

## Enhanced Detectability of Community Structure in Multilayer Networks through Layer Aggregation

Dane Taylor,<sup>1,\*</sup> Saray Shai,<sup>1</sup> Natalie Stanley,<sup>1,2</sup> and Peter J. Mucha<sup>1</sup>

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

<sup>2</sup>*Curriculum in Bioinformatics and Computational Biology, University of North Carolina, Chapel Hill, North Carolina 27599, USA*

(Received 16 November 2015; revised manuscript received 17 March 2016; published 2 June 2016)

Many systems are naturally represented by a multilayer network in which edges exist in multiple layers that encode different, but potentially related, types of interactions, and it is important to understand limitations on the detectability of community structure in these networks. Using random matrix theory, we analyze detectability limitations for multilayer (specifically, multiplex) stochastic block models (SBMs) in which  $L$  layers are derived from a common SBM. We study the effect of layer aggregation on detectability for several aggregation methods, including summation of the layers' adjacency matrices for which we show the detectability limit vanishes as  $O(L^{-1/2})$  with increasing number of layers,  $L$ . Importantly, we find a similar scaling behavior when the summation is thresholded at an optimal value, providing insight into the common—but not well understood—practice of thresholding pairwise-interaction data to obtain sparse network representations.

DOI: 10.1103/PhysRevLett.116.228301

The analysis of complex networks [1] has far-reaching applications ranging from social systems [2] to the brain [3]. Often, the system is a multilayer network (see reviews [4,5]), whereby network layers encode different classes of interactions, such as categorical social ties [6], types of critical infrastructure [7], or a network at different instances in time [8]. In principle, the multilayer framework offers a more comprehensive representation of a data set or system, as compared to an aggregation of layers that produces a simplified model but does so at the cost of information loss. For example, neglecting the layered structure can lead to severe and unintended consequences regarding structure [9] and dynamics [10–12], which can fundamentally differ between single-layer and multilayer networks [13,14].

However, layer aggregation also implements an information processing that can be beneficial. Network layers are often correlated with one another and can encode redundant information [15]. In some cases a multilayer representation is an over-modeling, which can negatively impact the computational and memory requirements for storage and analysis. In such situations, it is beneficial to seek a more concise representation in which certain layers are aggregated [16,17]. Identifying sets of repetitive layers amounts to a clustering problem, and it is closely related to the topic of clustering networks in an ensemble [17,18]. Much remains to be studied regarding *when* layer aggregation is appropriate and *how* it should be implemented.

We study the effect of layer aggregation on community structure in multilayer networks with layers drawn from a common stochastic block model (SBM). SBMs are a paradigmatic model [19] for complex structure in networks and are particularly useful for studying limitations on

detectability—that is, if the community structure is too weak, it cannot be found upon inspection of the network [20–25]. Recently, the detectability limit has been explored for networks with degree heterogeneity [26] and hierarchical structure [27,28], for temporal networks [29], and for the detection of communities using multiresolution methods [30]. Despite growing interest in multilayer SBMs [31–35] (which we note, focus on *multiplex* networks in which nodes are identical in every layer and edges are restricted to connecting nodes in the same layer [4,5]), the effect of layer aggregation on detectability limitations has yet to be explored outside the infinite layer limit [35].

To this end, we study detectability limitations for multilayer SBMs with layers following a single SBM and find layer aggregation to significantly influence detectability. When the aggregate network corresponds to the summation of the adjacency matrices encoding the network layers, aggregation always improves detectability. The detectability limit vanishes with increasing number of layers,  $L$ , and decays as  $O(L^{-1/2})$ . Because the summation of  $L$  adjacency matrices can often yield a weighted and dense network—which increases the complexity of community detection [36]—we also study binary adjacency matrices obtained by thresholding this summation at some value  $\tilde{L}$ . We find that the detectability limit is very sensitive to the choice of  $\tilde{L}$  and that there exist thresholds (e.g., mean edge probability for homogeneous communities) that are optimal in that the detectability limit also decays as  $O(L^{-1/2})$ . These results provide insight into the use of thresholding pairwise-interaction data to construct sparse networks—a practice that is commonplace but for which the effects are not well understood.

