities in temporal networks with summation-based binning into time windows. (a) Illustration of a temporal network with $L = 32$ time layers and hidden communities that persist across different time layers. The shaded region indicates a bin, or time window, of size $w \leq L$ at time t for which the layers will be aggregated, which is a process that can be used to discretize and/or smooth the network data. The bin contains layers $\mathcal{W}_w(t) = \{t - (w - 1)/2, \dots, t + (w - 1)/2\}$. (b) We illustrate by color the values $|\langle \mathbf{v}^{(r)}, \mathbf{u}^{(r)} \rangle|^2$ for the aggregation of layers across bins $\mathcal{W}_w(t)$ for each of the four communities $r \in \{1, 2, 3, 4\}$. In particular, we show Eq. (10) under the variable substitutions $T_r(\mathcal{W}_w(t)) \mapsto T$ and $w \mapsto L$, where $T_r(\mathcal{W}_w(t))$ is the number of layers in which community r is present in bin $\mathcal{W}_w(t)$. Layer aggregation across each bin was implemented by summation. We study a temporal network with $N = 10^4$, $L = 32$, $p = 0.01$, $\sigma_p = 0.001$, and we show results for several bin widths $w \in \{1, 3, 5, 7, 9\}$. The hidden communities all contain $K_r = 8$ nodes and have different persistent lengths T_r , as depicted in panel (a). The green arrows indicate, for each r , the bin location and w value at which $|\langle \mathbf{v}^{(r)}, \mathbf{u}^{(r)} \rangle|^2$ obtains its maximum.

