OPTIMAL ADAPTIVITY OF SIGNED-POLYGON STATISTICS FOR NETWORK TESTING

By Jiashun Jin[1], Zheng Tracy Ke[2] and Shengming Luo[1]

[1]Department of Statistics & Data Science, Carnegie Mellon University, jiashun@stat.cmu.edu; shengmil@andrew.cmu.edu

[2]Department of Statistics, Harvard University, zke@fas.harvard.edu

Given a symmetric social network, we are interested in testing whether it has only one community or multiple communities. The desired tests should (a) accommodate severe degree heterogeneity, (b) accommodate mixed-memberships, (c) have a tractable null distribution, and (d) adapt automatically to different levels of sparsity, and achieve the optimal phase diagram. How to find such a test is a challenging problem.

We propose the Signed Polygon as a class of new tests. Fixing $m \geq 3$, for each m-gon in the network, define a score using the centered adjacency matrix. The sum of such scores is then the m-th order Signed Polygon statistic. The Signed Triangle (SgnT) and the Signed Quadrilateral (SgnQ) are special examples of the Signed Polygon.

We show that both the SgnT and SgnQ tests satisfy (a)-(d), and especially, they work well for both very sparse and less sparse networks. Our proposed tests compare favorably with existing tests. For example, the EZ and GC tests behave unsatisfactorily in the less sparse case and do not achieve the optimal phase diagram. Also, many existing tests do not allow for severe heterogeneity or mixed-memberships, and they behave unsatisfactorily in our settings.

The analysis of the SgnT and SgnQ tests is delicate and extremely tedious, and the main reason is that we need a unified proof that covers a wide range of sparsity levels and a wide range of degree heterogeneity. For lower bound theory, we use a phase transition framework, which includes the standard minimax argument, but is more informative. The proof uses classical theorems on matrix scaling.

1. Introduction.

Given a symmetrical social network, we are interested in the *global testing problem* where we use the adjacency matrix of the network to test whether it has only one community or multiple communities. A good understanding of the problem is useful for discovering non-obvious social groups and patterns [5, 14], measuring diversity of individual nodes [15], determining stopping time in a recursive community detection scheme [32, 43]. It may also help understand other related problems such as membership estimation [42] and estimation of the number of communities [39, 41].

Natural networks have several characteristics that are ubiquitously found:

- *Severe degree heterogeneity*. The distribution of the node degrees usually has a power-law tail, implying severe degree heterogeneity.
- *Mixed-memberships*. Communities are tightly woven clusters of nodes where we have more edges within than between [17, 38]. Communities are rarely non-overlapping, and some nodes may belong to more than one community (and thus have mixed-memberships).
- *Sparsity*. Many networks are sparse. The sparsity levels may range significantly from one network to another, and may also range significantly from one node to another (due to severe degree heterogeneity).

MSC2020 subject classifications: Primary 62H30, 91C20; secondary 62P25..

Keywords and phrases: asymptotic normality, DCBM, DCMM, lower bound, matrix scaling, optimal phase diagram, phase transition, SBM, Signed Quadrilateral, Signed Triangle, Sinkhorn's theorem, sparsity.

Phase transition is a well-known optimality framework [13, 22, 33, 37]. It is related to the minimax framework but can be more informative in many cases. Conceptually, for the global testing problem, in the two-dimensional phase space with the two axes calibrating the "sparsity" and "signal strength," respectively, there is a "Region of Possibility" and a "Region of Impossibility." In "Region of Possibility," any alternative is separable from the null. In "Region of Impossibility," any alternative is inseparable from the null.

If a test is able to automatically adapt to different levels of sparsity and separate any given alternative in the "Region of Possibility" from the null, then we call it "optimally adaptive."

We are interested in finding tests that satisfy the following requirements.

(R1) Applicable to networks with severe degree heterogeneity.
(R2) Applicable to networks with mixed-memberships.
(R3) The asymptotic null distribution is easy to track, so the rejection regions are easy to set.
(R4) Optimally adaptive: We desire a single test that is able to adapt to different levels of sparsity and is optimally adaptive.

1.1. *The DCMM model.* We adopt the *Degree Corrected Mixed Membership (DCMM)* model [42, 24]. Denote the adjacency matrix by A, where

$$A_{ij} = \begin{cases} 1, & \text{if node } i \text{ and node } j \text{ have an edge,} \\ 0, & \text{otherwise.} \end{cases} \tag{1.1}$$

Conventionally, self-edges are not allowed so all the diagonal entries of A are 0. In DCMM, we assume there are K perceivable communities $\mathcal{C}_1, \mathcal{C}_2, \ldots, \mathcal{C}_K$, and each node is associated with a mixed-membership weight vector $\pi_i = (\pi_i(1), \pi_i(2), \ldots, \pi_i(K))'$ where for $1 \leq k \leq K$ and $1 \leq i \leq n$,

$$\pi_i(k) = \text{the weight node } i \text{ puts on community } k. \tag{1.2}$$

Moreover, for a $K \times K$ symmetric nonnegative matrix P which models the community structure, and positive parameters $\theta_1, \theta_2, \ldots, \theta_n$ which model the degree heterogeneity, we assume the upper triangular entries of A are independent Bernoulli variables satisfying

$$\mathbb{P}(A_{ij} = 1) = \theta_i \theta_j \cdot \pi_i' P \pi_j \equiv \Omega_{ij}, \qquad 1 \leq i < j \leq n, \tag{1.3}$$

where Ω denotes the matrix $\Theta \Pi P \Pi' \Theta$, with Θ being the $n \times n$ diagonal matrix $\mathrm{diag}(\theta_1, \ldots, \theta_n)$ and Π being the $n \times K$ matrix $[\pi_1, \pi_2, \ldots, \pi_n]'$. For identifiability (see [24] for more discussion), we assume

$$\text{all diagonal entries of } P \text{ are } 1. \tag{1.4}$$

When $K = 1$, (1.4) implies $P = 1$, and so $\Omega_{ij} = \theta_i \theta_j$, $1 \leq i, j \leq n$.

Write for short $\mathrm{diag}(\Omega) = \mathrm{diag}(\Omega_{11}, \Omega_{22}, \ldots, \Omega_{nn})$, and let W be the matrix where for $1 \leq i, j \leq n$, $W_{ij} = A_{ij} - \Omega_{ij}$ if $i \neq j$ and $W_{ij} = 0$ otherwise. In matrix form, we have

$$A = \Omega - \mathrm{diag}(\Omega) + W, \qquad \text{where} \quad \Omega = \Theta \Pi P \Pi' \Theta. \tag{1.5}$$

DCMM includes three models as special cases, each of which is well-known and has been studied extensively recently.

- *Degree Corrected Block Model (DCBM)* [28]. If we do not allow mixed-memberships (i.e., each weight vector π_i is degenerate with one entry being nonzero), then DCMM reduces to the DCBM.
- *Mixed Membership Stochastic Block Model (MMSBM)* [1]. DCBM further reduces to MMSBM if $\theta_1 = \ldots = \theta_n (= \sqrt{\alpha_n})$. In this special cse, $\Omega = \alpha_n \Pi P \Pi'$, and for identifiability, (1.4) is too strong, so we relax it to that the average of the diagonals of P is 1.

- *Stochastic Block Model (SBM)* [20]. MMSBM further reduces to the classical SBM if additionally we do not allow mixed-memberships.

Under DCMM, the global testing problem is the problem of testing

$$H_0^{(n)}: K = 1 \qquad \text{vs.} \qquad H_1^{(n)}: K \geq 2. \tag{1.6}$$

The seeming simplicity of the two hypotheses is deceiving, as both of them are highly composite, consisting of many different parameter configurations.

1.2. *Phase transition: a preview of our main results.* Let $\lambda_1, \lambda_2, \ldots, \lambda_K$ be the first K eigenvalues of Ω, arranged in the descending order in magnitude. We can view (a) $\sqrt{\lambda_1}$ both as the sparsity level and the noise level [23] (i.e., spectral norm of the noise matrix W), (b) $|\lambda_2|$ as the signal strength, so that $|\lambda_2|/\sqrt{\lambda_1}$ is the Signal-to-Noise Ratio (SNR), and (c) $|\lambda_2|/\lambda_1$ as a measure of dissimilarity between different communities (Example 1 below illustrates why it measures 'dissimilarity'). We note that [19, 12] also pointed out that $|\lambda_2|/\sqrt{\lambda_1}$ is a reasonable metric of SNR.

Now, in the two-dimensional *phase space* where the x-axis is $\sqrt{\lambda_1}$ which measures the sparsity level, and the y-axis is $|\lambda_2|/\lambda_1$ which measures the community dissimilarity, we have two regions.

- *Region of Possibility ($1 \ll \sqrt{\lambda_1} \ll \sqrt{n}$, $|\lambda_2|/\sqrt{\lambda_1} \to \infty$).* For any alternative hypothesis in this region, it is possible to distinguish it from any null hypothesis, by the Signed Polygon tests to be introduced.
- *Region of Impossibility ($1 \ll \sqrt{\lambda_1} \ll \sqrt{n}$, $|\lambda_2|/\sqrt{\lambda_1} \to 0$).* In this region, any alternative hypothesis is inseparable from the null hypothesis, provided with some mild conditions.

See Figure 1 (left panel). Also, see Sections 2 and 3 for our main theorems on *Possibility* and *Impossibility*, respectively. Note that the figure is only for illustration purpose, where the cases of $|\lambda_2| = c_0\sqrt{\lambda_1}$ for some constant $c_0 > 0$ are compressed in the separating boundary of two regions (red curve). The Signed Polygon test satisfies all requirements (R1)-(R4) above. Since the test is able to separate all alternatives (ranging from very sparse to less sparse) in the Region of Possibility from the null, it is *optimally adaptive*.

Remark 1. A stronger version of the phase transition is that for a constant $c_0 > 0$, the Region of Possibility and Region of Impossibility are given by $|\lambda_2|/\sqrt{\lambda_1} > c_0$ and $|\lambda_2|/\sqrt{\lambda_1} < c_0$, respectively. For the broad setting we consider, this is an open problem, though for some special cases, there are some interesting works (e.g., [19]); see Remark 11.

It is instructive to consider a special DCMM model, which is a generalization of the symmetric SBM [36] to the case with degree heterogeneity.

Example 1 *(A special DCMM).* Let $e_1, \ldots, e_K$ be the standard basis of $\mathbb{R}^K$. Fixing a positive vector $\theta \in \mathbb{R}^n$ and a scalar $b_n \in (0,1)$, we assume

$$P = (1 - b_n)I_K + b_n 1_K 1_K', \qquad \pi_i \text{ are iid sampled from } e_1, \ldots, e_K. \tag{1.7}$$

In this model, $(1 - b_n)$ measures the "dissimilarity" between different communities (it quantifies how well we can tell whether two nodes i and j are from the same community or not; note that $b_n = 1$ corresponds to the null case where all communities are indistinguishable) and $\|\theta\|$ measures the sparsity level. In this model, $\lambda_1 \sim (1 + (K-1)b_n)\|\theta\|^2$ and $\lambda_k \sim (1 - b_n)\|\theta\|^2$, $2 \leq k \leq K$. The sparsity level is $\sqrt{\lambda_1} \asymp \|\theta\|$, the community dissimilarity is characterized by $\lambda_2/\lambda_1 \asymp (1 - b_n)$, and the SNR is $|\lambda_2|/\sqrt{\lambda_1} \asymp \|\theta\|(1 - b_n)$. The Region of Possibility and Region of Impossibility are given by $\{1 \ll \|\theta\| \ll \sqrt{n}, \|\theta\|(1 - b_n) \to \infty\}$ and $\{1 \ll \|\theta\| \ll \sqrt{n}, \|\theta\|(1 - b_n) \to 0\}$, respectively. See Figure 1 (right panel).

Remark 2. As the phase transition is hinged on $\lambda_2/\sqrt{\lambda_1}$, one may think that the statistic $\hat{\lambda}_2/\sqrt{\hat{\lambda}_1}$ is optimally adaptive, where $\hat{\lambda}_k$ is the k-th largest (in magnitude) eigenvalue of A. This is however not true, because the consistency of $\hat{\lambda}_2$ for estimating λ_2 can not be guaranteed in our range of interest, unless with strong conditions on θ_{max} [23].

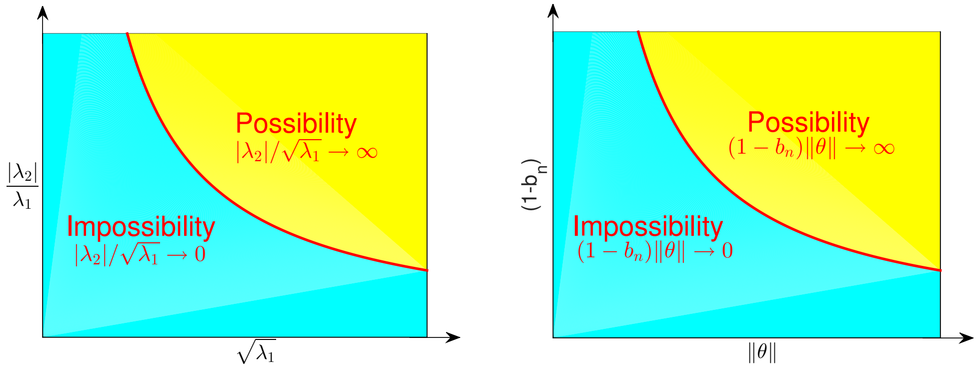
Possibility
$|\lambda_2|/\sqrt{\lambda_1} \to \infty$
Impossibility
$|\lambda_2|/\sqrt{\lambda_1} \to 0$
$\frac{|\lambda_2|}{\lambda_1}$
$\sqrt{\lambda_1}$
Possibility
$(1-b_n)\|\theta\| \to \infty$
Impossibility
$(1-b_n)\|\theta\| \to 0$
(1-b_n)
$\|\theta\|$

FIG 1. *Left: Phase transition. In Region of Impossibility, any alternative hypothesis is indistinguishable from a null hypothesis, provided that some mild conditions hold. In Region of Possibility, the Signed Polygon test is able to separate any alternative hypothesis from a null hypothesis asymptotically. Right: Phase transition for the special DCMM model in Example 1, where $\sqrt{\lambda_1} \asymp \|\theta\|$, $|\lambda_2|/\lambda_1 \asymp (1-b_n)$, and $|\lambda_2|/\sqrt{\lambda_1} \asymp (1-b_n)\|\theta\|$.*

1.3. *Literature review, the Signed Polygon, and our contribution.* Recently, the global testing problem has attracted much attention and many interesting approaches have been proposed. To name a few, Mossel et al. [36] and Banerjee and Ma [3] (see also [4]) considered a special case of the testing problem, where they assume a simple null of Erdos-Renyi random graph model and a special alternative which is an SBM with two equal-sized communities. They provided the asymptotic distribution of the log-likelihood ratio within the contiguous regime. Since the likelihood ratio test statistic is NP-hard to compute, [3] introduced an approximation by linear spectral statistics. Lei [31] also considered the SBM model and studied the problem of testing whether $K = K_0$ or $K > K_0$, where K_0 is a pre-specified integer. His approach is based on the Tracy-Widom law of extreme eigenvalues and requires delicate random matrix theory. Unfortunately, these works have been focused on the SBM (which allows neither severe degree heterogeneity nor mixed membership). Therefore, despite the elegant theory in these works, it remains unclear how to extend their ideas to our settings.