We begin by describing the multilayer SBM. We consider  $N$  nodes divided into  $K$  communities, and we denote by  $c_i \in \{1, \dots, K\}$  the community index for each node  $i \in \{1, \dots, N\}$ . The multiplex network is defined by  $L$  layers encoded by a set of adjacency matrices,  $\{\mathbf{A}^{(l)}\}$ , where  $A_{ij}^{(l)} = 1$  if  $(i, j)$  is an edge in layer  $l$  and  $A_{ij}^{(l)} = 0$  otherwise. The probability of edge  $(i, j)$  in layer  $l$  is given by  $\Pi_{c_i c_j} \in [0, 1]$ , where  $\Pi$  is a  $K \times K$  matrix.

The detectability of community structure relates to the ability to recover the nodes' community labels  $\{c_i\}$ . To connect with previous research [21,23–25], we focus on the case of  $K = 2$  communities of equal size with edge probabilities  $\Pi_{11} = \Pi_{22} = p_{\text{in}}$  and  $\Pi_{12} = \Pi_{21} = p_{\text{out}}$ . Below, we will simultaneously refer to these respective probabilities as  $p_{\text{in,out}}$ . We assume  $p_{\text{in}} \geq p_{\text{out}}$  to study “assortative” communities in which there is a prevalence of edges between nodes in the same community [37].

It has been shown for the large network  $N \rightarrow \infty$  limit that there exists a detectability limit characterized [23,24] by the solution curve  $(\Delta^*, \rho)$  to

$$N\Delta = \sqrt{4N\rho}, \quad (1)$$

where  $\Delta = p_{\text{in}} - p_{\text{out}}$  is the difference in probability and  $\rho = (p_{\text{in}} + p_{\text{out}})/2$  is the mean edge probability. For given  $\rho$ , the communities are detectable only when the presence of community structure is sufficiently strong, i.e.,  $\Delta > \Delta^*$ . Equation (1) describes a phase transition that has been obtained via complementary analyses—Bayesian inference [23] and random matrix theory [24]—and represents a critical point that is independent of the community detection method (see [23] and footnote 11 in [24]). We further note that Eq. (1) was derived for sparse networks [i.e., constant  $\rho N$  so that  $\rho = O(N^{-1})$ ]. Here, we must consider the full range of densities,  $\rho \in [0, 1]$ , to allow for aggregated networks that are potentially dense [i.e.,  $\rho = O(1)$  as  $N \rightarrow \infty$ ].

In this Letter, we study the behavior of  $\Delta^*$  for two methods of aggregating layers. We define the *summation* network corresponding to the weighted adjacency matrix  $\bar{\mathbf{A}} = \sum_l \mathbf{A}^{(l)}$  as well as a family of *thresholded* networks with unweighted adjacency matrices  $\{\hat{\mathbf{A}}^{(\tilde{L})}\}$  that are obtained by applying a threshold  $\tilde{L} \in \{1, \dots, L\}$  to the entries of  $\bar{\mathbf{A}}$ . Specifically, we define  $\hat{A}_{ij}^{(\tilde{L})} = 1$  if  $\bar{A}_{ij} \geq \tilde{L}$  and  $\hat{A}_{ij}^{(\tilde{L})} = 0$  otherwise. Of particular interest are the limiting cases  $\tilde{L} = L$  and  $\tilde{L} = 1$ , which respectively correspond to applying logical AND and OR operations to the original multiplex data  $\{A_{ij}^{(l)}\}$  for fixed  $(i, j)$ . We refer to these thresholded networks as the AND and OR networks, respectively.

We study the detectability limit for the layer-aggregated networks using random matrix theory [38,39]. This approach is particularly suited for detectability analysis since community labels  $\{c_i\}$  can be identified using spectral

partitioning and phase transitions [24,27,28] in detectability correspond to the disappearance of gaps between isolated eigenvalues (whose corresponding eigenvectors reflect community structure) and bulk eigenvalues [which arise due to stochasticity and whose  $N \rightarrow \infty$  limiting distribution is given by a spectral density  $P(\lambda)$ ]. We develop theory based on the modularity matrix  $\bar{\mathbf{B}}_{ij} = \bar{A}_{ij} - \rho L$  [40]. Note that we do not use the configuration model as the null model. Instead, since all nodes are identical under the SBM, the appropriate null model is Erdős-Rényi with repeated edges allowed in which so that the expected number of edges between any pair of nodes is  $\rho L$ .