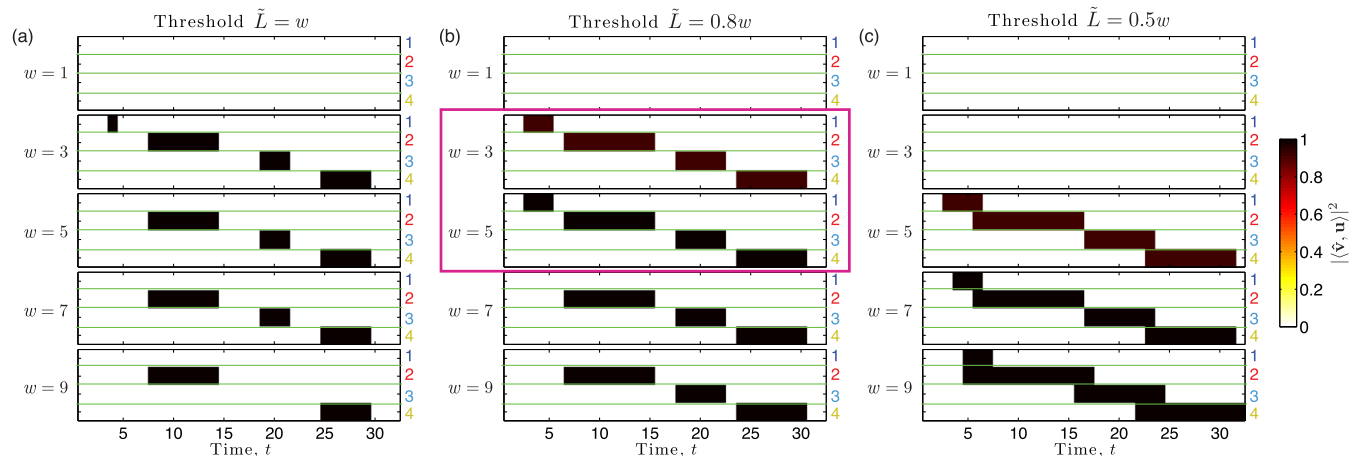


FIG. 8. Detectability of small communities in temporal networks with time-window binning by summation and thresholding. We illustrate by color the values $|\langle \hat{\mathbf{v}}^{(r)}, \mathbf{u}^{(r)} \rangle|^2$ given by Eq. (16) for each of the four communities $r \in \{1, 2, 3, 4\}$ with the variable substitutions $T_r(\mathcal{W}_w(t)) \mapsto T$ and $w \mapsto L$ into Eqs. (13)–(18). Results are shown for bins of width $w \in \{1, 3, 5, 7, 9\}$ for a temporal network with $N = 10^4$ nodes, $L = 32$ time layers, and hidden communities as depicted in Fig. 7(a). The communities each contain $K_r = K = 8$ nodes and have different persistence lengths T_r . Layer aggregation across each bin was implemented by summation and thresholding at \tilde{L} . Panels (a)–(c) respectively indicate the choices $\tilde{L} = w$, $\tilde{L} = 0.8w$, and $\tilde{L} = 0.5w$. The violet box in panel (b) indicates combinations of thresholds and bin sizes that yield accurate detection of all four communities. We stress, however, that since the detectability-limit criterion given by Eq. (18) depends on a complex interplay between the community and network characteristics, one should not, in general, expect there to exist a *single* best combination for *all* communities.

we plot $|\langle \hat{\mathbf{v}}^{(r)}, \mathbf{u}^{(r)} \rangle|^2$ given by Eq. (16) with the variable substitutions $T_r(\mathcal{W}_w(t)) \mapsto T$ and $w \mapsto L$ into Eqs. (13)–(18). Results reflect the aggregation of layers into bins $\mathcal{W}_w(t)$ for each of the four communities $r \in \{1, 2, 3, 4\}$ and with bin sizes $w \in \{1, 3, 5, 7, 9\}$. Panels (a)–(c) indicate results for different thresholds, $\tilde{L} \in \{w, 0.8w, 0.5w\}$.

Our first observation for Fig. 8 is that none of the communities can be detected (for any threshold) if the layers are analyzed in isolation (see results for window size $w = 1$). This result is similar to that shown in Fig. 7(b) for summation without thresholding (i.e., whenever $w = 1$, we find $|\langle \hat{\mathbf{v}}^{(r)}, \mathbf{u}^{(r)} \rangle|^2 = |\langle \mathbf{v}^{(r)}, \mathbf{u}^{(r)} \rangle|^2 = 0$). In other words, the detectability of communities is only made possible through layer aggregation.

Our next observation is that the values $|\langle \hat{\mathbf{v}}^{(r)}, \mathbf{u}^{(r)} \rangle|^2$ are either zero or close to one, which is in sharp contrast to the values of $|\langle \mathbf{v}^{(r)}, \mathbf{u}^{(r)} \rangle|^2$ shown in Fig. 7(b), which can be observed to obtain many values across the range $[0, 0.7]$. In other words, in this experiment, the use of thresholding as a filter allows small communities to be either strongly detected or not detected—there is no middle ground for weak detection (which is the case for layer aggregation without thresholding). This is important since error rates for community detection vanish as $|\langle \hat{\mathbf{v}}^{(r)}, \mathbf{u}^{(r)} \rangle|^2 \rightarrow 1$ [27].

Our final observation is that different threshold values enhance the detectability of different communities. For example, community 1 is detectable when $w = 3$ for $\tilde{L} \geq 0.8w$ but not for $\tilde{L} = 0.5w$ [compare panels (a) and (b) to panel (c)]. Similarly, community 3 is detectable when $w = 9$ for $\tilde{L} \leq 0.8w$ but not for $\tilde{L} = w$ [compare panels (b)