Along a different line, graphlet counts (GC) have been frequently used for hypothesis testing in non-parametric and parametric network models. This includes the EZ test [16] and GC test [25]. Other interesting works include [6, 7, 35]. In particular, [25] suggested a general recipe for constructing test statistics and showed that both GC and EZ tests have competitive power in a broad setting. Unfortunately, it turns out that in the less sparse case, the variance of the GC test statistic is much larger than expected, which largely hurts the power of the test. The underlying reason is that GC tests use *non-centered* cycle counts. If, however, we use *centered* cycle counts, we can largely reduce the variances and have a more powerful test. A similar phenomenon was discovered by Bubeck et al. [10] for the SBM setting.

This motivate a class of new tests which we call *Signed Polygon*, including the Signed Triangle (SgnT) and the Signed Quadrilateral (SgnQ). The Signed Polygon statistics are related to the Signed Cycle statistics, first introduced by Bubeck et al. [10] and later generalized by Banerjee [2]. Both Signed Polygon and Signed Cycle recognize that using centered-cycle counts may help reduce the variance, but there are some major differences. The study of the Signed Cycles has been focused on the SBM and similar models, where under the null, $\mathbb{P}(A_{ij} = 1) = \alpha$, $1 \leq i \neq j \leq n$, and α is the only unknown parameter. In this case, a natural approach to centering the adjacency matrix A is to first estimate α using the whole matrix A (say, $\hat{\alpha}$), and then subtract all off-diagonal entries of A by $\hat{\alpha}$. However, under the null of our setting, $\mathbb{P}(A_{ij} = 1) = \theta_i\theta_j$, $1 \leq i \neq j \leq n$, and there are n different unknown parameters $\theta_1, \theta_2, \ldots, \theta_n$. In this case, how to center the matrix A is not only unclear but also *worrisome*, especially when the network is very sparse, because we have to use limited data to estimate a large number of unknown parameters. Also, for any approaches we may have, the analysis is seen to be much harder than that of the previous case. Note that the ways how two statistics are defined over the centered adjacency matrix are also different; see Section 1.4 and [10, 2].

In the Signed Polygon, we use a new approach to estimate $\theta_1, \theta_2, \ldots, \theta_n$ under the null, and use the estimates to center the matrix A. To our surprise, data limitation (though a challenge) does not ruin the idea: even for very sparse networks, the estimation errors of $\theta_1, \theta_2, \ldots, \theta_n$ only have a negligible effect. The main contributions of the paper are as follows.

- Discover the phase transition for global testing in the broad DCMM setting by identifying the Regions of Impossibility and Possibility.
- Propose the Signed Polygon as a class of new tests that are appropriate for networks with severe degree heterogeneity and mixed-memberships.
- Prove that the Signed Triangle and Signed Quadrilateral tests satisfy all the requirements (R1)-(R4), and especially that they are optimally adaptive and perform well for all networks in the Region of Possibility, ranging from very sparse ones to the least sparse ones.

To show the success of the Signed Polygon test for the whole Region of Possibility is very subtle and extremely tedious. The main reason is that we hope to cover the *whole spectrum* of degree heterogeneity and sparsity levels. Crude bounds may work in one case but not another, and many seemingly negligible terms turn out to be non-negligible (see Sections 1.4 and 4). The lower bound argument is also very subtle. Compared to work on SBM where there is only one unknown parameter under the null, our null has n unknown parameters. The difference provides a lot of freedom in constructing inseparable hypothesis pairs, and so the Region of Impossibility in our setting is much wider than that for SBM. Our construction of inseparable hypothesis pairs uses theorems on non-negative matrix scaling, a mathematical area pioneered by Sinkhorn [40] and Olkin [34] among others (e.g., [9, 27]).

1.4. *The Signed Polygon statistic.* Recall that A is the adjacency matrix of the network. Introduce a vector $\hat{\eta}$ by ($\mathbf{1}_n$ denotes the vector of 1's)

$$\hat{\eta} = (1/\sqrt{V})\, A\mathbf{1}_n, \qquad \text{where } V = \mathbf{1}_n' A \mathbf{1}_n. \tag{1.8}$$

Fixing $m \geq 3$, the order-m *Signed Polygon* statistic is defined by (notation: $(dist)$ is short for "distinct", which means any two of $i_1, \ldots, i_m$ are unequal)

$$U_n^{(m)} = \sum_{i_1, i_2, \ldots, i_m (dist)} (A_{i_1 i_2} - \hat{\eta}_{i_1}\hat{\eta}_{i_2})(A_{i_2 i_3} - \hat{\eta}_{i_2}\hat{\eta}_{i_3}) \ldots (A_{i_m i_1} - \hat{\eta}_{i_m}\hat{\eta}_{i_1}). \tag{1.9}$$

When $m = 3$, we call it the Signed-Triangle (SgnT) statistic:

$$T_n = \sum_{i_1, i_2, i_3 (dist)} (A_{i_1 i_2} - \hat{\eta}_{i_1}\hat{\eta}_{i_2})(A_{i_2 i_3} - \hat{\eta}_{i_2}\hat{\eta}_{i_3})(A_{i_3 i_1} - \hat{\eta}_{i_3}\hat{\eta}_{i_1}). \tag{1.10}$$

When $m = 4$, we call it the Signed-Quadrilateral (SgnQ) statistic:

$$Q_n = \sum_{i_1, i_2, i_3, i_4 (dist)} (A_{i_1 i_2} - \hat{\eta}_{i_1}\hat{\eta}_{i_2})(A_{i_2 i_3} - \hat{\eta}_{i_2}\hat{\eta}_{i_3})(A_{i_3 i_4} - \hat{\eta}_{i_3}\hat{\eta}_{i_4})(A_{i_4 i_1} - \hat{\eta}_{i_4}\hat{\eta}_{i_1}). \tag{1.11}$$

For analysis, we focus on T_n and Q_n, but our main results are extendable to general m.

The key to understanding and analyzing the Signed Polygon is the *Ideal Signed Polygon*. Introduce a *non-stochastic counterpart* of $\hat{\eta}$ by

$$\eta^* = (1/\sqrt{v_0})\, \Omega \mathbf{1}_n, \qquad \text{where } v_0 = \mathbf{1}_n' \Omega \mathbf{1}_n. \tag{1.12}$$

Define the order-m *Ideal Signed Polygon* statistic by

$$\widetilde{U}_n^{(m)} = \sum_{i_1, i_2, \ldots, i_m (dist)} (A_{i_1 i_2} - \eta^*_{i_1}\eta^*_{i_2})(A_{i_2 i_3} - \eta^*_{i_2}\eta^*_{i_3}) \ldots (A_{i_m i_1} - \eta^*_{i_m}\eta^*_{i_1}). \tag{1.13}$$

We expect to see that $\hat{\eta} \approx \mathbb{E}[\hat{\eta}] \approx \eta^*$. We can view $\widetilde{U}_n^{(m)}$ as the oracle version of $U_n^{(m)}$, with η^* given. We can also view $U_n^{(m)}$ as the *plug-in* version of $\widetilde{U}_n^{(m)}$, where we replace η^* by $\hat{\eta}$.

For implementation, it is desirable to rewrite T_n and Q_n in matrix forms, which allows us to avoid using an for-loop and compute much faster (say, in MATLAB or R). For any two matrices $M, N \in \mathbb{R}^{n,n}$, let $\mathrm{tr}(M)$ be the trace of M, $\mathrm{diag}(M) = \mathrm{diag}(M_{11}, M_{22}, \ldots, M_{nn})$, and $M \circ N$ be the Hadamard product of M and N (i.e., $M \circ N \in \mathbb{R}^{n,n}$, $(M \circ N)_{ij} = M_{ij}N_{ij}$). Denote $\widetilde{A} = A - \hat{\eta}\hat{\eta}'$. The following theorem is proved in the supplementary material.

THEOREM 1.1. *We have* $T_n = \mathrm{tr}(\widetilde{A}^3) - 3\mathrm{tr}(\widetilde{A} \circ \widetilde{A}^2) + 2\mathrm{tr}(\widetilde{A} \circ \widetilde{A} \circ \widetilde{A})$ *and* $Q_n = \mathrm{tr}(\widetilde{A}^4) - 4\mathrm{tr}(\widetilde{A} \circ \widetilde{A}^3) + 8\mathrm{tr}(\widetilde{A} \circ \widetilde{A} \circ \widetilde{A}^2) - 6\mathrm{tr}(\widetilde{A} \circ \widetilde{A} \circ \widetilde{A} \circ \widetilde{A}) - 2\mathrm{tr}(\widetilde{A}^2 \circ \widetilde{A}^2) + 2 \cdot 1_n'[\mathrm{diag}(\widetilde{A})(\widetilde{A} \circ \widetilde{A})\mathrm{diag}(\widetilde{A})]1_n + 1_n'[\widetilde{A} \circ \widetilde{A} \circ \widetilde{A} \circ \widetilde{A}]1_n$. *The complexity of computing both* T_n *and* Q_n *is* $O(n^2\bar{d})$*, where* $\bar{d}$ *is the average degree of the network.*

Compared to the EZ and GC tests [16, 25], the computational complexity of SgnT and SgnQ is of the same order.

Remark 3. The computational complexity of $U_n^{(m)}$ remains as $O(n^2\bar{d})$ for larger m. Similarly as that in Theorem 1.1, the main complexity of $U_n^{(m)}$ comes from computing $\widetilde{A}^m$. Since we can compute $\widetilde{A}^m$ with $\widetilde{A}^m = \widetilde{A}^{m-1}\widetilde{A}$ and recursive matrix multiplications, each time with a complexity of $O(n^2\bar{d})$, the overall complexity is $O(n^2\bar{d})$.

Remark 4 (*Connection to the Signed Cycle*). In the more idealized SBM or MMSBM model, we do not have degree heterogeneity, and $\Omega = \alpha_n \mathbf{1}_n \mathbf{1}_n'$ under the null, where α_n is the only unknown parameter. In this simple setting, it makes sense to estimate α_n by $\hat{\alpha}_n = \bar{d}/(n-1)$, where $\bar{d}$ is the average degree. This gives rise to the *Signed Cycle* statistics [2, 10]: $C_n^{(m)} = \sum_{i_1, i_2, \ldots, i_m (dist)} (A_{i_1 i_2} - \hat{\alpha}_n)(A_{i_2 i_3} - \hat{\alpha}_n) \ldots (A_{i_m i_1} - \hat{\alpha}_n)$. Bubeck et al. [10] first proposed $C_n^{(3)}$ for a global testing problem in a model similar to MMSBM. Although their test statistic is also called the Signed Triangle, it is different from our SgnT statistic (1.10), because their tests are only applicable to models without degree heterogeneity. The analysis of the Signed Polygon is also much more delicate than that of the Signed Cycle, as the error $(\hat{\alpha}_n - \alpha_n)$ is much smaller than the errors in $(\hat{\eta} - \eta^*)$.

It remains to understand (A) how the Signed Polygon manages to reduce variance, and (B) what are the analytical challenges.

Consider Question (A). We illustrate it with the Ideal Signed Polygon (1.13) and the null case. In this case, $\Omega = \theta\theta'$. It is seen $\eta^* = \theta$, $A_{ij} - \eta^*_i\eta^*_j = A_{ij} - \Omega_{ij} = W_{ij}$, for $i \neq j$ (see (1.5) for definition of W), and so $\widetilde{U}_n^{(m)} = \sum_{i_1, i_2, \ldots, i_m (dist)} W_{i_1 i_2} W_{i_2 i_3} \ldots W_{i_m i_1}$. Here, each term is an m-product of independent centered Bernoulli variables, and $W_{i_1 i_2} W_{i_2 i_3} \ldots W_{i_m i_1}$ and $W_{i_1' i_2'} W_{i_2' i_3'} \ldots W_{i_m' i_1'}$ are correlated only when $\{i_1, i_2, \ldots, i_m\}$ and $\{i_1', i_2', \ldots, i_m'\}$ are the vertices of the same polygon. Such a construction is known to be efficient in variance reduction (e.g., [10]).

In comparison, for an order-m GC statistic [25], $N_n^{(m)} = \sum_{i_1, i_2, \ldots, i_m (dist)} A_{i_1 i_2} A_{i_2 i_3} \ldots A_{i_m i_1}$ is the main term. Since here the Bernoulli variables are not centered, we can split $N_n^{(m)}$ into two uncorrelated terms: $N_n^{(m)} = \widetilde{U}_n^{(m)} + (N_n^{(m)} - \widetilde{U}_n^{(m)})$. Compared to the Signed Polygon, the additional variance comes from the second term, which is undesirably large in the less sparse case [29].

Remark 5. The above also explains why the order-2 Signed Polygon does not work well. In fact, when $m = 2$, $\widetilde{U}_n^{(m)} = \sum_{i_1 \neq i_2} W_{i_1 i_2}^2$ under the null, which has an unsatisfactory variance due to the square of the W-terms.

Consider Question (B). We discuss with the SgnQ statistic. Recall that η^* is a non-stochastic proxy of $\hat{\eta}$. For any $1 \leq i, j \leq n$ and $i \neq j$, we decompose $\eta_i^* \eta_j^* - \hat{\eta}_i \hat{\eta}_j = \delta_{ij} + r_{ij}$, where δ_{ij} is the main term, which is a linear function of $\hat{\eta}_i$ and $\hat{\eta}_j$, and r_{ij} is the remainder term. Introduce

$$\widetilde{\Omega} = \Omega - \eta^* (\eta^*)'. \tag{1.14}$$

We have $A_{ij} - \hat{\eta}_i \hat{\eta}_j = \widetilde{\Omega}_{ij} + W_{ij} + \delta_{ij} + r_{ij}$. After inserting this into Q_n, each 4-product is now the product of 4 bracketed terms, where each bracketed term is the sum of 4 terms. Expanding the brackets and re-organizing, Q_n splits into $4 \times 4 \times 4 \times 4 = 256$ *post-expansion* sums, each having the form $\sum_{i_1, i_2, i_3, i_4 (dist)} a_{i_1 i_2} b_{i_2 i_3} c_{i_3 i_4} d_{i_4 i_1}$, where a is a generic term which can be equal to either of the four terms $\widetilde{\Omega}$, W, δ, and r; same for b, c and d. While some of these terms may be equal to each other, the symmetry we can exploit is limited, due to (a) degree heterogeneity, (b) mixed-memberships, and (c) the underlying polygon structure. As a result, we still have more than 50 post-expansion sums to analyze.

The analysis of a post-expansion sum with the presence of one or more r-term is the most tedious of all, where we need to further decompose each r-term into three different terms. This requires analysis of more than 100 additional post-expansion sums. We may think most of the post-expansion sums are easy to control via a crude bound (e.g., by Cauchy-Schwarz inequality). Unfortunately, this is not the case, and many seemingly negligible terms turn out to be non-negligible. Here are some of the reasons.

- We wish to cover most interesting cases. A crude bound may be enough for some cases but not for others.
- We desire to have a *single* test that achieves the phase transition for the whole range of interest. Alternatively, we may want to find several tests, each covering a subset of cases of interest, but this is less appealing.

As a result, we have to analyze a large number of post-expansion sums, where the analysis is subtle, extremely tedious, and error-prone, involving delicate combinatorics, due to the underlying polygon structure. See Section 4.