We first study  $\Delta^*$  for the summation network. We analyze the distribution of real eigenvalues  $\{\lambda_i\}$  of  $\bar{\mathbf{B}}$  (in descending order) using methodology developed in [24,38]; we extend this work to networks that are multiplex and possibly dense. We outline our results here and provide further details in the Supplemental Material [41]. We begin by describing the statistical properties of entries  $\{\bar{A}_{ij}\}$ , which are independent random variables following a binomial distribution  $P(\bar{A}_{ij} = a) = f(a; L, \Pi_{c_i c_j})$ , where

$$f(a; L, p) = \binom{L}{a} p^a (1-p)^{L-a} \quad (2)$$

has mean  $Lp$  and variance  $Lp(1-p)$ . Provided that there is sufficiently large variance in the edge probabilities [i.e.,  $NLp(1-p) \gg 1$ ], we find that the limiting  $N \rightarrow \infty$  distribution of bulk eigenvalues for  $\bar{\mathbf{B}}$  is given by a semicircle distribution,

$$P(\lambda) = \frac{\sqrt{\lambda_2^2 - \lambda^2}}{\pi \lambda_2^2 / 2} \quad (3)$$

for  $|\lambda| < \lambda_2$  and  $P(\lambda) = 0$  otherwise, where

$$\lambda_2 = \sqrt{4NL[\rho(1-\rho) - \Delta^2/4]} \quad (4)$$

is the upper bound on the support of this spectral density and is the limiting  $N \rightarrow \infty$  value of the second-largest eigenvalue. The largest eigenvalue of  $\bar{\mathbf{B}}$  in the  $N \rightarrow \infty$  limit is an isolated eigenvalue

$$\lambda_1 = NL\Delta/2 + 2[\rho(1-\rho) - \Delta^2/4]/\Delta. \quad (5)$$

As we shall show,  $\Delta^* \rightarrow 0$  as  $N$  increases, and therefore the  $\Delta^2/4$  terms in Eqs. (4) and (5) are negligible near the detectability limit (i.e.,  $\Delta \approx \Delta^*$ ). The eigenvector  $\mathbf{v}$  corresponding to  $\lambda_1$  gives the spectral bipartition—the inferred community label of node  $i$  is determined by the sign of  $v_i$ —and provided that the largest eigenvalue corresponds to this isolated eigenvalue,  $\lambda_1$ , the eigenvector entries  $\{v_i\}$  are correlated with the community labels  $\{c_i\}$ . To obtain the detectability limit, we set  $\lambda_1 = \lambda_2$ , neglect the  $\Delta^2/4$  terms and simplify, yielding a modified detectability equation

$$NL\Delta = \sqrt{4NL\rho(1-\rho)}. \quad (6)$$

Note that Eq. (6) recovers Eq. (1) when  $L = 1$  and  $\rho \rightarrow 0$  [i.e., for sparse networks,  $\rho(1-\rho) \approx \rho$ ]. Defining  $p_{\text{in}}^* = \rho + \Delta^*/2$  and  $p_{\text{out}}^* = \rho - \Delta^*/2$ , we find for fixed  $\rho$  and increasing  $N$  and/or  $L$  that  $p_{\text{in,out}}^* \rightarrow \rho$  and  $\Delta^* \rightarrow 0$ , decaying as  $O(1/\sqrt{NL})$ .

We now study  $\Delta^*$  for the thresholded networks, which correspond to single-layer SBMs in which the community labels  $\{c_j\}$  are identical to those of the multilayer SBM, but there are new *effective* block edge probabilities

$$\hat{\Pi}_{nm}^{(\tilde{L})} = 1 - F(\tilde{L} - 1; L, \Pi_{nm}), \quad (7)$$

where  $F(a; L, p)$  is the cumulative distribution function for the binomial distribution  $f(a; L, p)$ . The effective probabilities for the AND and OR networks are  $\hat{\Pi}_{nm}^{(L)} = (\Pi_{nm})^L$  and  $\hat{\Pi}_{nm}^{(1)} = 1 - (1 - \Pi_{nm})^L$ , respectively. For the two-community SBM, the effective probabilities are  $\hat{p}_{\text{in,out}}^{(\tilde{L})} = 1 - F(\tilde{L} - 1; L, p_{\text{in,out}})$ ,  $\hat{\Delta}^{(\tilde{L})} = \hat{p}_{\text{in}}^{(\tilde{L})} - \hat{p}_{\text{out}}^{(\tilde{L})}$ , and  $\hat{\rho}^{(\tilde{L})} = (\hat{p}_{\text{in}}^{(\tilde{L})} + \hat{p}_{\text{out}}^{(\tilde{L})})/2$ . The modularity matrices for the thresholded networks become  $\hat{B}_{ij}^{(\tilde{L})} = \hat{A}_{ij}^{(\tilde{L})} - \hat{\rho}^{(\tilde{L})}$ . We identify the detectability limit by substituting  $\hat{\Delta}^{(\tilde{L})} \mapsto \Delta$  and  $\hat{\rho}^{(\tilde{L})} \mapsto \rho$  into Eq. (6) (with  $L = 1$ ) and numerically finding a solution  $(\Delta^*, \rho)$  using a root-finding algorithm. Note that the detectability equation holds for the effective probabilities,  $N\hat{\Delta}^{(\tilde{L})} = \sqrt{4N\hat{\rho}^{(\tilde{L})}(1-\hat{\rho}^{(\tilde{L})})}$ , and not the single-layer probabilities,  $N\Delta \neq \sqrt{4N\rho(1-\rho)}$ .