and (c) to panel (a)]. Interestingly, in this experiment, we were able to identify a combination of parameters (\tilde{L}, w) that allows accurate detection of all four communities—that is, $|\langle \hat{\mathbf{v}}^{(r)}, \mathbf{u}^{(r)} \rangle|^2 \approx 1$ for bin $\mathcal{W}_w(t)$ only when community r is present in time layer t [i.e., $t \in T_r$]; otherwise, $|\langle \hat{\mathbf{v}}^{(r)}, \mathbf{u}^{(r)} \rangle|^2 \approx 0$. We highlight these values of (\tilde{L}, w) in panel (b) with a violet box. However, we stress that these “best” values for (\tilde{L}, w) arise in this experiment because the communities are relatively similar in size (i.e., $K_r \in [3, 9]$) and density (i.e., $\rho_r = 1$). In general, one should not expect there to exist one choice of parameters (\tilde{L}, w) to work well for all communities since the detectability-limit criterion given by Eq. (18) depends on a complex interplay between the network and community parameters $\{p_l\}$, ρ_L , T_r , K_r , L , and \tilde{L} .

V. DISCUSSION

There is considerable need to better understand how network preprocessing affects network-analysis methodologies. Herein, we studied how different methods for layer aggregation affect the detectability of small-scale communities in multilayer networks (including multilayer representations of temporal networks). Small-community detection is widely used for anomaly detection in network data [23–28]; in cybersecurity, for example, it allows detection of harmful events such as attacks [23], intrusions [24], and fraud [25]. Understanding limitations on small-community detection provides insight towards the detectability of these harmful activities. Despite most

networks inherently changing in time, previous theory for limitations on small-community detection have been restricted to single-layer networks [26,27] or summation-based aggregation [11]. We highlight that our model and analysis generalizes these previous works in several ways: (i) A community has edge probability $\rho \in (0, 1]$ and is not necessarily a clique, (ii) a community can persist across a subset of layers, (iii) the mean edge probability p_l can vary across network layers, and (iv) the multilayer or temporal network can simultaneously contain several communities.

Motivated in this way, we developed random matrix theory [27,40] to analyze detectability phase transitions in which the dominant eigenvectors of modularity matrices associated with layer-aggregated multilayer networks localize onto communities, thereby allowing their detection. We developed theory for when a community with $K_r \ll N$ nodes is hidden (i.e., planted) in $T_r \leq L$ layers of a multilayer network with N nodes and L layers. We found a detectability phase transition to occur for a given community r when its size K_r surpasses a detectability limit. When layers are aggregated by summation, the detectability limit K_r^* is given by Eq. (11) and has the scaling behavior $K_r^* \propto \sqrt{NL}/T_r$. Surprisingly, if L is allowed to vary, this implies that summation-based aggregation enhances community detection even if the community exists in a vanishing fraction T_r/L of layers, provided that T_r/L decays more slowly than $\mathcal{O}(L^{-1/2})$. This result is surprising since layer aggregation still benefits community detection despite the fact that most layers carry no information about the community.

We also introduced and studied the utility of layer aggregation with thresholding as a nonlinear data filter to enhance small-community detection. Our analysis [particularly, Eq. (18)] revealed that in addition to implementing sparsification and dichotomization, thresholding can allow super-resolution community detection, whereby the detectability limit decreases by several orders of magnitude (see Fig. 6). In particular, we showed in Sec. III B that \hat{K}_r^* decays exponentially with \sqrt{L}/T_r for clique detection in layer-aggregated sparse networks filtered by threshold $\tilde{L} = T_r$.

To illustrate practical implications of our results, in Sec. IV we presented an experiment involving the detection of small communities in a time-varying network, highlighting the following key insights:

- (i) Aggregating time layers into appropriate-sized bins can allow the detection of small communities that would otherwise be undetectable (that is, if the layers were considered in isolation or if all layers were aggregated).
- (ii) Layer aggregation by summation enhances community detection if the community persists across sufficiently many [specifically, $\mathcal{O}(L^{1/2})$] layers; otherwise, it can obscure detection.

- (iii) Layer aggregation with thresholding is a filter that can allow super-resolution community detection of small communities that are otherwise too small for detection.
- (iv) The threshold that best enhances the detection of a small community depends on many parameters, and the detection of multiple communities should, in general, utilize multiple thresholds.