Remark 6. In Signed Polygon (1.9), we estimate Ω by $\hat{\eta}\hat{\eta}' = (\mathbf{1}_n' A \mathbf{1}_n)^{-1} A \mathbf{1}_n \mathbf{1}_n' A$ for the null. Alternatively, we may use a spectral approach and estimate Ω by $\hat{\lambda}_1 \hat{\xi}_1 \hat{\xi}_1'$, where $\hat{\lambda}_1$ and $\hat{\xi}_1$ are the first eigenvalue and eigenvector of A, respectively. Unfortunately, even in the more idealized SBM case, this estimate may be unsatisfactory for sparse networks (e.g., [11, Section 2.2]). In fact, for our main results to hold, we need to have $|\hat{\lambda}_1 - \lambda_1| \leq C\|\theta\|$ with large probability, but the best concentration inequality we have is $|\hat{\lambda}_1 - \lambda_1| \leq C\sqrt{\theta_{max}\|\theta\|_1}$ with large probability [24, Lemma C.1]. In the presence of severe degree heterogeneity, we often have $\sqrt{\theta_{max}\|\theta\|_1} \gg \|\theta\|$. Also, unlike $\hat{\eta}\hat{\eta}'$ in our proposal, $\hat{\lambda}_1 \hat{\xi}_1 \hat{\xi}_1'$ is not an explicit function of A, so the alternative version of the Signed Polygon statistic is much harder to analyze.

1.5. *Organization of the paper.* Section 2 focuses on the Region of Possibility and contains the upper bound argument. Section 3 focuses on the Region of Impossibility and contains the lower bound argument. Section 4 presents the key proof ideas, with the proof of secondary lemmas deferred to the supplementary material. Section 5 presents the numerical study, and Section 6 discusses extensions and connections.

For any $q > 0$ and $\theta \in \mathbb{R}^n$, $\|\theta\|_q$ denotes the ℓ^q-norm of θ (when $q = 2$, we drop the subscript for simplicity). Also, θ_{min} and θ_{max} denote $\min\{\theta_1, \ldots, \theta_n\}$ and $\max\{\theta_1, \ldots, \theta_n\}$, respectively. For any $n > 1$, $\mathbf{1}_n \in \mathbb{R}^n$ denotes the vector of 1's. For two positive sequences $\{a_n\}_{n=1}^{\infty}$ and $\{b_n\}_{n=1}^{\infty}$, we write $a_n \sim b_n$ if $\lim_{n\to\infty} a_n/b_n = 1$, and we write $a_n \asymp b_n$ if for sufficiently large n, there are two constants $c_2 > c_1 > 0$ such that $c_1 \leq a_n/b_n \leq c_2$. We use $\sum_{i_1,i_2,\ldots,i_m(dist)}$ to denote the sum over all $(i_1,\ldots,i_m)$ such that $1 \leq i_k \leq n$ and $i_k \neq i_\ell$ for $1 \leq k \neq \ell \leq m$. We use $C > 0$ as a generic constant that may vary from occurrence to occurrence. For constants that need to be more specific, we use c_0, c_1, etc.

2. The Signed Polygon test and the upper bound.

For reasons aforementioned, we focus on the SgnT statistic T_n and SgnQ statistic Q_n, but the ideas are extendable to general Signed Polygon statistics. In this section, we study the upper bound. In detail, in Section 2.1, we establish the asymptotic normality of both test statistics. In Sections 2.2-2.3, we discuss the power of the two tests. We show that if $|\lambda_2|/\sqrt{\lambda_1} \to \infty$ and some mild regularity conditions hold, then for each of the two tests, the sum of Type I and Type II errors tends to 0 as $n \to \infty$. The lower bound is studied in Section 3, where we show that for an alternative hypothesis setting with $|\lambda_2|/\sqrt{\lambda_1} \to 0$, we can always pair it with a null setting so that two hypotheses are asymptotically inseparable.

In a DCMM model, $\Omega = \Theta\Pi P \Pi'\Theta$, where $\Theta = \mathrm{diag}(\theta_1,\ldots,\theta_n)$, and Π is the $n \times K$ membership matrix $[\pi_1, \pi_2, \ldots, \pi_n]'$. We assume as $n \to \infty$,

$$\|\theta\| \to \infty, \quad \theta_{max} \to 0, \quad \text{and} \quad (\|\theta\|^2/\|\theta\|_1)\sqrt{\log(\|\theta\|_1)} \to 0. \tag{2.15}$$

The first condition is necessary. In fact, if $\|\theta\| \to 0$, then the alternative is indistinguishable from the null, as suggested by lower bounds in Section 3. The second one is mild as we usually assume $\theta_{max} \leq C$. This is due to that under DCMM, P has unit diagonal entries and $\theta_i\theta_j(\pi_i' P \pi_j)$ is a probability for all $i \neq j$. The last one is weaker than that of $\theta_{max}\sqrt{\log(n)} \to 0$, and is very mild. It is assumed mostly for technical reasons and is not required in many cases (e.g, the dense case where all $\theta_i = O(1)$). Moreover, introduce $G = \|\theta\|^{-2}\Pi'\Theta^2\Pi \in \mathbb{R}^{K\times K}$. This matrix is properly scaled and it can be shown that $\|G\| \leq 1$ (Appendix E, supplemental material). When the null is true, $K = P = G = 1$, and we do not need any additional condition. When the alternative is true, we assume

$$\frac{\max_{1\leq k\leq K}\{\sum_{i=1}^n \theta_i\pi_i(k)\}}{\min_{1\leq k\leq K}\{\sum_{i=1}^n \theta_i\pi_i(k)\}} \leq C, \qquad \|G^{-1}\| \leq C, \qquad \|P\| \leq C; \tag{2.16}$$

Here, $C > 0$ is a generic constant; see Section 1.5. The conditions are mild. Take the first two for example. When there is no mixed membership, they only require the K classes to be relatively balanced.

2.1. *Asymptotic normality of the null.*

Theorems 2.1-2.2 are proved in the supplement.

THEOREM 2.1 (Limiting null of the SgnT statistic). *Consider the testing problem (1.6) under the DCMM model (1.1)-(1.4), where the condition (2.15) is satisfied. Suppose the null hypothesis is true. As $n \to \infty$, $\mathbb{E}[T_n] = o(\|\theta\|^3)$, $\mathrm{Var}(T_n) \sim 6\|\theta\|^6$, and $(T_n - \mathbb{E}[T_n])/\sqrt{\mathrm{Var}(T_n)} \longrightarrow N(0,1)$ in law.*

THEOREM 2.2 (Limiting null of the SgnQ statistic). *Consider the testing problem (1.6) under the DCMM model (1.1)-(1.4), where the condition (2.15) is satisfied. Suppose the null hypothesis is true. As $n \to \infty$, $\mathbb{E}[Q_n] = (2 + o(1))\|\theta\|^4$, $\mathrm{Var}(Q_n) \sim 8\|\theta\|^8$, and $(Q_n - \mathbb{E}[Q_n])/\sqrt{\mathrm{Var}(Q_n)} \longrightarrow N(0,1)$ in law.*

Note that under the null, the limiting distributions of $T_n/\sqrt{\mathrm{Var}(T_n)}$ and $Q_n/\sqrt{\mathrm{Var}(Q_n)}$ are $N(0,1)$ and $N(1/\sqrt{2},1)$, respectively. To appreciate the difference, recall that the Signed

Polygon can be viewed as a plug-in statistic, where we replace η^* in the Ideal Signed Polygon by $\hat{\eta}$. Under the null, the effect of the plug-in is negligible for SgnT but not for SgnQ, so the two limiting distributions are different. See Section 4 for details.

2.2. *The level-α SgnT and SgnQ tests.* By Theorems 2.1 and 2.2, the null variances of the two statistics depend on $\|\theta\|^2$. To use the two statistics as tests, we need to estimate $\|\theta\|^2$. For $\hat{\eta}$ and η^* defined in (1.8) and (1.12), respectively, we have $\hat{\eta} \approx \eta^*$ and $\eta^* = \theta$ under the null. A reasonable estimator for $\|\theta\|^2$ under the null is therefore $\|\hat{\eta}\|^2$. We propose to estimate $\|\theta\|^2$ with $(\|\hat{\eta}\|^2 - 1)$, which corrects the bias and is slightly more accurate than $\|\hat{\eta}\|^2$. The following lemma is proved in the supplementary material.

LEMMA 2.1 (Estimation of $\|\theta\|^2$). *Consider the testing problem* (1.6) *under the DCMM model* (1.1)-(1.4)*, where the condition* (2.15) *holds when either hypothesis is true and condition (2.16) holds when the alternative is true. Then, under both hypotheses, as $n \to \infty$ $(\|\hat{\eta}\|^2 - 1)/\|\eta^*\|^2 \to 1$ in probability, where $\|\eta^*\|^2 = (\mathbf{1}_n'\Omega^2\mathbf{1}_n)/(\mathbf{1}_n'\Omega\mathbf{1}_n)$. Furthermore, $\|\eta^*\|^2 = \|\theta\|^2$ under $H_0^{(n)}$ and $\|\eta^*\|^2 \asymp \|\theta\|^2$ under $H_1^{(n)}$.*

Combining Lemma 2.1 with Theorem 2.1 gives

$$T_n/\sqrt{6(\|\hat{\eta}\|^2 - 1)^3} \longrightarrow N(0,1), \qquad \text{in law.} \tag{2.17}$$

Fix $\alpha \in (0,1)$. We propose the following SgnT test, which is a two-sided test where we reject the null hypothesis if and only if

$$|T_n| \geq z_{\alpha/2}\sqrt{6}(\|\hat{\eta}\|^2 - 1)^{3/2}, \quad z_{\alpha/2}\text{: upper } (\alpha/2)\text{-quantile of } N(0,1). \tag{2.18}$$

Similarly, combining Theorem 2.2 and Lemma 2.1, we have

$$[Q_n - 2(\|\hat{\eta}\|^2 - 1)^2]/\sqrt{8(\|\hat{\eta}\|^2 - 1)^4} \longrightarrow N(0,1), \qquad \text{in law.} \tag{2.19}$$

With the same α, we propose the following SgnQ test, which is a one-sided test where we reject the null hypothesis if and only if

$$Q_n \geq \big(2 + z_\alpha\sqrt{8}\big)(\|\hat{\eta}\|^2 - 1)^2, \quad z_\alpha\text{: upper } \alpha\text{-quantile of } N(0,1). \tag{2.20}$$

As a result, for both tests we just defined, the levels satisfy

$$\mathbb{P}_{H_0^{(n)}}(\text{Reject the null}) \to \alpha, \qquad \text{as } n \to \infty.$$

Figure 2 shows the histograms of $T_n/\sqrt{6(\|\hat{\eta}\|^2 - 1)^3}$ (left) and $(Q_n - 2(\|\hat{\eta}\|^2 - 1)^2)/(\sqrt{8(\|\hat{\eta}\|^2 - 1)^4})$ (right) under a null and an alternative simulated from DCMM. Recall that in DCMM, $\Omega = \theta\theta'$ under the null and $\Omega = \Theta\Pi P\Pi\Theta$, where $\Theta = \mathrm{diag}(\theta_1, \ldots, \theta_n)$. For the null, we take $n = 2000$ and draw θ_i from $\mathrm{Pareto}(12, 3/8)$ and scale θ to have an ℓ^2-norm of 8. For the alternative, we let $(n, K) = (2000, 2)$, P be the matrix with 1 on the diagonal and 0.6 on the off-diagonal, rows of Π equal to $\{1, 0\}$ and $\{0, 1\}$ half by half, and with the same θ as in the null but (to make it harder to separate from the null) rescaled to have an ℓ^2-norm of 9. The results confirm the limiting null of $N(0,1)$ for both tests.

2.3. *Power analysis of the SgnT and SgnQ tests.* The matrices Ω and $\widetilde{\Omega}$ play a key role in power analysis. Recall that Ω is defined in (1.3) where $\mathrm{rank}(\Omega) = K$, and $\widetilde{\Omega} = \Omega - \eta^*(\eta^*)'$ is defined in (1.14) with $\eta^* = \Omega\mathbf{1}_n/\sqrt{\mathbf{1}_n'\Omega\mathbf{1}_n}$ as in (1.12). Recall that $\lambda_1, \lambda_2, \ldots, \lambda_K$ are the K nonzero eigenvalues of Ω. Let $\xi_1, \xi_2, \ldots, \xi_K$ be the corresponding eigenvectors. The following theorems are proved in the supplemental material.

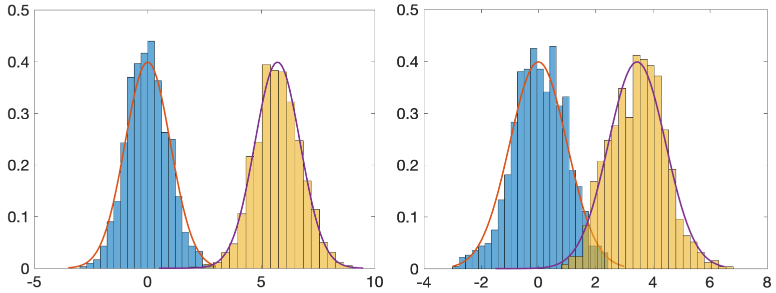

FIG 2. *Left: histograms of the SgnT test statistics in (2.17) for the null (blue) and the alternative (yellow). Empirical mean and SD under the null:* 0.04 *and* 0.94. *Right: same but for SgnQ test statistic in (2.19). Empirical mean and SD under the null:* -0.02 *and* 0.92. *Repetition:* 1000 *times. See setting details in the main text.*

THEOREM 2.3 (Limiting behavior the SgnT statistic (alternative)). *Consider the testing problem (1.6) under the DCMM model (1.1)-(1.4). Suppose the alternative hypothesis is true, and the conditions (2.15)-(2.16) hold. As* $n \to \infty$, $\mathbb{E}[T_n] = \mathrm{tr}(\widetilde{\Omega}^3) + o((|\lambda_2|/\lambda_1)^3\|\theta\|^6) + o(\|\theta\|^3)$ *and* $\mathrm{Var}(T_n) \le C\big(\|\theta\|^6 + (\lambda_2/\lambda_1)^4\|\theta\|^4\|\theta\|_3^6\big)$.

THEOREM 2.4 (Limiting behavior of the SgnQ statistic (alternative)). *Consider the testing problem (1.6) under the DCMM model (1.1)-(1.4). Suppose the alternative hypothesis is true and the conditions (2.15)-(2.16) hold. As* $n \to \infty$, $\mathbb{E}[Q_n] = \mathrm{tr}(\widetilde{\Omega}^4) + o((\lambda_2/\lambda_1)^4\|\theta\|^8) + o(\|\theta\|^4)$ *and* $\mathrm{Var}(Q_n) \le C\big(\|\theta\|^8 + C(\lambda_2/\lambda_1)^6\|\theta\|^8\|\theta\|_3^6\big)$.

We conjecture that both T_n and Q_n are asymptotically normal under the alternative. In fact, asymptotic normality is easy to establish for the Ideal SgnT and Ideal SgnQ. To establish results for the real SgnT and real SgnQ, we need very precise characterization of the plug-in effect. For reasons of space, we leave them to the future.

Consider the SgnT test (2.18) first. By Theorem 2.3 and Lemma 2.1, under the alternative,

$$\text{the mean and variance of } \frac{T_n}{\sqrt{6(\|\hat{\eta}\|^2-1)^3}} \text{ are } \frac{\mathrm{tr}(\widetilde{\Omega}^3)}{\sqrt{6\|\eta^*\|^6}} \text{ and } \sigma_n^2, \text{ respectively,} \tag{2.21}$$

where σ_n^2 denotes the asymptotic variance, which satisfies that

$$\sigma_n^2 \le \begin{cases} C, & \text{if } |\lambda_2/\lambda_1| \ll \sqrt{\|\theta\|/\|\theta\|_3^3}, \\ C(\lambda_2/\lambda_1)^4 \cdot (\|\theta\|_3^6/\|\theta\|^2), & \text{if } |\lambda_2/\lambda_1| \gg \sqrt{\|\theta\|/\|\theta\|_3^3}. \end{cases} \tag{2.22}$$

If we fix the degree heterogeneity vector θ and let (λ_2/λ_1) range, there is a *phase change* in the variance. We shall call:

- the case of $|\lambda_2/\lambda_1| \le C\sqrt{\|\theta\|/\|\theta\|_3^3}$ as the *weak signal* case for SgnT.
- the case of $|\lambda_2/\lambda_1| \gg \sqrt{\|\theta\|/\|\theta\|_3^3}$ as the *strong signal* case for SgnT.