In Figs. 1(a) and 1(b), we show  $\Delta^*$  versus the mean edge probability  $\rho$  for the different aggregation methods: (i) a single layer (red dot-dashed curves), which is identical in panels (a) and (b); (ii) the summation network (blue dashed curves), for which the curve in (b) corresponds to the curve in panel (a) rescaled by a factor of  $1/2$ ; and (iii) thresholded networks (solid curves), which shift left-to-right with increasing  $\tilde{L}$ . This is evident by comparing  $\Delta^*$  for the AND ( $\tilde{L} = L$ , gold circles) and OR ( $\tilde{L} = 1$ , cyan squares) networks. We find when  $\rho$  is large that the AND (OR) network has a relatively small (large) detectability limit; in contrast, when  $\rho$  is small the AND (OR) network has a relatively large (small) detectability limit. In other words, aggregating layers using the AND (OR) operation is beneficial for dense (sparse) networks.

It is interesting to ask if there are choices of  $\rho$  and  $\tilde{L}$  for which the detectability limit vanishes as  $O(L^{-1/2})$  with increasing  $L$ —that is, a behavior similar to that of the summation network. To this end, we study the threshold  $\tilde{L} = \lceil \rho L \rceil$ , which we numerically observe to be the best  $\tilde{L}$  for most values of  $\rho$ . This choice is also convenient as it only requires knowledge of the mean edge probability,  $\rho$ , which is easy to obtain in practice. In Fig. 1(c), we plot  $\Delta^*$

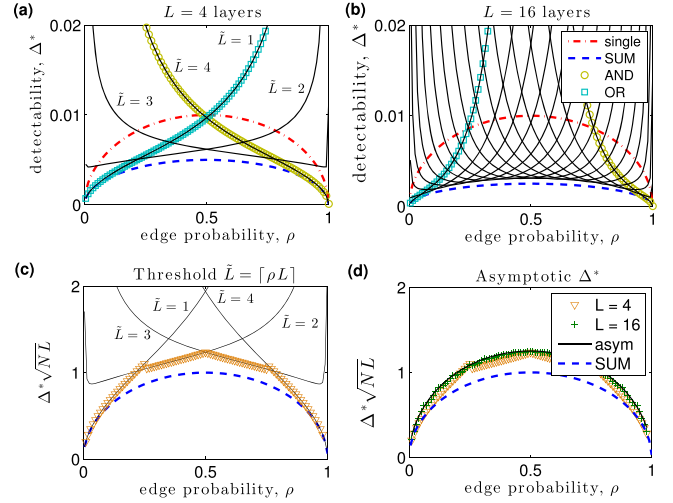


FIG. 1. Layer aggregation enhances the detectability of community structure. (a),(b) We plot the detectability limit  $\Delta^*$  versus mean edge probability  $\rho$  for a single network layer (red dot-dashed curves), the aggregate network obtained by summation (blue dashed curves), and aggregate networks obtained by thresholding this summation at  $\tilde{L} \in \{1, 2, 3, 4\}$  (solid curves). Gold circles and cyan squares highlight  $\tilde{L} = L$  and  $\tilde{L} = 1$ , which we refer to as AND and OR networks, respectively. Results are shown for  $N = 10^4$  nodes with (a)  $L = 4$  and (b)  $L = 16$  layers. (c) For  $L = 4$ , we show  $\Delta^*$  versus  $\rho$  for the optimal threshold  $\tilde{L} = \lceil \rho L \rceil$  (orange triangles), which lies on the solution curves for  $\tilde{L} \in \{1, \dots, L\}$  (solid curves). (d) We show  $\Delta^*$  for  $\tilde{L} = \lceil \rho L \rceil$  with  $L \in \{4, 16\}$ . These piecewise-continuous solutions collapse onto the asymptotic solution  $\Delta_{\text{asym}}^*$  (black curve) as  $L$  increases. In panels (c),(d), we additionally plot  $\Delta^*$  for the summation network (blue dashed curves).