We have thus provided a theoretical framework supporting how small-community detection in temporal network data can be improved through network preprocessing in which network layers are binned into time windows and are aggregated using summation with thresholding. This filtering, however, should not be approached as a “one-size-fits-all” procedure. In particular, we find that there exist optimal time window sizes w and layer-aggregation strategies that, in general, are unique to each community (i.e., depending on its size, density, persistence across the layers, etc.). While it is important to consider a range of window sizes and layer-aggregation methods, this leads to an unavoidable trade-off between computational cost and sufficient exploration of different parameters.

Before concluding, we discuss implications of our work regarding the topic of eigenvector localization in complex networks, which is an important topic in network science [44,45] for the study of centrality [46–48], spatial analysis [49], and core-periphery structure [50,51]. In particular, there is growing interest in extending these ideas to time-varying [52] and multilayer networks [53]. Recently, Ref. [54] showed that an Anderson-localization-type transition occurs for material transport on several real-world networks (e.g., interconnected ponds of melting sea ice, porous human bone, and resistor networks) and noted that they did not observe the wave interference and scattering effects that typically occur for Anderson localization (a widely studied phenomenon in which eigenfunctions localize onto defects in disordered materials [55,56]). Reference [54] found the phase transition to coincide with a phase transition in network connectivity due to eigenvector localization onto different connected components. Our work complements these findings, showing that a similar localization phenomenon can be brought on by small communities—that is, localization does not necessarily require network fragmentation. (We note in passing that connected components can be interpreted as one, and perhaps the strictest, notion of a community.) Future research should further explore the connection between community-based and connected-component-based eigenvector localization on networks, and their relationship to Anderson localization in materials. (See Refs. [57,58] for related research using network-based models for disordered and composite materials.)

Finally, we highlight other extensions to our work that would be interesting to pursue. Motivated by applications for data fusion, recent research [11] considered weighted

averaging of adjacency matrices, allowing them to optimize the weights for the different network layers. It would be interesting to extend our research to weighted averages, which should be fairly straightforward by redefining $\langle \cdot \rangle$ in Eqs. (9)–(11) with weights. We leave open the joint optimization of weighting and thresholding. Finally, it would also be interesting to use our method to study the temporal behavior of communities [59], such as a set of nodes that form a recurring community in different time windows (i.e., periodically or stochastically).

ACKNOWLEDGMENTS

D. T. and P. J. M. were supported by the Eunice Kennedy Shriver National Institute of Child Health & Human Development of the National Institutes of Health (Grant No. R01HD075712) and a James S. McDonnell Foundation 21st Century Science Initiative–Complex Systems Scholar Award (No. 220020315). R. S. C. was supported by the Assistant Secretary of Defense for Research and Engineering under Air Force Contract No. FA8721-05-C-0002 and/or No. FA8702-15-D-0001. Interpretations, opinions, and conclusions of this work are those of the authors and do not reflect the official position of these funding agencies.