It remains to derive a more explicit formula for $\mathrm{tr}(\widetilde{\Omega}^3)$. Recall that λ_k and ξ_k are the k-th eigenvalue and eigenvector of Ω, $1 \le k \le K$, respectively. Define $\Lambda \in \mathbb{R}^{(K-1)\times(K-1)}$ and $h \in \mathbb{R}^{K-1}$ by $\Lambda = \mathrm{diag}(\lambda_2, \lambda_3, \ldots, \lambda_K)$ and $h_k = (\mathbf{1}_n'\xi_{k+1})/(\mathbf{1}_n'\xi_1)$, $1 \le k \le K-1$. It can be shown that $\mathbf{1}_n'\xi_1 \neq 0$ and $\|h\|_\infty \le C$ so the vector h is well-defined. In the special case of $\|h\|_\infty = o(1)$ (this happens when the angle between $\mathbf{1}_n$ and ξ_1 is small):

- We can show that $\mathrm{tr}(\widetilde{\Omega}^3) \approx \sum_{k=2}^K \lambda_k^3$.
- Motivated by these, we say "signal cancellation" happens when $|\mathrm{tr}(\widetilde{\Omega}^3)| \ll \sum_{k=2}^K |\lambda_k|^3$.

Therefore, "signal cancellation" may happen if the $(K-1)$ eigenvalues $\lambda_2, \lambda_3, \ldots, \lambda_K$ have different signs. In fact, in the extreme case, we can have $\sum_{k=2}^K \lambda_k^3 = 0$, though $\sum_{k=2}^K |\lambda_k|^3$ is very large (e.g., [25, Section 3.3]). Normally, the "signal cancellation" is found for odd-order moment-based statistics (e.g., $3rd$, $5th$, …, moment), but not for even-order moment methods (in fact, the SgnQ test won't experience such "signal cancellation").

Fortunately, "signal cancellation" is only possible when $\lambda_2, \lambda_3, \ldots, \lambda_K$ have different signs, and can be avoided in some special cases. We propose the following conditions.

CONDITION 2.1. *(a) $\lambda_2, \lambda_3, \ldots, \lambda_K$ have the same signs, (b) $K = 2$, and (c) $|\lambda_2|/\lambda_1 \to 0$, and $|\mathrm{tr}(\Lambda^3) + 3h'\Lambda^3 h + 3(h'\Lambda h)(h'\Lambda^2 h) + (h'\Lambda h)^3| \geq C\sum_{k=2}^K |\lambda_k|^3$.*

In (a)-(b), $\lambda_2, \ldots, \lambda_K$ have the same signs. Condition (c) is based on more delicate analysis; see the proof of Lemma 2.2 for details.

While the above discussion is motivated by the case of $\|h\|_\infty = o(1)$, the idea continues to be valid for more general cases. The following is proved in the supplementary material.

LEMMA 2.2 (Analysis of $\mathrm{tr}(\widetilde{\Omega}^3)$). *Suppose conditions of Theorem 2.3 hold. Under the alternative hypothesis,*

- *If $|\lambda_2|/\lambda_1 \to 0$, then $\mathrm{tr}(\widetilde{\Omega}^3) = \mathrm{tr}(\Lambda^3) + 3h'\Lambda^3 h + 3(h'\Lambda h)(h'\Lambda^2 h) + (h'\Lambda h)^3 + o(|\lambda_2|^3)$.*
- *If $\lambda_2, \lambda_3, \ldots, \lambda_K$ have the same signs, then*

$$|\mathrm{tr}(\widetilde{\Omega}^3)| \geq \begin{cases} \sum_{k=2}^K |\lambda_k|^3 + o(|\lambda_2|^3), & \textit{if } |\lambda_2/\lambda_1| \to 0, \\ C|\lambda_2|^3, & \textit{if } |\lambda_2/\lambda_1| \geq C. \end{cases}$$

- *In the special case where $K = 2$, the vector h is a scalar, and*

$$|\mathrm{tr}(\widetilde{\Omega}^3)| \begin{cases} = [(h^2+1)^3 + o(1)] \cdot |\lambda_2|^3, & \textit{if } |\lambda_2|/\lambda_1 \to 0, \\ \geq C|\lambda_2|^3, & \textit{if } |\lambda_2/\lambda_1| \geq C. \end{cases}$$

As a result, when either one of (a)-(c) holds, $|\mathrm{tr}(\widetilde{\Omega}^3)| \geq C\sum_{k=2}^K |\lambda_k|^3$.

It can be shown $\|\eta^*\| \asymp \sqrt{\lambda_1} \asymp \|\theta\|$. We combine Lemma 2.2 with (2.21)-(2.22). In the *weak signal* case, $\frac{\mathbb{E}[T_n]}{\sqrt{\mathrm{Var}(T_n)}} \geq \frac{C(\sum_{k=2}^K |\lambda_k|^3)}{\|\theta\|^3} \geq C\big(\lambda_1^{-\frac{3}{2}} \sum_{k=2}^K |\lambda_k|^3\big)$. In the *strong signal* case, since $(\lambda_2/\lambda_1)^2 \leq \lambda_1^{-2}(\sum_{k=2}^K |\lambda_k|^3)^{\frac{2}{3}}$, we have $\frac{\mathbb{E}[T_n]}{\sqrt{\mathrm{Var}(T_n)}} \geq \frac{C(\sum_{k=2}^K |\lambda_k|^3)}{\lambda_1^{-2}(\sum_{k=2}^K |\lambda_k|^3)^{\frac{2}{3}} \|\theta\|_3^3 \|\theta\|^2} \geq \frac{C\|\theta\|^3}{\|\theta\|_3^3}\big(\lambda_1^{-\frac{3}{2}} \sum_{k=2}^K |\lambda_k|^3\big)^{\frac{1}{3}}$, where it should be noted that in our setting, $\|\theta\|^3/\|\theta\|_3^3 \to \infty$. As a result, in both cases, the power of the SgnT test → 1 as long as $\lambda_1^{-3/2} \sum_{k=2}^K |\lambda_k|^3 \to \infty$. This is validated in the following theorem, which is proved in the supplemental material.

THEOREM 2.5 (Power of the SgnT test). *Under the conditions of Theorem 2.3, for any fixed $\alpha \in (0,1)$, consider the SgnT test in (2.18). Suppose one of the cases in Condition 2.1 holds. As $n \to \infty$, if $\lambda_1^{-1/2}\big(\sum_{k=2}^K |\lambda_k|^3\big)^{1/3} \to \infty$, then the Type I error $\to \alpha$, and the Type II error $\to 0$.*

Next, consider the SgnQ test (2.20). By Theorem 2.4 and Lemma 2.1, under the alternative, the mean and variance of $[Q_n - 2(\|\hat{\eta}\|^2 - 1)^2]/\sqrt{8(\|\hat{\eta}\|^2 - 1)^4}$ are $\mathrm{tr}(\widetilde{\Omega}^4)/\sqrt{8\|\eta^*\|^8}$ and σ_n^2, respectively, where σ_n^2 denotes the asymptotic variance and satisfies

$$\sigma_n^2 \leq \begin{cases} C, & \text{if } |\lambda_2/\lambda_1| \ll \|\theta\|_3^{-1}, \\ C(\lambda_2/\lambda_1)^6 \cdot \|\theta\|_3^6, & \text{if } |\lambda_2/\lambda_1| \gg \|\theta\|_3^{-1}. \end{cases}$$

Similar to the SgnT test, if we fix the degree heterogeneity vector θ and let (λ_2/λ_1) range, there is a *phase change* in the variance. We shall call:

- the case of $|\lambda_2/\lambda_1| \leq C\|\theta\|_3^{-1}$ as the *weak signal* case for SgnQ.
- the case of $|\lambda_2/\lambda_1| \gg \|\theta\|_3^{-1}$ as the *strong signal* case for SgnQ.

We now analyze $\mathrm{tr}(\widetilde{\Omega}^4)$. The following lemma is proved in the supplementary material.

LEMMA 2.3 (Analysis of $\mathrm{tr}(\widetilde{\Omega}^4)$). *Suppose the conditions of Theorem 2.4 hold. Under the alternative hypothesis,*

- *If* $|\lambda_2|/\lambda_1 \to 0$, *then* $\mathrm{tr}(\widetilde{\Omega}^4) = \mathrm{tr}(\Lambda^4) + (q'\Lambda q)^4 + 2(h'\Lambda^2 h)^2 + 4(h'\Lambda h)^2(h'\Lambda^2 h) + 4h'\Lambda^4 h + 4(h'\Lambda h)(h'\Lambda^3 h) + o(\lambda_2^4) \gtrsim \sum_{k=2}^{4} \lambda_k^4$.
- *If* $|\lambda_2|/\lambda_1 \geq C$, *then* $\mathrm{tr}(\widetilde{\Omega}^4) \geq C\sum_{k=2}^{K} \lambda_k^4$.
- *In the special case of* $K = 2$, h *is a scalar and* $\mathrm{tr}(\widetilde{\Omega}^4) = [(h^2+1)^4 + o(1)] \cdot \lambda_2^4$.

As a result, the SgnQ test has no issue of "signal cancellation", and it always holds that $\mathrm{tr}(\widetilde{\Omega}^4) \geq C\sum_{k=2}^{K} \lambda_k^4$. Then, in the *weak signal* case, we have $\frac{\mathbb{E}[Q_n]}{\sqrt{\mathrm{Var}(Q_n)}} \geq \frac{C(\sum_{k=2}^{K} \lambda_k^4)}{\|\theta\|^4} \geq C\big(\lambda_1^{-2}\sum_{k=2}^{K} \lambda_k^4\big)$. In the *strong signal* case, since $(\lambda_2/\lambda_1)^3 \leq \lambda_1^{-3}(\sum_{k=2}^{K} \lambda_k^4)^{\frac{3}{4}}$, we have $\frac{\mathbb{E}[Q_n]}{\sqrt{\mathrm{Var}(Q_n)}} \geq \frac{C(\sum_{k=2}^{K} \lambda_k^4)}{\lambda_1^{-3}(\sum_{k=2}^{K} \lambda_k^4)^{\frac{3}{4}}\|\theta\|_3^3\|\theta\|^4} \geq \frac{C\|\theta\|^3}{\|\theta\|_3^3}\big(\lambda_1^{-2}\sum_{k=2}^{K} \lambda_k^4\big)^{\frac{1}{4}}$, where $\|\theta\|^3/\|\theta\|_3^3 \to \infty$. So, in both cases, the power of the SgnQ test goes to 1 if $\lambda_1^{-2}\sum_{k=2}^{K} \lambda_k^4 \to \infty$. This is validated in Theorem 2.6, which is proved in the supplemental material.

THEOREM 2.6 (Power of the SgnQ test). *Under the conditions of Theorem 2.4, for any fixed* $\alpha \in (0,1)$, *consider the SgnQ test in* (2.20). *As* $n \to \infty$, *if* $\lambda_1^{-1/2}\big(\sum_{k=2}^{K} \lambda_k^4\big)^{1/4} \to \infty$, *then the Type I error* $\to \alpha$, *and the Type II error* $\to 0$.

In summary, Theorem 2.5 and Theorem 2.6 imply that as long as

$$|\lambda_2|/\sqrt{\lambda_1} \to \infty, \tag{2.23}$$

the levels of SgnT and SgnQ tests tend to α as expected, and their powers tend to 1. The SgnT test requires mild conditions to avoid "signal cancellation", but the SgnQ test has no such issue (such an advantage of SgnQ test is confirmed by numerical study in Section 5).

Remark 7. Practically, we prefer to fix α, say, $\alpha = 5\%$. If we allow the level α to change with n, then when (2.23) holds, there is a sequence of α_n that tends to 0 slowly enough such that $|\lambda_2|/(z_{\alpha_n/2} \cdot \sqrt{\lambda_1}) \to \infty$. As a result, for either of the two tests, the Type I error $\to 0$ and the power $\to 1$, so the sum of Type I and Type II errors $\to 0$.

Example 1 (contd). For this example, $\lambda_1 \sim (1 + (K-1)b_n)\|\theta\|^2$, and $\lambda_k \sim (1 - b_n)\|\theta\|^2$, $k = 2, 3, \ldots, K$. The condition (2.23) of $|\lambda_2|/\sqrt{\lambda_1} \to \infty$ translates to $(1 - b_n)\|\theta\| \to \infty$. See Section 1.2 and also Section 3 for more discussion.

3. Optimal adaptivity, lower bound, and Region of Impossibility.

We now focus on the Region of Impossibility, where $|\lambda_2|/\sqrt{\lambda_1} \to 0$. We first present a standard minimax lower bound, from which we can conclude that there is a sequence of hypothesis pairs (one alternative and one null) that are asymptotically indistinguishable. However, this does not answer the question whether *all alternatives* in the Region of Impossibility are indistinguishable from the null. To answer this question, we need much more sophisticated study; see Section 3.2.

3.1. *Minimax lower bound.* Given an integer $K \geq 1$, a constant $c_0 > 0$, and two positive sequences $\{\alpha_n\}_{n=1}^{\infty}$ and $\{\beta_n\}_{n=1}^{\infty}$, we define a class of parameters for DCMM (recall that $\Omega = \Theta\Pi P\Pi'\Theta$, $G = \|\theta\|^{-2}\Pi'\Theta^2\Pi$ and is properly scaled, and λ_k is the k-th largest eigenvalue of Ω in magnitude):

$$\mathcal{M}_n(K, c_0, \alpha_n, \beta_n) = \left\{ \begin{array}{r} (\theta, \Pi, P) : \theta_{\max} \leq \beta_n, \|\theta\|^{-1} \leq \beta_n, \|\theta\|^2\|\theta\|_1^{-1}\sqrt{\log(\|\theta\|_1)} \leq \beta_n, \\ \frac{\max_k\{\sum_{i=1}^n \theta_i\pi_i(k)\}}{\min_k\{\sum_{i=1}^n \theta_i\pi_i(k)\}} \leq c_0, \|G^{-1}\| \leq c_0, |\lambda_2|/\sqrt{\lambda_1} \geq \alpha_n \end{array} \right\}.$$

For the null case, $K = P = \pi_i = 1$, and the above defines a class of θ, which we write for short by $\mathcal{M}_n(1, c_0, \alpha_n, \beta_n) = \mathcal{M}_n^*(\beta_n)$.

THEOREM 3.1 (Minimax lower bound). *Fix $K \geq 2$, a constant $c_0 > 0$, and any sequences $\{\alpha_n\}_{n=1}^{\infty}$ and $\{\beta_n\}_{n=1}^{\infty}$ such that $\alpha_n \to 0$ and $\beta_n \to 0$ as $n \to \infty$. Then, as $n \to \infty$,*

$$\inf_{\psi}\Big\{ \sup_{\theta \in \mathcal{M}_n^*(\beta_n)} \mathbb{P}(\psi = 1) + \sup_{(\theta,\Pi,P) \in \mathcal{M}_n(K, c_0, \alpha_n, \beta_n)} \mathbb{P}(\psi = 0) \Big\} \to 1,$$

where the infimum is taken over all possible tests ψ.