versus  $\rho$  for  $L = 4$  and  $\tilde{L} = \lceil \rho L \rceil$  (orange triangles), which lies along the solution curves for  $\tilde{L} \in \{1, \dots, L\}$  (solid curves). In Fig. 1(d), we plot  $\Delta^*$  for threshold  $\tilde{L} = \lceil \rho L \rceil$  with  $L = 4$  (orange triangles) and  $L = 64$  (green crosses). These curves align due to the rescaling of the vertical axis by  $\sqrt{NL}$ . In fact, we find in the large  $L$  limit that these solutions  $\Delta^*$  collapse onto a single curve  $(\Delta_{\text{asym}}^*, \rho)$  that solves

$$NL\Delta = \sqrt{2\pi NL\rho(1-\rho)}, \quad (8)$$

which we plot by the black line in Fig. 1(d). To obtain Eq. (8), we use the central limit theorem [42] to approximate  $\hat{p}_{\text{in,out}}^{(\lceil \rho L \rceil)} \approx \hat{p}_{\text{in,out}}^{(\text{asym})} = 1 - G[L\rho; Lp_{\text{in,out}}, Lp_{\text{in,out}}(1-p_{\text{in,out}})]$ , where  $G(p; \mu, \sigma^2) = \frac{1}{2} + \frac{1}{2} \text{erf}[(p - \mu)/\sigma\sqrt{2}]$  is the value of the cumulative distribution function of the normal distribution with mean  $\mu$  and variance  $\sigma^2$  evaluated at  $p$ . In particular, we approximate  $\hat{\Delta}^{(\lceil \rho L \rceil)} \approx \hat{\Delta}^{(\text{asym})} = \text{erf}(\Delta\sqrt{L}/\sqrt{8\rho(1-\rho)})$  and  $\hat{\rho}^{(\lceil \rho L \rceil)} \approx \hat{\rho}^{(\text{asym})} = 1/2$ . Equation (8) is recovered after substituting  $\hat{\Delta}^{(\text{asym})} \mapsto \Delta$  and  $\hat{\rho}^{(\text{asym})} \mapsto \rho$  into Eq. (6) with  $L = 1$  and using the

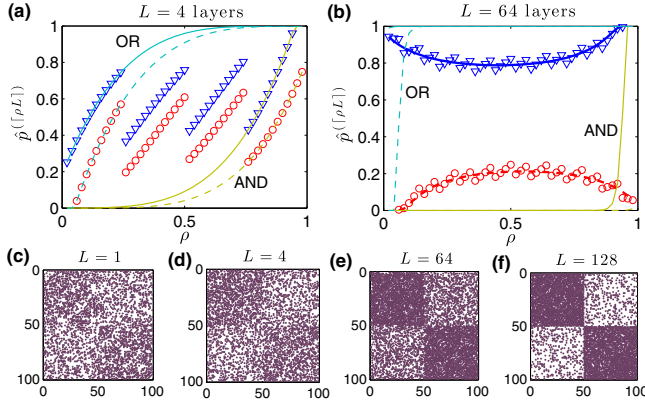


FIG. 2. Effective edge probabilities for layer aggregation at an optimal threshold. (a),(b) The summation and thresholding at  $\tilde{L} = \lceil \rho L \rceil$  of  $L$  adjacency matrices yields a new SBM with effective edge probabilities  $\hat{p}_{\text{in}}^{(\lceil \rho L \rceil)}$  (blue triangles) and  $\hat{p}_{\text{out}}^{(\lceil \rho L \rceil)}$  (red circles). Results are for  $\Delta = 0.1$  (i.e.,  $p_{\text{in,out}} = \rho \pm 0.05$ ) with (a)  $L = 4$  and (b)  $L = 64$  layers. We also show effective probabilities for the AND (gold curves) and OR (cyan curves) networks. (Solid and dashed curves give  $\hat{p}_{\text{in}}^{(\tilde{L})}$  and  $\hat{p}_{\text{out}}^{(\tilde{L})}$ , respectively.) Note for the larger  $L$  value in (b) that  $\hat{p}_{\text{in}}^{(\lceil \rho L \rceil)}$  and  $\hat{p}_{\text{out}}^{(\lceil \rho L \rceil)}$  have separated and aligned with the asymptotic probabilities  $\hat{p}_{\text{in}}^{(\text{asym})}$  (blue solid curve) and  $\hat{p}_{\text{out}}^{(\text{asym})}$  (red dashed curve), respectively. (c)–(f) Adjacency matrices of thresholded networks with  $\rho = 0.3$ ,  $\Delta = 0.1$ ,  $\tilde{L} = \lceil \rho L \rceil$  and various  $L$ .

first-order expansion  $\text{erf}^{-1}(N^{-1/2}) \approx \sqrt{\pi/4N}$ . Importantly, Eq. (8) implies that  $\Delta^*$  decays as  $O(1/\sqrt{NL})$  for thresholded networks with  $\tilde{L} = \lceil \rho L \rceil$ .