-
- [1] M. E. J. Newman, *The Structure and Function of Complex Networks*, *SIAM Rev.* **45**, 167 (2003).
- [2] K. Lewis, J. Kaufman, M. Gonzalez, A. Wimmer, and N. Christakis, *Tastes, Ties, and Time: A New Social Network Dataset Using Facebook.com*, *Soc. Networks* **30**, 330 (2008).
- [3] P. Holme and J. Saramäki, *Temporal Networks*, *Phys. Rep.* **519**, 97 (2012).
- [4] S. Boccaletti, G. Bianconi, R. Criado, C. Del Genio, J. Gómez-Gardenes, M. Romance, I. Sendina-Nadal, Z. Wang, and M. Zanin, *The Structure and Dynamics of Multilayer Networks*, *Phys. Rep.* **544**, 1 (2014).
- [5] M. Kivelä, A. Arenas, M. Barthelemy, J. P. Gleeson, Y. Moreno, and M. A. Porter, *Multilayer Networks*, *J. Complex Netw.* **2**, 203 (2014).
- [6] G. Menichetti, D. Remondini, and G. Bianconi, *Correlations Between Weights and Overlap in Ensembles of Weighted Multiplex Networks*, *Phys. Rev. E* **90**, 062817 (2014).
- [7] M. De Domenico, V. Nicosia, A. Arenas, and V. Latora, *Structural Reducibility of Multilayer Networks*, *Nat. Commun.* **6**, 6864 (2015).
- [8] K. K. Kleineberg, M. Boguna, M. A. Serrano, and F. Papadopoulos, *Hidden Geometric Correlations in Real Multiplex Networks*, *Nat. Phys.* **12**, 1076 (2016).
- [9] N. Stanley, S. Shai, D. Taylor, and P. J. Mucha, *Clustering Network Layers with the Strata Multilayer Stochastic Block Model*, *IEEE Trans. Network Sci. Eng.* **3**, 95 (2016).
- [10] D. Taylor, S. Shai, N. Stanley, and P. J. Mucha, *Enhanced Detectability of Community Structure in Multilayer Networks through Layer Aggregation*, *Phys. Rev. Lett.* **116**, 228301 (2016).
- [11] H. Nayar, B. A. Miller, K. Geyer, R. S. Caceres, S. T. Smith, and R. R. Nadakuditi, *Improved Hidden Clique Detection by Optimal Linear Fusion of Multiple Adjacency Matrices*, *49th Asilomar Conference on Signals, Systems and Computers*, 1520 (IEEE, New York, 2015).
- [12] P. J. Mucha, T. Richardson, K. Macon, M. A. Porter, and J. P. Onnela, *Community Structure in Time-Dependent, Multiscale, and Multiplex Networks*, *Science* **328**, 876 (2010).
- [13] D. S. Bassett, N. F. Wymbs, M. A. Porter, P. J. Mucha, J. M. Carlson, and S. T. Grafton, *Dynamic Reconfiguration of Human Brain Networks During Learning*, *Proc. Natl. Acad. Sci. U.S.A.* **108**, 7641 (2011).
- [14] F. Chung and W. Zhao, *A Sharp PageRank Algorithm with Applications to Edge Ranking and Graph Sparsification*, in *Proceedings of the 2010 International Workshop on Algorithms and Models for the Web-Graph* (Nature Publishing, London, 2010), pp. 2–14.
- [15] S. M. Hill *et al.*, *Inferring Causal Molecular Networks: Empirical Assessment through a Community-Based Effort*, *Nat. Methods* **13**, 310 (2016).
- [16] A. Clauset, C. Moore, and M. E. J. Newman, *Hierarchical Structure and the Prediction of Missing Links in Network*, *Nature (London)* **453**, 98 (2008).
- [17] M. E. J. Newman, *Measurement Errors in Network Data*, arXiv:1703.07376.
- [18] S. Fortunato, *Community Detection in Graphs*, *Phys. Rep.* **486**, 75 (2010).
- [19] M. Rosvall and C. T. Bergstrom, *Maps of Random Walks on Complex Networks Reveal Community Structure*, *Proc. Natl. Acad. Sci. U.S.A.* **105**, 1118 (2008).
- [20] A. Lancichinetti, S. Fortunato, and F. Radicchi, *Benchmark Graphs for Testing Community Detection Algorithms*, *Phys. Rev. E* **78**, 046110 (2008).
- [21] M. Sales-Pardo, R. Guimera, A. A. Moreira, and L. A. N. Amaral, *Extracting the Hierarchical Organization of Complex Systems*, *Proc. Natl. Acad. Sci. U.S.A.* **104**, 15224 (2007).
- [22] J. Moody, *Peer Influence Groups: Identifying Dense Clusters in Large Networks*, *Soc. Networks* **23**, 261 (2001).
- [23] D. Mavroeidis, L. Batina, T. van Laarhoven, and E. Marchiori, *PCA, Eigenvector Localization and Clustering for Side-Channel Attacks on Cryptographic Hardware Devices*, in *Joint European Conference on Machine Learning and Knowledge Discovery in Databases* (Springer, Berlin, 2012), pp. 253–268.
- [24] Q. Ding, N. Katenka, P. Barford, E. Kolaczyk, and M. Crovella, *Intrusion as (Anti) Social Communication: Characterization and Detection*, in *Proceedings of the 18th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (ACM, New York, 2012), pp. 886–894.
- [25] S. Chen and A. Gangopadhyay, *A Novel Approach to Uncover Health Care Frauds Through Spectral Analysis*, in *IEEE International Conference on Healthcare Informatics* (IEEE, New York, 2013), pp. 499–504.
- [26] N. Alon, M. Krivelevich, and B. Sudakov, *Finding a Large Hidden Clique in a Random Graph*, *Random Struct. Algorithm* **13**, 457 (1998).