Theorem 3.1 says that in the Region of Impossibility, *there exists a sequence of alternatives* that are inseparable from the null. This does not show what we desire, that is *any sequence in the Region of Impossibility* is inseparable from the null. This is discussed in the next section.

3.2. *Region of Impossibility.* Recall that under DCMM, $\Omega = \Theta\Pi P\Pi'\Theta$ and $\Pi = [\pi_1, \pi_2, \ldots, \pi_n]'$. Since our model is a mixed-membership latent variable model, in order to characterize the *least favorable configuration*, it is conventional to use a *random mixed-membership (RMM) model* for the matrix Π, while (Θ, P) are still non-stochastic. In detail,

- Let $V = \{x \in \mathbb{R}^K, x_k \geq 0, \sum_{k=1}^K x_k = 1\}$.
- Let $V_0 = \{e_1, e_2, \ldots, e_K\}$, where e_k is the k-th Euclidean basis vector.

In DCMM-RMM, we fix a distribution F defined over V and assume $\pi_i \overset{iid}{\sim} F$ where $h \equiv \mathbb{E}[\pi_i]$. If we further restrict that F is defined over V_0, then the network has no mixed-membership, and DCMM-RMM reduces to DCBM-RMM.

The desired result is to show that, for any given P and F, there is a sequence of hypothesis pairs (a null and an alternative)

$$H_0^{(n)}: \quad \Omega = \theta\theta', \qquad \text{and} \qquad H_1^{(n)}: \quad \Omega = \widetilde{\Theta}\Pi P\Pi'\widetilde{\Theta}, \tag{3.24}$$

where $\widetilde{\Theta} = \mathrm{diag}(\widetilde{\theta}_1, \widetilde{\theta}_1, \ldots, \widetilde{\theta}_n)$ and $\widetilde{\theta}_i$ can be different from θ_i, such that the two hypotheses within each pair are asymptotically indistinguishable from each other, provided that under the alternative $|\lambda_2|/\sqrt{\lambda_1} \to 0$.

Here, since Ω depends on π_i, λ_k is random, and it is more convenient to translate the condition of $|\lambda_2|/\sqrt{\lambda_1} \to 0$ to the condition of

$$\|\theta\| \cdot |\mu_2(P)| \to 0, \tag{3.25}$$

where $\mu_k(P)$ is the k-th largest eigenvalue of P in magnitude. The equivalence of two conditions are justified in Appendix F.1 of the supplement. Condition (2.16) can also be ensured with high probability, by assuming that all entries of $\mathbb{E}[\pi_i]$ are at the order of $O(1)$.

Under the DCBM, the desired result can be proved satisfactorily. The key is the following lemma, which is in the line of Sinkhorn's beautiful work on scalable matrices [40] (see also [9, 27, 34]) and is proved in the supplement.

LEMMA 3.1. *Fix a matrix $A \in \mathbb{R}^{K,K}$ with strictly positive diagonal entries and non-negative off-diagonal entries, and a strictly positive vector $h \in \mathbb{R}^K$, there exists a diagonal matrix $D = \mathrm{diag}(d_1, d_2, \ldots, d_K)$ such that $DADh = 1_K$ and $d_k > 0$, $1 \le k \le K$.*

In detail, consider a DCBM-RMM setting where $\pi_i \overset{iid}{\sim} F$ and F is supported over V_0 (with possibly unequal probabilities on the K points). Recall $h = \mathbb{E}[\pi_i]$. By Lemma 3.1, there is a unique diagonal matrix D such that $DPDh = 1_K$. Let

$$\widetilde{\theta}_i = d_k \cdot \theta_i, \qquad \text{if } \pi_i = e_k, \qquad 1 \le i \le n,\ 1 \le k \le K. \tag{3.26}$$

The following theorem is proved in the supplementary material.

THEOREM 3.2 (Region of Impossibility (DCBM)). *Fix $K > 1$ and a distribution F defined over V_0. Consider a sequence of DCBM model pairs indexed by n:*

$$H_0^{(n)} : \Omega = \theta\theta' \qquad \textit{and} \qquad H_1^{(n)} : \Omega = \widetilde{\Theta}\Pi P \Pi' \widetilde{\Theta},$$

where $\pi_i \overset{iid}{\sim} F$ and $\widetilde{\Theta} = \mathrm{diag}(\widetilde{\theta}_1, \widetilde{\theta}_2, \ldots, \widetilde{\theta}_n)$ with $\widetilde{\theta}_i$ defined as in (3.26). If $\theta_{max} \le c_0$ for a constant $c_0 < 1$, $\min_{1 \le k \le K}\{h_k\} \ge C$, and $\|\theta\| \cdot |\mu_2(P)| \to 0$, then for each pair of two hypotheses, the χ^2-distance between the two joint distributions tends to 0, as $n \to \infty$.

To generalize this to RMM-DCMM, we fix a distribution F defined over V. Given a set of (Θ, P, Π) with $\Theta = \mathrm{diag}(\theta_1, \theta_2, \ldots, \theta_n)$ and $\pi_i \overset{iid}{\sim} F$, let $\widetilde{h}_D = \mathbb{E}[D^{-1}\pi_i / \|D^{-1}\pi_i\|_1]$ for any diagonal matrix $D \in \mathbb{R}^{K \times K}$ with positive diagonals. We assume that there is a D such that

$$DPD\widetilde{h}_D = 1_K, \qquad \min_{1 \le k \le K}\{\widetilde{h}_{D,k}\} \ge C. \tag{3.27}$$

When such a D exists, we let

$$\widetilde{\theta}_i = \theta_i / \|D^{-1}\pi_i\|_1, \qquad 1 \le i \le n. \tag{3.28}$$

When the support of F is restricted to V_0, all realizations of π_i are degenerate (i.e., one entry is 1, and other entries are 0), so $\widetilde{h}_D = h$, $\widetilde{\theta}_i$ is the same as that in (3.26), and (3.27) holds by Lemma 3.1. Under DCMM-RMM, π_i's are not degenerate. We conjecture that (3.27) continues to hold generally (we can show it for the cases of $K = 2, 3$; the proof is elementary so is omitted). The following theorem is proved in the supplementary material.

THEOREM 3.3 (Region of Impossibility (DCMM)). *Fix $K > 1$ and a distribution F defined over V. Consider a sequence of DCMM model pairs indexed by n:*

$$H_0^{(n)} : \Omega = \theta\theta' \qquad \textit{and} \qquad H_1^{(n)} : \Omega = \widetilde{\Theta}\Pi P \Pi' \widetilde{\Theta},$$

where $\pi_i \overset{iid}{\sim} F$ and $\widetilde{\Theta} = \mathrm{diag}(\widetilde{\theta}_1, \widetilde{\theta}_2, \ldots, \widetilde{\theta}_n)$ with $\widetilde{\theta}_i$ defined as in (3.28). If (3.27) holds, $\theta_{max} \le c_0$ for a constant $c_0 < 1$, and $\|\theta\| \cdot |\mu_2(P)| \to 0$, then for each pair of two hypotheses, the χ^2-distance between the two joint distributions tends to 0, as $n \to \infty$.

One of the main strengths of Theorems 3.2-3.3 is that this lower bound is valid for an arbitrary choice of $\theta \in \mathbb{R}^n_+$. This is stronger than the standard minimax lower bound.

In Theorem 3.3, we try to be as general as we can so Π is given (and we are not allowed to change it in our construction). For any P and F, by Lemma 3.1, there is a unique positive diagonal matrix D such that $DPDh = 1_K$ where $h = \mathbb{E}[\pi_i]$. We now consider a special case where we allow Π to depend on D in our construction. In this case, Condition (3.27) can be removed. Let $\widetilde{\Pi} = [\widetilde{\pi}_1, \widetilde{\pi}_2, \ldots, \widetilde{\pi}_n]'$ and $\widetilde{\Theta} = \mathrm{diag}(\widetilde{\theta}_1, \widetilde{\theta}_2, \ldots, \widetilde{\theta}_n)$, with

$$\widetilde{\pi}_i = D\pi_i / \|D\pi_i\|_1, \qquad \widetilde{\theta}_i = \|D\pi_i\|_1 \cdot \theta_i. \tag{3.29}$$

THEOREM 3.4 (Region of Impossibility (DCMM with flexible Π)). *Fix $K > 1$ and a distribution F defined over $\mathcal{V}$. Consider a sequence of DCMM model pairs indexed by n: $H_0^{(n)} : \Omega = \theta\theta'$ and $H_1^{(n)} : \Omega = \widetilde{\Theta}\widetilde{\Pi}P\widetilde{\Pi}'\widetilde{\Theta}$, where $\widetilde{\Pi}$ and $\widetilde{\Theta}$ are defined as in (3.29). If $\theta_{max} \leq c_0$ for a constant $c_0 < 1$, $\min_{1\leq k\leq K}\{h_k\} \geq C$, and $\|\theta\| \cdot |\mu_2(P)| \to 0$, then for each pair of two hypotheses, the χ^2-distance between the two joint distributions tends to 0, as $n \to \infty$.*

Finally, we consider the case where we require that the null and the alternative have perfectly matching Θ matrix (up to an overall scaling). This is especially of interest when we consider SBM or MMSBM models where we have little freedom in choosing the Θ matrix. In this case, in order that the two hypotheses are indistinguishable, the expected node degrees under the alternative have to match those under the null. For each $1 \leq i \leq n$, conditional on π_i and neglecting the effect of no self edges, the expected degree of node i equals to $\|\theta\|_1 \cdot \theta_i$ and $\|\theta\|_1 \cdot (\pi_i' Ph) \cdot \theta_i$ under the null and under the alternative, respectively, where $\{\pi_j\}_{j\neq i} \overset{iid}{\sim} F$ and $h = \mathbb{E}[\pi_j]$. For the expected degrees to match under any realized π_i, it is necessary that

$$Ph = q_n 1_K, \qquad \text{for some scaling parameter } q_n > 0. \tag{3.30}$$

THEOREM 3.5 (Region of Impossibility (DCMM with matching Θ)). *Fix $K > 1$ and a distribution F defined over $\mathcal{V}$. Consider a sequence of DCMM model pairs indexed by n: $H_0^{(n)} : \Omega = q_n \cdot \theta\theta'$ and $H_1^{(n)} : \Omega = \Theta\Pi P\Pi'\Theta$, where $\Theta = \mathrm{diag}(\theta_1, \theta_2, \ldots, \theta_n)$, $\pi_i \overset{iid}{\sim} F$, and (P, h, q_n) satisfy (3.30). If $\theta_{max} \leq c_0$ for a constant $c_0 < 1$, $\min_{1\leq k\leq K}\{h_k\} \geq C$, and $\|\theta\| \cdot |\mu_2(P)| \to 0$, then for each pair of two hypotheses, the χ^2-distance between the two joint distributions tends to 0, as $n \to \infty$.*

Theorems 3.4-3.5 are proved in the supplementary material.

Example 1 (contd). In Example 1, π_i is drawn from $e_1, e_2, \ldots, e_K$ with equal probabilities, and $P = (1 - b_n)I_K + b_n 1_K 1_K'$. Therefore, $h = \mathbb{E}[\pi_i] = (1/K)1_K$. In this case, all conditions of Theorem 3.5 hold. Note $q_n = (1/K) + (K-1)b_n/K$ and $\mu_2(P) = (1 - b_n)$.

Remark 8 (Least favorable configuration of LDA-DCMM). The Dirichlet model is often used for mixed-memberships [1]. Consider the model pairs $H_0^{(n)} : \Omega = q_n\theta\theta'$ and $H_1^{(n)}$: $\Omega = \Theta\Pi P\Pi'\Theta$ and where $\pi_i \overset{iid}{\sim} \mathrm{Dir}(\alpha)$ ($\mathrm{Dir}(\alpha)$: Dirichlet distribution with parameters $\alpha = (\alpha_1, \ldots, \alpha_K)'$). By Theorem 3.5, as long as $P\alpha \propto 1_K$, the null and alternative hypotheses are asymptotically indistinguishable if $(1 - q_n)\|\theta\| \to 0$. One can easily construct P such that $P\alpha \propto 1_K$. For example, $P = (1 - q_n)MM' + q_n 1_K 1_K'$, where $M \in \mathbb{R}^{K\times(K-1)}$ is a matrix whose columns are from $Span^{\perp}(\alpha)$ and satisfy $\mathrm{diag}(MM') = I_K$.

3.3. *Optimal adaptivity.* Recall that $\sqrt{\lambda_1}$, $|\lambda_2|/\lambda_1$, and $|\lambda_2|/\sqrt{\lambda_1}$ can be viewed as a measure for the sparsity, community dissimilarity, and SNR, respectively. Combining Theorems 2.1-2.4, Theorems 3.2-3.5, and Remark 7 in Section 2.3, in the two-dimensional phase space where the x-axis is $\sqrt{\lambda_1}$ and the y-axis is the $|\lambda_2|/\lambda_1$, we have a partition to two regions, the Region of Possibility and the Region of Impossibility.

- **Region of Impossibility** ($1 \ll \sqrt{\lambda_1} \ll \sqrt{n}$, $|\lambda_2|/\sqrt{\lambda_1} = o(1)$). In this region, any DCBM alternative is asymptotically inseparable from the null, and up to a mild condition, any DCMM alternative is also asymptotically inseparable from the null.
- **Region of Possibility** ($1 \ll \sqrt{\lambda_1} \ll \sqrt{n}$, $|\lambda_2|/\sqrt{\lambda_1} \to \infty$). In this region, asymptotically, any alternative is completely separable form any null.

The SgnQ test is optimally adaptive: for any alternative in the Region of Possibility, the test is able to separate it from the null with a sum of Type I and Type II errors tending to 0. The

SgnT test is also optimally adaptive, provided that some mild conditions hold to avoid signal cancellation. To the best of our knowledge, the Signed Polygon is the only known test that is both applicable to general DCMM (where we allow severe degree heterogeneity and arbitrary mixed-memberships) and optimally adaptive. The EZ and GC tests are the only other tests we know that are applicable to general DCMM, but their variances are unsatisfactorily large for the less sparse case, so they are not optimally adaptive. See [29] for details.

Remark 9. Most existing lower bound results [36, 2, 16] are within the standard minimax framework, where they focus on a particular sequence of alternative (e.g., the off-diagonals of P are equal). In our case, the standard minimax theorem only implies that in the Region of Impossibility, there is a sequence of alternative that are inseparable from the null. Our results (Theorems 3.2-3.5) shed new light on the Region of Impossibility, saying that for each alternative, we can pair it with a null such that two hypotheses are asymptotically inseparable.

Remark 10. Existing minimax lower bounds [36, 4, 2] are largely focused on the SBM. Though a least favorable scenario for SBM is least favorable for DCMM, the former does not provide much insight on how the least favorable configurations and the phase transition depend on the degree heterogeneity and mixed-memberships. Moreover, our results (see also [19]) suggest that $\|\theta\|$, not $\|\theta\|_1$, determines the separating boundary. In the SBM case, $\theta_1 = \ldots = \theta_n$ and $\|\theta\|_1 = \sqrt{n}\|\theta\|$, so it is hard to tell which of the two norms decides the boundary. In DCMM, there is no simple relationship between $\|\theta\|_1$ and $\|\theta\|$, and we can tell this clearly.