In Fig. 2, we illustrate the limiting  $L \rightarrow \infty$  behavior for thresholded networks with  $\tilde{L} = \lceil \rho L \rceil$ . In panels (a) and (b), we plot  $\hat{p}_{\text{in}}^{(\lceil \rho L \rceil)}$  (blue triangles) and  $\hat{p}_{\text{out}}^{(\lceil \rho L \rceil)}$  (red circles) versus  $\rho$  for  $\Delta = 0.1$  with (a)  $L = 4$  and (b)  $L = 64$ . We also plot the effective probabilities  $\hat{p}_{\text{in}}^{(\tilde{L})}$  (solid curves) and  $\hat{p}_{\text{out}}^{(\tilde{L})}$  (dashed curves) for the AND (gold curves) and OR (cyan curves) networks. In panel (b), we additionally plot the limiting effective probabilities  $\hat{p}_{\text{in}}^{(\text{asym})}$  (blue solid curve) and  $\hat{p}_{\text{out}}^{(\text{asym})}$  (red dashed curve). Comparing panel (b) to (a), one can observe that as  $L$  increases, the piecewise-continuous solutions  $\hat{p}_{\text{in,out}}^{(\lceil \rho L \rceil)}$  separate and align with the respective asymptotic solutions  $\hat{p}_{\text{in,out}}^{(\text{asym})}$ .

In Figs. 2(c)–(f), we illustrate adjacency matrices  $\hat{\mathbf{A}}^{(\lceil \rho L \rceil)}$  of thresholded networks with  $\rho = 0.3$  and  $\Delta = 0.1$  for various  $L$ . We note that the community structure is undetectable for  $L = 1$  since  $\Delta^* = 0.1095$ , whereas it is detectable (and visually apparent) for  $L = 128$ . Comparing (c)–(f) illustrates the  $L \rightarrow \infty$  limiting behavior of  $\hat{\mathbf{A}}^{(\lceil \rho L \rceil)}$ . Specifically, application of Hoeffding’s inequality [43] (and using that  $p_{\text{in,out}} - \rho = \pm \Delta/2$ ) yields  $p_{\text{in}}^{(\lceil \rho L \rceil)} \geq 1 - e^{-L\Delta^2/2}$  and  $p_{\text{out}}^{(\lceil \rho L \rceil)} \leq e^{-L\Delta^2/2}$ , which implies that  $\hat{p}_{\text{in}}^{(\lceil \rho L \rceil)} \rightarrow 1$  and

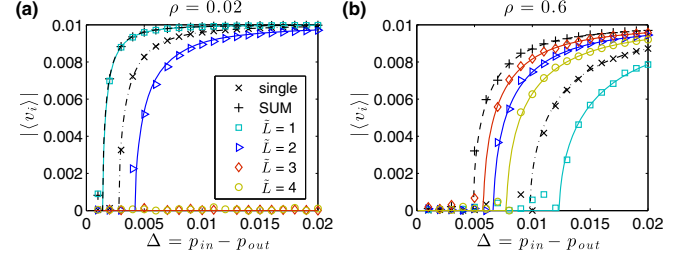


FIG. 3. Phase transition at  $\Delta^*$  for the dominant eigenvector  $\mathbf{v}$  of the modularity matrix. We show observed (symbols) and predicted values given by Eq. (9) (curves) for the mean eigenvector entry  $|\langle v_i \rangle|$  within a community for  $N = 10^4$  and  $L = 4$ .

$\hat{p}_{\text{out}}^{(\lceil \rho L \rceil)} \rightarrow 0$  with increasing  $L$  so that  $\hat{A}_{ij}^{(\lceil \rho L \rceil)} \rightarrow \delta_{c_i c_j}$ , where  $\delta_{nm}$  is the Kronecker delta function.