- [27] R. R. Nadakuditi, *On Hard Limits of Eigen-Analysis Based Planted Clique Detection*, *IEEE Statistical Signal Processing Workshop* (IEEE, New York, 2012), p. 129.
- [28] B. A. Miller, M. S. Beard, P. J. Wolfe, and N. T. Bliss, *Spectral Framework for Anomalous Subgraph Detection*, *IEEE Trans. Signal Process.* **63**, 4191 (2015).
- [29] A. Ghasemian, P. Zhang, A. Clauset, C. Moore, and L. Peel, *Detectability Thresholds and Optimal Algorithms for Community Structure in Dynamic Networks*, *Phys. Rev. X* **6**, 031005 (2016).
- [30] T. Kawamoto and Y. Kabashima, *Detectability of the Spectral Method for Sparse Graph Partitioning*, *Europhys. Lett.* **112**, 40007 (2015).
- [31] A. Decelle, F. Krzakala, C. Moore, and L. Zdeborová, *Inference and Phase Transitions in the Detection of Modules in Sparse Networks*, *Phys. Rev. Lett.* **107**, 065701 (2011).
- [32] R. R. Nadakuditi and M. E. J. Newman, *Graph Spectra and the Detectability of Community Structure in Networks*, *Phys. Rev. Lett.* **108**, 188701 (2012).
- [33] F. Radicchi, *Detectability of Communities in Heterogeneous Networks*, *Phys. Rev. E* **88**, 010801 (2013).
- [34] T. P. Peixoto, *Eigenvalue Spectra of Modular Networks*, *Phys. Rev. Lett.* **111**, 098701 (2013).
- [35] P. Y. Chen and A. O. Hero, *Phase Transitions in Spectral Community Detection*, *IEEE Trans. Signal Process.* **63**, 4339 (2015).
- [36] S. Fortunato and M. Barthelemy, *Resolution Limit in Community Detection*, *Proc. Natl. Acad. Sci. U.S.A.* **104**, 36 (2007).
- [37] P. Y. Chen and A. O. Hero, *Multilayer Spectral Graph Clustering via Convex Layer Aggregation: Theory and Algorithms*, arXiv:1708.02620.
- [38] B. D. O. Anderson and J. B. Moore, *Optimal Filtering* (Prentice-Hall, Englewood Cliffs, New Jersey, 1979).
- [39] M. E. J. Newman and M. Girvan, *Finding and Evaluating Community Structure in Networks*, *Phys. Rev. E* **69**, 026113 (2004).
- [40] F. Benaych-Georges and R. R. Nadakuditi, *The Eigenvalues and Eigenvectors of Finite, Low Rank Perturbations of Large Random Matrices*, *Adv. Math.* **227**, 494 (2011).
- [41] Z. Bai and J. W. Silverstein, *Spectral Analysis of Large Dimensional Random Matrices* (Springer, New York, 2010).
- [42] M. Capitaine, C. Donati-Martin, and D. Féral, *The Largest Eigenvalues of Finite Rank Deformation of Large Wigner Matrices: Convergence and Nonuniversality of the Fluctuations*, *Ann. Prob.* **37**, 1 (2009).
- [43] W. Hoeffding, *Probability Inequalities for Sums of Bounded Random Variables*, *J. Am. Stat. Assoc.* **58**, 13 (1963).