Remark 11. A sharper version of the phase transition is that there exists a constant $c_0 > 0$ such that the Region of Possibility and Region of Impossibility are given by $|\lambda_2|/\sqrt{\lambda_1} > c_0$ and $|\lambda_2|/\sqrt{\lambda_1} < c_0$, respectively. In some special cases, this kind of results exist for community detection (a related but different problem). For example, [19] considered a setting where (i) there is no mixed-membership, (ii) for some constants $a, b > 0$, $P(k,\ell) = a$ if $k = \ell$ and b otherwise, (iii) the communities have equal size, and (iv) for a constant $\phi > 0$, $\{\sqrt{n}\theta_i\}_{i=1}^n$ are iid drawn from a fixed distribution supported in $[\phi, \infty)$. They showed that, when $(a-b)^2\mathbb{E}\|\theta\|^2 < K(a+b)$, it is impossible to reconstruct the community label matrix Π. Moreover, in the special case of $K = 2$, [18] (also, see [12]) showed that when $(a-b)^2\mathbb{E}\|\theta\|^2 > 2(a+b)$, it is possible to construct an estimate of Π that is positively correlated with the true community labels. By connecting $(a, b, \mathbb{E}\|\theta\|^2)$ with eigenvalues, it is seen that these results give a sharp phase transition at $c_0 = 1$, in the special case where $K = 2$ and (i)-(iv) hold. For more general settings, whether such a sharp phase transition exists is unclear: a slight change in conditions (i)-(iv) may affect the lower bounds, and the optimal tests (for the sharp phase transition) are hard to find as they usually need to adapt to specific features of the model. Also, technically, allowing for mixed-memberships makes the lower bound much harder to study, and allowing for unequal community sizes and unequal off-diagonal entries in P requires an application of DAD theorem in lower bound construction (which is not needed in [19]). Moreover, [12, 18, 19] are for community detection and our paper is on global testing. For general DCMM settings, it is unclear whether the phase transitions for two problems are the same.

4. The behavior the SgnQ test statistics.

In this section, we study the SgnQ test statistic Q_n and explain how to prove Theorems 2.2, 2.4, and 2.6. We introduce a proxy SgnQ test statistic Q_n^* and an Ideal SgnQ test statistic $\widetilde{Q}_n$. Write $Q_n = \widetilde{Q}_n + (Q_n^* - \widetilde{Q}_n) + (Q_n - Q_n^*)$. We study the three terms on the RHS in Sections 4.1-4.3, respectively. Given these results, the proofs of Theorems 2.2, 2.4, and 2.6 are straightforward and contained in Appendix B. The study of the SgnT test statistic T_n is similar and contained in Appendix A, where we also prove Theorems 2.1, 2.3, and 2.5.

Recall that the SgnQ statistic Q_n is defined as

$$Q_n = \sum_{i_1,i_2,i_3,i_4 (dist)} (A_{i_1 i_2} - \hat{\eta}_{i_1}\hat{\eta}_{i_2})(A_{i_2 i_3} - \hat{\eta}_{i_2}\hat{\eta}_{i_3})(A_{i_3 i_4} - \hat{\eta}_{i_3}\hat{\eta}_{i_4})(A_{i_4 i_1} - \hat{\eta}_{i_4}\hat{\eta}_{i_1}),$$

where $\hat{\eta} = A\mathbf{1}_n/\sqrt{V}$, with $V = \mathbf{1}_n' A\mathbf{1}_n$. In Section 1.4, we have introduced the following non-stochastic proxy of $\hat{\eta}$: $\eta^* = \Omega\mathbf{1}_n/\sqrt{v_0}$, where $v_0 = \mathbf{1}_n'\Omega\mathbf{1}_n$. We now introduce another (stochastic) proxy $\tilde{\eta}$ by

$$\tilde{\eta} = A\mathbf{1}_n/\sqrt{v}, \qquad \text{where } v = \mathbb{E}[\mathbf{1}_n' A\mathbf{1}_n] = \mathbf{1}_n'(\Omega - \text{diag}(\Omega))\mathbf{1}_n. \tag{4.31}$$

Denoting the mean of $\tilde{\eta}$ by η, it is seen that

$$\eta = ([\Omega - \text{diag}(\Omega)]\mathbf{1}_n)/\sqrt{\mathbf{1}_n'(\Omega - \text{diag}(\Omega))\mathbf{1}_n}. \tag{4.32}$$

Here, η and η^* are close to each other but η^* has a more explicit form. For example, under the null hypothesis, $\Omega = \theta\theta'$, and it is seen that $\eta^* = \theta$. Recall that $A = \Omega - \text{diag}(\Omega) + W$ and $\widetilde{\Omega} = \Omega - \eta^*(\eta^*)'$. Fix $1 \leq i, j \leq n$ and $i \neq j$. First, we write

$$A_{ij} - \hat{\eta}_i\hat{\eta}_j = (A_{ij} - \eta_i^*\eta_j^*) + (\eta_i^*\eta_j^* - \hat{\eta}_i\hat{\eta}_j) = \widetilde{\Omega}_{ij} + W_{ij} + (\eta_i^*\eta_j^* - \hat{\eta}_i\hat{\eta}_j).$$

Second, we write $\eta_i^*\eta_j^* - \hat{\eta}_i\hat{\eta}_j = \delta_{ij} + r_{ij}$, where

$$\delta_{ij} = \eta_i(\eta_j - \tilde{\eta}_j) + \eta_j(\eta_i - \tilde{\eta}_i) \tag{4.33}$$

is the linear approximation term of $(\eta_i^*\eta_j^* - \hat{\eta}_i\hat{\eta}_j)$ and $r_{ij} \equiv (\eta_i^*\eta_j^* - \hat{\eta}_i\hat{\eta}_j) - \delta_{ij}$ is the remainder term. By definition and elementary algebra,

$$r_{ij} = (\eta_i^*\eta_j^* - \eta_i\eta_j) - (\eta_i - \tilde{\eta}_i)(\eta_j - \tilde{\eta}_j) + (1 - \frac{v}{V})\hat{\eta}_i\hat{\eta}_j. \tag{4.34}$$

It is shown that r_{ij} is of a smaller order than that of δ_{ij}. The remainder term can be shown to have a negligible effect over T_n and Q_n, in terms of the variances of T_n and Q_n, respectively; see Theorem 4.3.

Let X be the symmetric matrix where all diagonal entries are 0 and for $1 \leq i, j \leq n$ but $i \neq j$, $X_{ij} = A_{ij} - \hat{\eta}_i\hat{\eta}_j$, or equivalently,

$$X_{ij} = \widetilde{\Omega}_{ij} + W_{ij} + \delta_{ij} + r_{ij}. \tag{4.35}$$

If we omit the remainder term, then we have a proxy of X, denoted by X^*, where all diagonal entries of X^* are 0, and for $1 \leq i, j \leq n$ but $i \neq j$,

$$X_{ij}^* = \widetilde{\Omega}_{ij} + W_{ij} + \delta_{ij}. \tag{4.36}$$

If we further omit the δ term, then we have another proxy of X, denoted by $\widetilde{X}$, where all diagonal entries of $\widetilde{X}$ are 0, and for $1 \leq i, j \leq n$ but $i \neq j$,

$$\widetilde{X}_{ij} = \widetilde{\Omega}_{ij} + W_{ij}. \tag{4.37}$$

With the above notations, we can rewrite Q_n as $Q_n = \sum_{i_1,i_2,i_3,i_4 (dist)} X_{i_1i_2}X_{i_2i_3}X_{i_3i_4}X_{i_4i_1}$. We introduce the *Proxy SgnQ test statistic* and *Ideal SgnQ test statistic* by

$$Q_n^* = \sum_{i_1,i_2,i_3,i_4 (dist)} X_{i_1i_2}^*X_{i_2i_3}^*X_{i_3i_4}^*X_{i_4i_1}^*, \quad \widetilde{Q}_n \quad = \sum_{i_1,i_2,i_3,i_4 (dist)} \widetilde{X}_{i_1i_2}\widetilde{X}_{i_2i_3}\widetilde{X}_{i_3i_4}\widetilde{X}_{i_4i_1}.$$

The Ideal SgnQ test statistic $\widetilde{Q}_n$ is the same as that defined in (1.13). Using these notations, we partition Q_n as $Q_n = \widetilde{Q}_n + (Q_n^* - \widetilde{Q}_n) + (Q_n - Q_n^*)$. In Sections 4.1-4.3, we study the three terms on the RHS respectively.

4.1. *The behavior of the Ideal SgnQ test statistics.* In view of (4.37), the Ideal SgnQ test statistic $\widetilde{Q}_n$ is written as

$$\widetilde{Q}_n = \sum_{i_1,i_2,i_3,i_4(dist)} (\widetilde{\Omega}_{i_1i_2} + W_{i_1i_2})(\widetilde{\Omega}_{i_2i_3} + W_{i_2i_3})(\widetilde{\Omega}_{i_31i_4} + W_{i_3i_4})(\widetilde{\Omega}_{i_4i_1} + W_{i_4i_1}). \tag{4.38}$$

Under the null, $\Omega = \theta\theta'$ and $\eta^* = \theta$. By definition, $\widetilde{\Omega}_{ij} = 0$, and the statistic reduces to $\widetilde{Q}_n = \sum_{i_1,i_2,i_3,i_4(dist)} W_{i_1i_2}W_{i_2i_3}W_{i_3i_4}W_{i_4i_1}$. The RHS is the sum of a large number of uncorrelated terms, with each term being a 4-product of independent centered-Bernoulli variables. It can be shown that the statistic is asymptotically normal, with $\mathbb{E}[\widetilde{Q}_n] = 0$ and $\mathrm{Var}(\widetilde{Q}_n) \sim 8\|\theta\|^8$.

Consider the alternative hypothesis. In the RHS of (4.38), expanding the bracket and re-arranging, we have $2 \times 2 \times 2 \times 2 = 16$ post-expansion sums, each having the form of $\sum_{i_1,i_2,i_3,i_4(dist)} a_{i_1i_2}b_{i_2i_3}c_{i_3i_4}d_{i_4i_1}$, where a is a generic notation which may either equal to $\widetilde{\Omega}$ or W; same for b, c, and (d). For example, $\sum_{i_1,i_2,i_3,i_4(dist)} W_{i_1i_2}\widetilde{\Omega}_{i_2i_3}W_{i_3i_4}W_{i_4i_1}$ is one of the 16 post-expansion sums, corresponding to $b = \widetilde{\Omega}$, and $a = c = d = W$. Note that each of 16 post-expansion sums is the sum of many 4-product, where the number of the $\widetilde{\Omega}$ factors in each product is the same; denote this number (which can be 0, 1, 2, 3, or 4) by $N_{\widetilde{\Omega}}$. Similarly, the number of the W factors in each product are also the same. Denote it by N_W, we have $N_{\widetilde{\Omega}} + N_W = 4$. For the example above, $(N_{\widetilde{\Omega}}, N_W) = (1,3)$.

According to $(N_{\widetilde{\Omega}}, N_W)$, we can group the 16 post-expansion sums into 6 different types. Table 1 presents the mean and variance of each type (Recall that $\lambda_1, \ldots, \lambda_K$ are the K eigenvalues of Ω, arranged in descending order in magnitude. In Table 1, $\alpha = |\lambda_2|/\lambda_1$. In the alternative, we assume $|\lambda_2|/\sqrt{\lambda_1} \to \infty$, which translates to $\alpha\|\theta\| \to \infty$ since $\sqrt{\lambda_1} \asymp \|\theta\|$).

TABLE 1
The 6 *different types of the* 16 *post-expansion sums of* $\widetilde{Q}_n$ *(*$\|\theta\|_q$ *is the* ℓ^q*-norm of* θ *(the subscript is dropped when* $q = 2$*). In our setting,* $\alpha\|\theta\| \to \infty$*, and* $\|\theta\|_4^4 \ll \|\theta\|_3^3 \ll \|\theta\|^2 \ll \|\theta\|_1$.

Type	#	$(N_{\widetilde{\Omega}}, N_W)$	Examples	Mean	Variance
I	1	(0, 4)	$\sum_{i,j,k,\ell(dist)} W_{ij}W_{jk}W_{k\ell}W_{\ell i}$	0	$\asymp \|\theta\|^8$
II	4	(1, 3)	$\sum_{i,j,k,\ell(dist)} \widetilde{\Omega}_{ij}W_{jk}W_{k\ell}W_{\ell i}$	0	$\leq C\alpha^2\|\theta\|^4\|\theta\|_3^6 = o(\|\theta\|^8)$
IIIa	4	(2, 2)	$\sum_{i,j,k,\ell(dist)} \widetilde{\Omega}_{ij}\widetilde{\Omega}_{jk}W_{k\ell}W_{\ell i}$	0	$\leq C\alpha^4\|\theta\|^6\|\theta\|_3^6 = o(\alpha^6\|\theta\|^8\|\theta\|_3^6)$
IIIb	2	(2, 2)	$\sum_{i,j,k,\ell(dist)} \widetilde{\Omega}_{ij}W_{jk}\widetilde{\Omega}_{k\ell}W_{\ell i}$	0	$\leq C\alpha^4\|\theta\|_3^{12} = o(\|\theta\|^8)$
IV	4	(3, 1)	$\sum_{i,j,k,\ell(dist)} \widetilde{\Omega}_{ij}\widetilde{\Omega}_{jk}\widetilde{\Omega}_{k\ell}W_{\ell i}$	0	$\leq \alpha^6\|\theta\|^8\|\theta\|_3^6$
V	1	(4, 0)	$\sum_{i,j,k,\ell(dist)} \widetilde{\Omega}_{ij}\widetilde{\Omega}_{jk}\widetilde{\Omega}_{k\ell}\widetilde{\Omega}_{\ell i}$	$\sim \mathrm{tr}(\widetilde{\Omega}^4)$	0

From the table, among all 16 post-expansion sums, the total mean is $\sim \mathrm{tr}(\widetilde{\Omega}^4)$, and the total variance $\leq C\|\theta\|^8 + C(|\lambda_2|/\lambda_1)^6\|\theta\|^8\|\theta\|_3^6$, with Type I sum and Type IV sum being the major contributors. The following theorem is proved in the supplementary material.

THEOREM 4.1 (Ideal SgnQ test statistic). *Consider the testing problem* (1.6) *under the DCMM model* (1.1)-(1.4)*, where the condition* (2.16) *is satisfied under the alternative hypothesis. Suppose* $\theta_{\max} \to 0$ *and* $\|\theta\| \to \infty$ *as* $n \to \infty$*, and suppose* $|\lambda_2|/\sqrt{\lambda_1} \to \infty$ *under the alternative hypothesis. Then, under the null hypothesis, as* $n \to \infty$, $\mathbb{E}[\widetilde{Q}_n] = 0$, $\mathrm{Var}(\widetilde{Q}_n) = 8\|\theta\|^8 \cdot [1 + o(1)]$*, and* $(\widetilde{Q}_n - \mathbb{E}[\widetilde{Q}_n])/\sqrt{\mathrm{Var}(\widetilde{Q}_n)} \longrightarrow N(0,1)$ *in law. Furthermore, under the alternative hypothesis, as* $n \to \infty$, $\mathbb{E}[\widetilde{Q}_n] = \mathrm{tr}(\widetilde{\Omega}^4) + o(\|\theta\|^4)$ *and* $\mathrm{Var}(\widetilde{T}_n) \leq C[\|\theta\|^8 + (|\lambda_2|/\lambda_1)^6\|\theta\|^8\|\theta\|_3^6]$.