We conclude by studying the dominant eigenvector  $\mathbf{v}$  of the appropriate modularity matrix, which undergoes a phase transition at  $\Delta^*$ :  $\{v_i\}$  and the community labels  $\{c_i\}$  are uncorrelated for  $\Delta < \Delta^*$ , whereas they are correlated for  $\Delta > \Delta^*$ . Using methodology developed in [38], we find that the entries  $\{v_i\}$  within a community are Gaussian distributed with mean

$$|\langle v_i \rangle| = \sqrt{\frac{1}{N}} \sqrt{1 - \frac{\lambda_2^2}{(NL\Delta)^2}}, \quad (9)$$

which we use as an order parameter to observe the phase transition. In Fig. 3, we depict observed (symbols) and predicted values given by Eq. (9) (curves) of  $|\langle v_i \rangle|$  for a single layer ( $\times$  symbols), the summation network ( $+$  symbols), and thresholded networks (open symbols). We focus on a range of  $\Delta$  that contains  $\Delta^*$  for most aggregation methods. Note for the thresholded networks that there is no simple ordering to  $\Delta^*$ , which can be deduced by examining Fig. 1(a) for  $\rho \in \{0.02, 0.6\}$ . Finally, we note that finite-size effects amplify disagreement between observed and predicted values near the phase transitions.

In this Letter, we studied limitations on community detection for multilayer networks with layers drawn from a common SBM. As an illustrative model, we analyzed the effect of layer aggregation on the detectability limit  $\Delta^*$  for two equal-sized communities. When layers are aggregated by summation, we analytically showed that  $\Delta^*$  vanishes as  $O(L^{-1/2})$ . When layers are aggregated by thresholding this summation,  $\Delta^*$  depends on the choice of threshold,  $\tilde{L}$ . For  $\tilde{L} = \lceil \rho L \rceil$ , we analytically found  $\Delta^*$  to also vanish as  $O(L^{-1/2})$ . We note that our analysis also describes layer aggregation by taking the mean,  $L^{-1} \sum_l \mathbf{A}^{(l)}$ , since the multiplication of a matrix by a constant simply scales the eigenvalues by that constant. Thus, our results are in excellent agreement with previous work [35] that proved the consistency of spectral clustering via the mean adjacency matrix.

Finally, it is commonplace to threshold pairwise-interaction data to construct network representations that are sparse and unweighted and can be studied at a lower computational cost. Our research provides insight into this common—yet not well understood—practice. It would be interesting to extend this work to allow the SBMs of layers to be correlated [25] (that is, rather than identical) or organized into “strata” [17] (i.e., layers within a single stratum are similar, but they differ across strata). We are currently extending our analysis to hierarchical SBMs [27].

The authors were supported by the NIH (R01HD075712, T32GM067553, T32CA201159), the James S. McDonnell Foundation (#220020315) and the UNC Lineberger Comprehensive Cancer Center with funding provided by the University Cancer Research Fund via the State of North Carolina. The content does not necessarily represent the views of the funding agencies. We thank the reviewers for their helpful comments.