- [44] J. A. Méndez-Bermúdez, A. Alcazar-Lopez, A. J. Martinez-Mendoza, F. A. Rodrigues, and T. K. DM. Peron, *Universality in the Spectral and Eigenfunction Properties of Random Networks*, *Phys. Rev. E* **91**, 032122 (2015).
- [45] R. Pastor-Satorras and C. Castellano, *Distinct Types of Eigenvector Localization in Networks*, *Sci. Rep.* **6**, 18847 (2016).
- [46] T. Martin, X. Zhang, and M. E. J. Newman, *Localization and Centrality in Networks*, *Phys. Rev. E* **90**, 052808 (2014).
- [47] T. Kawamoto, *Localized Eigenvectors of the Non-Backtracking Matrix*, *J. Stat. Mech.* (2016) 023404.
- [48] H. Nassar, K. Kloster, and D. F. Gleich, *Strong Localization in Personalized Page Rank Vectors*, *Algorithms and Models for the Web Graph*, in *Proceedings of the 2015 Workshop on Algorithms for the Web-Graph* (Springer, Cham, 2015), Vol. 9479, pp. 190–202.
- [49] M. Cucuringu, V. D. Blondel, and P. Van Dooren, *Extracting Spatial Information from Networks with Low-Order Eigenvectors*, *Phys. Rev. E* **87**, 032803 (2013).
- [50] P. Barucca, D. Tantari, and F. Lillo, *Centrality Metrics and Localization in Core-Periphery Networks*, *J. Stat. Mech.* **2**, 023401 (2016).
- [51] S. Suweis, *Effect of Localization on the Stability of Mutualistic Ecological Networks*, *Nat. Commun.* **6**, 10179 (2015).
- [52] D. Taylor, S. A. Myers, A. Clauset, M. A. Porter, and P. J. Mucha, *Eigenvector-Based Centrality Measures for Temporal Networks*, *Multiscale Modeling Sim.* **15**, 537 (2017).
- [53] J. A. Méndez-Bermúdez, G. F. de Arruda, F. A. Rodrigues, and Y. Moreno, *Scaling Properties of Multilayer Random Networks*, *Phys. Rev. E* **96**, 012307 (2017).
- [54] N. B. Murphy, E. Cherkaev, and K. M. Golden, *Anderson Transition for Classical Transport in Composite Materials*, *Phys. Rev. Lett.* **118**, 036401 (2017).
- [55] P. W. Anderson, *Localized Magnetic States in Metals*, *Phys. Rev. Lett.* **124**, 41 (1961).
- [56] E. Abrahams, P. W. Anderson, D. C. Licciardello, and T. V. Ramakrishnan, *Scaling Theory of Localization: Absence of Quantum Diffusion in Two Dimensions*, *Phys. Rev. Lett.* **42**, 673 (1979).
- [57] F. Shi, S. Wang, M. G. Forest, and P. J. Mucha, *Percolation-Induced Exponential Scaling in the Large Current Tails of Random Resistor Networks*, *Multiscale Modeling Sim.* **11**, 1298 (2013).
- [58] F. Shi, S. Wang, M. G. Forest, P. J. Mucha, and R. Zhou, *Network-Based Assessments of Percolation-Induced Current Distributions in Sheared Rod Macromolecular Dispersions*, *Multiscale Modeling Sim.* **12**, 249 (2014).
- [59] V. Sekara, A. Stopczynski, and S. Lehmann, *Fundamental Structures of Dynamic Social Networks*, *Proc. Natl. Acad. Sci. U.S.A.* **113**, 9977 (2016).