4.2. *The behavior of* $(Q_n^* - \widetilde{Q}_n)$. The Proxy SgnQ test statistic is defined as $Q_n^* = \sum_{i_1,i_2,i_3,i_4 (dist)} X^*_{i_1 i_2} X^*_{i_2 i_3} X^*_{i_3 i_4} X^*_{i_4 i_1}$. Inserting $X^*_{ij} = \widetilde{\Omega}_{ij} + W_{ij} + \delta_{ij}$ and expanding every bracket, we similarly obtain $3 \times 3 \times 3 \times 3 = 81$ different post-expansion sums, where 15 of them do not involve any δ term. The sum of the remaining 65 terms is $(Q_n^* - \widetilde{Q}_n)$. For each of these 65 post-expansion sums, we are summing over many 4-products, where each of them has the same number of $\widetilde{\Omega}$ factors, W factors, and δ factors, which we denote by $N_{\widetilde{\Omega}}, N_W$, and N_δ, respectively. According to $(N_{\widetilde{\Omega}}, N_W, N_\delta)$, we divide the 65 post-expansion sums into 10 different types. See Table 2, where we recall that $\alpha = |\lambda_2|/\lambda_1$.

TABLE 2
The 10 *types of the post-expansion sums for* $(Q_n^* - \widetilde{Q}_n)$. *Notations: same as in Table 1.*

Type	#	$(N_\delta,N_{\widetilde{\Omega}},N_W)$	Examples	Abs. Mean	Variance
Ia	4	(1, 0, 3)	$\sum_{\substack{i,j,k,\ell \\ (dist)}} \delta_{ij} W_{jk} W_{k\ell} W_{\ell i}$	0	$\le C\|\theta\|^2\|\theta\|_3^6 = o(\|\theta\|^8)$
Ib	8	(1, 1, 2)	$\sum_{\substack{i,j,k,\ell \\ (dist)}} \delta_{ij} \widetilde{\Omega}_{jk} W_{k\ell} W_{\ell i}$	0	$\le C\alpha^2\|\theta\|^4\|\theta\|_3^6 = o(\|\theta\|^8)$
	4		$\sum_{\substack{i,j,k,\ell \\ (dist)}} \delta_{ij} W_{jk} \widetilde{\Omega}_{k\ell} W_{\ell i}$	0	$\le C\alpha^2\|\theta\|^4\|\theta\|_3^6 = o(\|\theta\|^8)$
Ic	8	(1, 2, 1)	$\sum_{\substack{i,j,k,\ell \\ (dist)}} \delta_{ij} \widetilde{\Omega}_{jk} \widetilde{\Omega}_{k\ell} W_{\ell i}$	$\le C\alpha^2\|\theta\|^6 = o(\alpha^4\|\theta\|^8)$	$\le \frac{C\alpha^4\|\theta\|^{10}\|\theta\|_3^3}{\|\theta\|_1} = o(\alpha^6\|\theta\|^8\|\theta\|_3^6)$
	4		$\sum_{\substack{i,j,k,\ell \\ (dist)}} \delta_{ij} \widetilde{\Omega}_{jk} W_{k\ell} \widetilde{\Omega}_{\ell i}$	0	$\le \frac{C\alpha^4\|\theta\|^4\|\theta\|_3^9}{\|\theta\|_1} = o(\|\theta\|^8)$
Id	4	(1, 3, 0)	$\sum_{\substack{i,j,k,\ell \\ (dist)}} \delta_{ij} \widetilde{\Omega}_{jk} \widetilde{\Omega}_{k\ell} \widetilde{\Omega}_{\ell i}$	0	$\le \frac{C\alpha^6\|\theta\|^{12}\|\theta\|_3^3}{\|\theta\|_1} = O(\alpha^6\|\theta\|^8\|\theta\|_3^6)$
IIa	4	(2, 0, 2)	$\sum_{\substack{i,j,k,\ell \\ (dist)}} \delta_{ij} \delta_{jk} W_{k\ell} W_{\ell i}$	$\le C\|\theta\|^4 = o(\alpha^4\|\theta\|^8)$	$\le C\|\theta\|^2\|\theta\|_3^6 = o(\|\theta\|^8)$
	2		$\sum_{\substack{i,j,k,\ell \\ (dist)}} \delta_{ij} W_{jk} \delta_{k\ell} W_{\ell i}$	$\le C\|\theta\|^4 = o(\alpha^4\|\theta\|^8)$	$\le \frac{C\|\theta\|^6\|\theta\|_3^3}{\|\theta\|_1} = o(\|\theta\|^8)$
IIb	8	(2, 1, 1)	$\sum_{\substack{i,j,k,\ell \\ (dist)}} \delta_{ij} \delta_{jk} \widetilde{\Omega}_{k\ell} W_{\ell i}$	0	$\le C\alpha^2\|\theta\|^4\|\theta\|_3^6 = o(\|\theta\|^8)$
	4		$\sum_{\substack{i,j,k,\ell \\ (dist)}} \delta_{ij} \widetilde{\Omega}_{jk} \delta_{k\ell} W_{\ell i}$	$\le C\alpha\|\theta\|^4 = o(\alpha^4\|\theta\|^8)$	$\le \frac{C\alpha^2\|\theta\|^8\|\theta\|_3^3}{\|\theta\|_1} = o(\|\theta\|^8)$
IIc	4	(2, 2, 0)	$\sum_{\substack{i,j,k,\ell \\ (dist)}} \delta_{ij} \delta_{jk} \widetilde{\Omega}_{k\ell} \widetilde{\Omega}_{\ell i}$	$\le C\alpha^2\|\theta\|^6 = o(\alpha^4\|\theta\|^8)$	$\le \frac{C\alpha^4\|\theta\|^{14}}{\|\theta\|_1^2} = o(\alpha^6\|\theta\|^8\|\theta\|_3^6)$
	2		$\le \sum_{\substack{i,j,k,\ell \\ (dist)}} \delta_{ij} \widetilde{\Omega}_{jk} \delta_{k\ell} \widetilde{\Omega}_{\ell i}$	$\frac{C\alpha^2\|\theta\|^8}{\|\theta\|_1^2} = o(\alpha^4\|\theta\|^8)$	$\le \frac{C\alpha^4\|\theta\|^8\|\theta\|_3^6}{\|\theta\|_1^2} = o(\|\theta\|^8)$
IIIa	4	(3, 0, 1)	$\sum_{\substack{i,j,k,\ell \\ (dist)}} \delta_{ij} \delta_{jk} \delta_{k\ell} W_{\ell i}$	$\le C\|\theta\|^4 = o(\alpha^4\|\theta\|^8)$	$\le \frac{C\|\theta\|^6\|\theta\|_3^3}{\|\theta\|_1} = o(\|\theta\|^8)$
IIIb	4	(3, 1, 0)	$\le \sum_{\substack{i,j,k,\ell \\ (dist)}} \delta_{ij} \delta_{jk} \delta_{k\ell} \widetilde{\Omega}_{\ell i}$	$\le \frac{C\alpha\|\theta\|^6}{\|\theta\|_1^3} = o(\alpha^4\|\theta\|^8)$	$\le \frac{C\alpha^2\|\theta\|^8\|\theta\|_3^3}{\|\theta\|_1} = o(\|\theta\|^8)$
IV	1	(4, 0, 0)	$\sum_{\substack{i,j,k,\ell \\ (dist)}} \delta_{ij} \delta_{jk} \delta_{k\ell} \delta_{\ell i}$	$\le C\|\theta\|^4 = o(\alpha^4\|\theta\|^8)$	$\le \frac{C\|\theta\|^{10}}{\|\theta\|_1^2} = o(\|\theta\|^8)$

We now analyze $Q_n^* - \widetilde{Q}_n$. Consider the null hypothesis first. Under the null, $\widetilde{\Omega}$ is a zero matrix, so the nonzero post-expansion sums only include Type Ia, Type IIa, Type IIIa, and Type IV. It is seen that $|\mathbb{E}[Q_n^* - \widetilde{Q}_n]| \le C\|\theta\|^4$, and $\mathrm{Var}(Q_n^* - \widetilde{Q}_n) = o(\|\theta\|^8)$. Note that $\|\theta\|^8$ is the order of $\mathrm{Var}(\widetilde{Q}_n)$ under the null. The difference between the variance of Q_n^* and the variance of $\widetilde{Q}_n$ is negligible, but the difference between the mean of Q_n^* and the mean of $\widetilde{Q}_n$ is non-negligible. With lengthy calculations (see the supplementary material), we can show that $\mathbb{E}[Q_n^* - \widetilde{Q}_n] \sim 2\|\theta\|^4$. Therefore, $(Q_n^* - 2\|\theta\|^4)$ and $\widetilde{Q}_n$ have a negligible difference under the null.

Consider the alternative hypothesis next. From Table 2, $|\mathbb{E}[Q_n^* - \widetilde{Q}_n]| \le C(|\lambda_2|/\lambda_1)^2\|\theta\|^6$, where the major contribution is from Type Ic and Type IIc post-expansion sums. Under our assumptions for the alternative, $|\lambda_2|/\sqrt{\lambda_1} \to \infty$ and $\lambda_1 \asymp \|\theta\|^4$. It is easy to see that $|\mathbb{E}[Q_n^* - \widetilde{Q}_n]| = o(\lambda_2^4)$, where λ_2^4 is the order of $\mathrm{tr}(\widetilde{\Omega}^4)$ and $\mathbb{E}[\widetilde{Q}_n]$; see Lemma 2.3

and Theorem 4.1. Additionally, $\|\theta\|^4 = O(\lambda_1^2) = o(\lambda_2^4)$, which is also of a smaller order of $\mathbb{E}[\widetilde{Q}_n]$. We conclude that $\big|\mathbb{E}[Q_n^* - \widetilde{Q}_n - 2\|\theta\|^4]\big| = o(\mathbb{E}[\widetilde{Q}_n])$. From the table, $\mathrm{Var}(Q_n^* - \widetilde{Q}_n) \leq C(|\lambda_2|/\lambda_1)^6\|\theta\|^{12}\|\theta\|_3^3/\|\theta\|_1 + o(\|\theta\|^8)$, with the major contribution from Type Id. Here, the second term is smaller than $\mathrm{Var}(\widetilde{Q}_n)$, and the first term is upper bounded by $C(|\lambda_2|/\lambda_1)^6\|\theta\|^8\|\theta\|_3^6$ (using the universal inequality of $\|\theta\|^4 \leq \|\theta\|_1\|\theta\|_3^3$), which has a comparable order as $\mathrm{Var}(\widetilde{Q}_n)$. It follows that $\mathrm{Var}(Q_n^* - \widetilde{Q}_n - 2\|\theta\|^4) = \mathrm{Var}(Q_n^* - \widetilde{Q}_n) \leq C\mathrm{Var}(\widetilde{Q}_n)$. Combining the above, we obtain that the SNR of $(Q_n^* - 2\|\theta\|^4)$ and $\widetilde{Q}_n$ are at the same order.

These results are summarized in Theorem 4.2, which is proved in the supplement.

THEOREM 4.2 (Proxy SgnQ test statistic). *Consider the testing problem* (1.6) *under the DCMM model* (1.1)-(1.4)*, where the condition* (2.16) *is satisfied under the alternative hypothesis. Suppose $\theta_{\max} \to 0$ and $\|\theta\| \to \infty$ as $n \to \infty$, and suppose $|\lambda_2|/\sqrt{\lambda_1} \to \infty$ under the alternative hypothesis. Then, under the null hypothesis, as $n \to \infty$, $\mathbb{E}[(Q_n^* - 2\|\theta\|^4) - \widetilde{Q}_n] = o(\|\theta\|^4)$ and $\mathrm{Var}(Q_n^* - \widetilde{Q}_n) = o(\|\theta\|^8)$. Furthermore, under the alternative hypothesis, $\mathbb{E}[(Q_n^* - 2\|\theta\|^4) - \widetilde{Q}_n] = o((|\lambda_2|/\lambda_1)^4\|\theta\|^8)$ and $\mathrm{Var}(Q_n^* - \widetilde{Q}_n) \leq C(|\lambda_2|/\lambda_1)^6\|\theta\|^8\|\theta\|_3^6 + o(\|\theta\|^8)$.*

4.3. *The behavior of $(Q_n - Q_n^*)$.* Recall that $Q_n = \sum_{i_1,i_2,i_3,i_4 (dist)} X_{i_1 i_2} X_{i_2 i_3} X_{i_3 i_4} X_{i_4 i_1}$, where $X_{ij} = \widetilde{\Omega}_{ij} + W_{ij} + \delta_{ij} + r_{ij}$ for any $i \neq j$. Similar to Sections 4.1-4.2, we first expand every bracket in the definitions and obtain $4 \times 4 \times 4 \times 4 = 256$. Out of the 256 post-expansion sums in Q_n, $3 \times 3 \times 3 \times 3 = 81$ of them do not involve any r term and are contained in Q_n^*; this leaves a total of $256 - 81 = 175$ different post-expansion sums in $(Q_n - Q_n^*)$. In the supplementary material, we investigate the order of mean and variance of each of the 175 post-expansion sums in $(Q_n - Q_n^*)$. The calculations are very tedious: although we expect these post-expansion sums to be of a smaller order than the post-expansion sums in Sections 4.1-4.2, it is impossible to prove this argument rigorously using only some crude bounds (such as Cauchy-Schwarz inequality). Instead, we still need to do calculations for each post-expansion sum; details are in the supplementary material.

THEOREM 4.3 (Real SgnQ test statistic). *Consider the testing problem* (1.6) *under the DCMM model* (1.1)-(1.4)*, where the condition* (2.16) *is satisfied under the alternative hypothesis. Suppose $\theta_{\max} \to 0$ and $\|\theta\| \to \infty$ as $n \to \infty$, and suppose $|\lambda_2|/\sqrt{\lambda_1} \to \infty$ under the alternative hypothesis. Then, under the null hypothesis, as $n \to \infty$, $|\mathbb{E}[Q_n - Q_n^*]| = o(\|\theta\|^4)$ and $\mathrm{Var}(Q_n - Q_n^*) = o(|\theta\|^8)$. Under the alternative hypothesis, as $n \to \infty$, $|\mathbb{E}[Q_n - Q_n^*]| = o((|\lambda_2|/\lambda_1)^4\|\theta\|^8)$ and $\mathrm{Var}(Q_n - Q_n^*) = o((|\lambda_2|/\lambda_1)^6\|\theta\|^8|\theta\|_3^6) + o(\|\theta\|^8)$.*

5. Simulations.

We investigate the numerical performance of two Signed Polygon tests, the SgnT test (2.18) and the SgnQ test (2.20). We also include the EZ test [16] and the GC test [25] for comparison. For reasons mentioned in [25], we use a two-sided rejection region for EZ and a one-sided rejection region for GC.

Given (n, K), a scalar $\beta_n > 0$ that controls $\|\theta\|$, a symmetric nonnegative matrix $P \in \mathbb{R}^{K\times K}$, a distribution $f(\theta)$ on $\mathbb{R}_+$, and a distribution $g(\pi)$ on the standard simplex of $\mathbb{R}^K$, we generate two network adjacency matrices A^{null} and A^{alt}, under the null and the alternative, respectively, as follows: (i) Generate $\tilde{\theta}_1, \tilde{\theta}_2, \ldots, \tilde{\theta}_n$ *iid* from $f(\theta)$. Let $\theta_i = \beta_n \cdot \tilde{\theta}_i/\|\tilde{\theta}\|$, $1 \leq i \leq n$. (ii) Generate $\pi_1, \pi_2, \ldots, \pi_n$ *iid* from $g(\pi)$. (iii) Let $\Omega^{alt} = \Theta\Pi P\Pi'\Theta'$, where $\Theta = \mathrm{diag}(\theta_1, \cdots, \theta_n)$ and $\Pi = [\pi_1, \pi_2, \ldots, \pi_n]'$. Generate A^{alt} from Ω^{alt} according to Model (1.1). (iv) Let $\Omega^{null} = (a'Pa) \cdot \theta\theta'$, where $a = \mathbb{E}_g\pi \in \mathbb{R}^K$ is the mean vector of $g(\pi)$. Generate A^{null} from Ω^{null} according to Model (1.1). The pair $(\Omega^{null}, \Omega^{alt})$ is constructed in a way such that the corresponding networks have approximately the same expected average degree. This is the most subtle case for distinguishing two hypotheses (see Section 3).