---

\*dane.r.taylor@gmail.com

- [1] M. E. J. Newman, *SIAM Rev.* **45**, 167 (2003).
- [2] J. Moody, *Am. Soc. Rev.* **69**, 213 (2004).
- [3] D. S. Bassett *et al.*, *Proc. Natl. Acad. of Sci. U.S.A.* **108**, 7641 (2011).
- [4] S. Boccaletti, G. Bianconi, R. Criado, C. I. del Genio, J. Gómez-Gardeñes, M. Romance, I. Sendiña-Nadal, Z. Wang, and M. Zanin, *Phys. Rep.* **544**, 1 (2014).
- [5] M. Kivelä, A. Arenas, M. Barthélemy, J. P. Gleeson, Y. Moreno, and M. A. Porter, *J. Complex Netw.* **2**, 203 (2014).
- [6] D. Krackhardt, *Soc. Networks* **9**, 109 (1987).
- [7] Y. Y. Haimes and P. Jiang, *J. Infrast. Sys.* **7**, 1 (2001).
- [8] P. Holme and J. Saramäki, *Phys. Rep.* **519**, 97 (2012).
- [9] P. J. Mucha, T. Richardson, K. Macon, M. A. Porter, and J.-P. Onnela, *Science* **328**, 876 (2010).
- [10] R. J. Sánchez-García, E. Cozzo, and Y. Moreno, *Phys. Rev. E* **89**, 052815 (2014).
- [11] A. Sole-Ribalta, M. De Domenico, N. E. Kouvaris, A. Díaz-Guilera, S. Gómez, and A. Arenas, *Phys. Rev. E* **88**, 032807 (2013).
- [12] C. D. Brummitt, R. M. D’Souza, and E. A. Leicht, *Proc. Natl. Acad. Sci. U.S.A.* **109**, E680 (2012).
- [13] A. Bashan, Y. Berezin, S. V. Buldyrev, and S. Havlin, *Nat. Phys.* **9**, 667 (2013).
- [14] F. Radicchi and A. Arenas, *Nat. Phys.* **9**, 717 (2013).
- [15] G. Menichetti, D. Remondini, and G. Bianconi, *Phys. Rev. E* **90**, 062817 (2014).
- [16] M. De Domenico, V. Nicosia, A. Arenas, V. Latora, *Nat. Commun.* **6**, 6864 (2015).
- [17] N. Stanley, S. Shai, D. Taylor, and P. J. Mucha, [arXiv:1507.01826](https://arxiv.org/abs/1507.01826).
- [18] J.-P. Onnela, D. J. Fenn, S. Reid, M. A. Porter, P. J. Mucha, M. D. Fricker, and N. S. Jones, *Phys. Rev. E* **86**, 036104 (2012).
- [19] A. Lancichinetti, S. Fortunato, and F. Radicchi, *Phys. Rev. E* **78**, 046110 (2008).
- [20] A. Lancichinetti and S. Fortunato, *Phys. Rev. E* **84**, 066122 (2011).
- [21] J. Reichardt and M. Leone, *Phys. Rev. Lett.* **101**, 078701 (2008).
- [22] D. Hu, P. Ronhovde, and Z. Nussinov, *Philo. Mag.* **92**, 406 (2012).
- [23] A. Decelle, F. Krzakala, C. Moore, and L. Zdeborová, *Phys. Rev. Lett.* **107**, 065701 (2011).
- [24] R. R. Nadakuditi and M. E. J. Newman, *Phys. Rev. Lett.* **108**, 188701 (2012).
- [25] E. Abbe, A. S. Bandeira, and G. Hall, *IEEE Trans. Inf. Theory*, **62**, 471 (2016).
- [26] F. Radicchi, *Phys. Rev. E* **88**, 010801 (2013).
- [27] T. P. Peixoto, *Phys. Rev. Lett.* **111**, 098701 (2013).
- [28] S. Sarkar, J. A. Henderson, and P. A. Robinson, *PLoS One*, **8**, e54383 (2013).
- [29] A. Ghasemian *et al.*, [arXiv:1506.06179](https://arxiv.org/abs/1506.06179).
- [30] T. Kawamoto and Y. Kabashima, *Europhys. Lett.* **112**, 40007 (2015).
- [31] T. Valles-Catala, F. A. Massucci, R. Guimera, and M. Sales-Pardo, *Phys. Rev. X* **6**, 011036 (2016).
- [32] S. Paul and Y. Chen, [arXiv:1506.02699](https://arxiv.org/abs/1506.02699).
- [33] P. Barbillon, S. Donnet, E. Lazega, and A. Bar-Hen, [arXiv:1501.06444](https://arxiv.org/abs/1501.06444).
- [34] T. P. Peixoto, *Phys. Rev. E* **92**, 042807 (2015).
- [35] Q. Han, K. S. Xu, and E. M. Airoldi, *Proceedings of the 32nd International Conference on Machine Learning, Lille, France, 2015* (JMLR: W&CP, 2015), Vol. 37, p. 1511.
- [36] C. Aicher, A. Z. Jacobs, and A. Clauset, *J. Complex Netw.* **3**, 221 (2015).
- [37] M. P. Rombach, M. A. Porter, J. H. Fowler, and P. J. Mucha, *SIAM J. Appl. Math.* **74**, 167 (2014).
- [38] F. Benaych-Georges and R. R. Nadakuditi, *Adv. Math.* **227**, 494 (2011).
- [39] R. R. Nadakuditi and M. E. J. Newman, *Phys. Rev. E* **87**, 012803 (2013).
- [40] M. E. J. Newman and M. Girvan, *Phys. Rev. E* **69**, 026113 (2004).
- [41] See Supplemental Material at <http://link.aps.org/supplemental/10.1103/PhysRevLett.116.228301> for a more detailed mathematical analysis of the eigenspectra and phase transition in detectability.
- [42] O. Kallenberg, *Foundations of Modern Probability* (Springer Science & Business Media, New York, 2006).
- [43] W. Hoeffding, *J. Am. Stat. Assoc.* **58**, 13 (1963).