It is of interest to explore different sparsity levels and also focus on the parameter settings where the SNR is neither too large nor too small. Therefore, for most experiments, we let $\beta_n = \|\theta\|$ range but fix the SNR at more or less the same level. See details below. For each parameter setting, we generate 200 networks under the null hypothesis and 200 networks under the alternative hypothesis, run all the four tests with a target level $\alpha = 5\%$, and then record the sum of percent of type I errors and percent of type II errors. For space limit, we do not report separately the percent of each type of errors but relegate these results to the supplementary material.

5.1. *Experiment 1.* We study the role of degree heterogeneity. Fix $(n, K) = (2000, 2)$. Let P be a 2×2 matrix with unit diagonal entries and all off-diagonal entries equal to b_n. Let $g(\pi)$ be the uniform distribution on $\{(0,1),(1,0)\}$. We consider three sub-experiments, Exp 1a-1c, where respectively we take $f(\theta)$ to be the following: (a) $\mathrm{Uniform}(2,3)$, (b) two-point distribution $0.95\delta_1 + 0.05\delta_3$, where δ_a is a point mass at a, and (c) $\mathrm{Pareto}(10, 0.375)$, where 10 is the shape parameter and 0.375 is the scale parameter. The degree heterogeneity is moderate in Exp 1a-1b, but more severe in Exp 1c. In such a setting, SNR is at the order of $\|\theta\|(1-b_n)$. Therefore, for each sub-experiment, we let $\beta_n = \|\theta\|$ vary while fixing the SNR to be $\|\theta\|(1-b_n) = 3.2$. The sum of Type I and Type II errors are displayed in Figure 3.

First, both the SgnQ test and the GC test are based on the counts of 4-cycles, but the GC test counts *non-centered* cycles and the SgnQ test counts *centered* cycles. As we pointed out in Section 1, counting *centered* cycles may have much smaller variances than counting *non-centered* cycles, especially in the less sparse case, and thus improves the testing power. This is confirmed by numerical results here, where the SgnQ test is consistently better than the GC test, significantly so in the less sparse case. Similarly, both the SgnT test and the EZ test are based on the counts of 3-cycles, but the EZ test counts *non-centered* cycles and the SgnT test counts *centered* cycles, and we expect that SgnT significantly improves EZ, especially in the less sparse case. This is also confirmed in the experiment.

Second, SgnQ and GC are order-4 graphlet counting statistics, and SgnT and EZ are order-3 graphlet counting statistics. In comparison, SgnQ significantly outperforms SgnT, and GC significantly outperforms EZ (in the more sparse case; see discussion below for the less sparse case). A possible explanation is that higher-order graphlet counting statistics have larger SNR. Investigation towards this direction is interesting, and we leave it to future study. Note that SgnQ is the best among all four tests.

Last, GC outperforms EZ in the more sparse case but underperforms EZ in the less sparse case. The reason for the latter is as follows. The biases of both tests are negligible in the more sparse case, but are non-negligible in the less sparse case, with that of GC much larger. In [29], we propose a bias correction method, where the performance of GC is significantly improved. However, GC continues to underperform SgnQ, because even with the bias corrected, it still has a variance that is unsatisfactorily large.

5.2. *Experiment 2.* We study the cases with larger K and a more complicated matrix of P. For some $b_n \in (0,1)$, let $\epsilon_n = \frac{1}{6}\min(1-b_n, b_n)$, and let P be the matrix with 1 on the diagonal and the off-diagonal entries iid drawn from $\mathrm{Unif}(b_n - \epsilon_n, b_n + \epsilon_n)$; once the P matrix is drawn, it is fixed throughout different repetitions. We consider two sub-experiments, Exp 2a and 2b. In Exp 2a, we take $(n, K) = (1000, 5)$, $f(\theta)$ to be $\mathrm{Pareto}(10, 0.375)$, and $g(\pi)$ to be the uniform distribution on $\{e_1, \cdots, e_K\}$ (the standard basis vectors of $\mathbb{R}^K$). We let β_n range but fix $\|\theta\|(1-b_n)$ at 4.5, so the SNR will not change drastically. In Exp 2b, we take $(n, K) = (3000, 10)$, $f(\theta)$ to be $0.95\delta_1 + 0.05\delta_3$, and $g(\pi) = 0.1\sum_{k=1}^{2}\delta_{e_k} +$

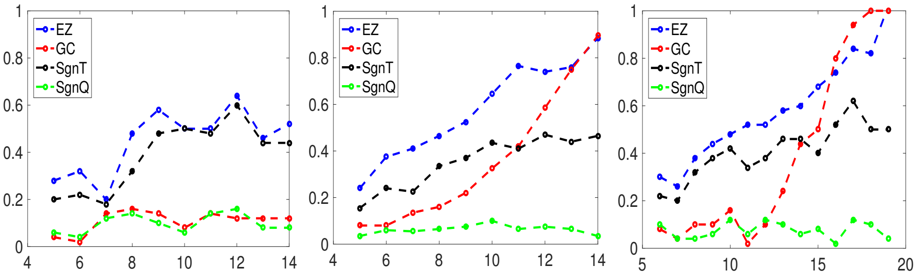

FIG 3. *From left to right: Experiment 1a, 1b, and 1c. The y-axis are the sum of Type I and Type II errors (testing level is fixed at* 5%*). The x-axis are $\|\theta\|$ or sparsity levels. Results are based on* 200 *repetitions.*

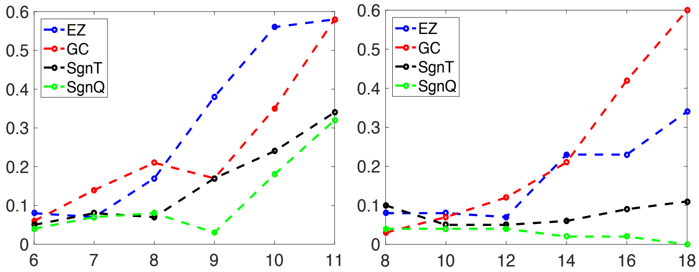

FIG 4. *From left to right: Experiment 2a and 2b. The y-axis are the sum of Type I and Type II errors (testing level is fixed at* 5%*). The x-axis are $\|\theta\|$ or sparsity levels. Results are based on* 200 *repetitions.*

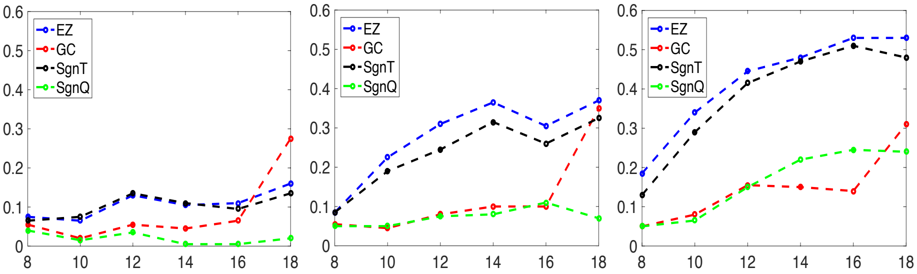

FIG 5. *From left to right: Experiment 3a, 3b, and 3c. The y-axis are the sum of Type I and Type II errors (testing level is fixed at* 5%*). The x-axis are $\|\theta\|$ or sparsity levels. Results are based on* 200 *repetitions.*

$0.15\sum_{k=3}^{6}\delta_{e_k}+0.05\sum_{k=7}^{10}\delta_{e_k}$ (so to have unbalanced community sizes). Similarly, we let β_n range but fix $\|\theta\|(1-b_n)=5.2$. The sum of Type I & II errors are shown in Figure 4.

In these examples, EZ and GC underperform SgnT and SgnQ, especially in the less sparse case, and the performances of SgnT and SgnQ are more similar to each other, compared to those in Experiment 1. In these examples, we have larger K, more complicated P, and unbalanced community sizes, and the performance of SgnT and SgnQ test statistics suggest that they are relatively robust.

5.3. *Experiment 3.* We investigate the role of mixed-membership. We have three sub-experiments, Exp 3a-3c. where the memberships are not-mixed, lightly mixed, and significantly mixed, respectively. For all sub-experiments, we take $(n,K)=(2000,3)$ and $f(\theta)$ to be $\mathrm{Unif}(2,3)$. For Exp 3a, we let $g_1(\pi)=0.4\delta_{e_1}+0.3\delta_{e_2}+0.3\delta_{e_3}$. In Exp 3b, we let $g_2(\pi)=0.3\sum_{k=1}^{3}\delta_{e_k}+0.1\cdot\mathrm{Dirichlet}$, and in Exp 3c, we let $g_3(\pi)=0.25\sum_{k=1}^{3}\delta_{e_k}+0.25\cdot\mathrm{Dirichlet}$, where Dirichlet represents the symmetric K-dimensional Dirichlet distribution. In Exp 3a-3b, we let β_n range while $(1-b_n)\|\theta\|$ is fixed at 4.2 so the SNR's are roughly the same. In Exp 3c, we also let β_n range but $(1-b_n)\|\theta\| = 4.5$ (the SNR's need to be slightly larger to counter the effect of mixed-membership, which makes the testing problem harder).

The sum of Type I and Type II errors are presented in Figure 5. First, the results confirm that mixed-memberships make the testing problem harder. For example, the value of $\|\theta\|(1-b_n)$ in Exp 3c is higher than that of Exp 3a-3b, but the testing errors are higher, due to that the memberships in Exp 3c are more mixed. Second, SgnQ consistently outperforms EZ and SgnT. Third, GC is comparable with SgnQ in the more sparse case, but performs unsatisfactorily in the less sparse case, for reasons explained before. Last, in these settings, SgnT is uniformly better than EZ, and more so when the memberships become more mixed.

5.4. *Experiment 4.* We vary the size of network and study its impact on testing errors. We fix $K=2$ and let P be a 2×2 matrix with unit diagonals and off-diagonals equal to b_n. Let $g(\pi)$ be the uniform distribution on $\{(0,1),(1,0)\}$ and let $f(\theta)$ be $\mathrm{Pareto}(8, 0.375)$. We let n ranges from $\{100, 300, 1000, 3000\}$. Note that in our data generating process, $\beta_n = \|\theta\|$ controls the sparsity level and $(1-b_n)\|\theta\|$ is the SNR. As n varies, we fix $\beta_n = 4$ and change b_n accordingly so that the SNR is fixed at 3. The results are in Table 3. This is a sparse setting, therefore, the biases in EZ and GC are negligible and they both control the Type I error well. The SgnT and SgnQ tests also control the Type I error well. In terms of the Type II errors, GC and SgnQ are better than EZ and SgnT. The results are relatively stable as n varies.

TABLE 3

Experiment 4. Numbers in each cell are Type I error, Type II error, and their sum.

n	100	300	1000	3000
EZ	(.025, .22, .245)	(.055, .26, .315)	(.05, .27, .32)	(.06, .275, .335)
GC	(.02, .02, .04)	(.06, .02, .08)	(.04, .005, .045)	(.04, .005, .045)
SgnT	(.01, .15, .16)	(.04, .14, .18)	(.065, .175, .24)	(.06, .14, .2)
SgnQ	(.05, .015, .02)	(.04, .005, .045)	(.04, 0, .04)	(.02, .005, .025)

6. Discussions. A closely related idea is to use $\|A - \hat\eta\hat\eta'\|$ as the test statistics. To see why this is a reasonable choice, consider the proxy test statistic $\|A - \eta^*(\eta^*)'\|$, where we recall that $\eta^* = \theta$ under the null; see (1.12). Therefore, $A - \eta^*(\eta^*)'$ is equal to W and $(\Omega - (\eta^*(\eta^*)') + W$, under the null and the alternative, respectively. The test has reasonable power, as $\|A - \eta^*(\eta^*)'\|$ is expected to be bigger in the alternative than in the null. Another related idea is to extend the Signed Polygon to address the problem of testing whether $K = k_0$ vs. $K > k_0$, where $k_0 > 1$ is a prescribed integer. Let $\hat\Omega = \sum_{k=1}^{k_0}\hat\lambda_k\hat\xi_k\hat\xi_k'$, where $\hat\lambda_k$ are the k-th eigenvalue of A, arranged in the descending order in magnitude, and $\hat\xi_k$ is the corresponding eigenvector. The Signed Polygon test statistic can then be extended to $U_{n,k_0}^{(m)} = \sum_{i_1,i_2,\dots,i_m(dist)}(A_{i_1i_2} - \hat\Omega_{i_1i_2})(A_{i_2i_3} - \hat\Omega_{i_2i_3})\dots(A_{i_mi_1} - \hat\Omega_{i_mi_1})$. See [26] for more discussion. It remains unclear whether these test statistics are optimally adaptive, and we leave the study to the future.

Another testing idea would be using the first eigenvalue of $\widetilde{A} = \hat\theta^{-1}A\hat\theta^{-1} - \hat b\mathbf{1}_n\mathbf{1}_n'$, for a reasonable estimate $\hat\theta$ for θ and a proper $\hat b$. Unfortunately, even if $\hat\theta = \theta$, the distribution of the test is unknown for general cases. In fact, this is essentially the approaches proposed in [8, 31]). Both papers showed that in the dense case of $\theta_1 = \theta_2 = \ldots = \theta_n = O(1)$, the largest eigenvalue of $\widetilde{A}$ (when standardized) converges to the Tracy-Widom law. Unfortunately, the approaches have been focused on the more idealized SBM model and the less sparse case where $\theta_1 = \theta_2 = \ldots = \theta_n = \sqrt{\alpha_n} \geq O(n^{-1/6})$, and the limiting distribution remains unknown for other cases.

The testing problem is also closely related to the problem of estimating K. In fact, we can cast the estimation problem as a sequential testing problem where we test $K = k_0$ vs. $K > k_0$ for $k_0 = 1, 2, 3, \ldots$, and estimate K to be the smallest k_0 where we accept the null.

Note also the lower bound argument for the global testing problem sheds useful insight for many other problems (e.g., estimating K, community detection, mixed-membership). Take the problem of estimating K for example. Given an alternative setting, if we can not distinguish it from some null setting, then the underlying parameter K is not estimable.

In a high level, these ideas, together with the Signed Polygon, are related to the ideas in [21] on testing $K = k_0$ vs. $K > k_0$, in [31] on goodness of fit, and in [30] on estimating K. However, the focus of these works are on the more idealized model where we don't have degree heterogeneity, and how to extend their ideas to the current setting remains unclear.

SUPPLEMENTARY MATERIAL

Additional Results and Technical Proofs. The supplemental material contains the results not reported in the main article due to space limit and the proofs of all theorems and lemmas. ().

REFERENCES

[1] AIROLDI, E., BLEI, D., FIENBERG, S. and XING, E. (2008). Mixed membership stochastic blockmodels. *J. Mach. Learn. Res.* **9** 1981–2014.

[2] BANERJEE, D. (2018). Contiguity and non-reconstruction results for planted partition models: the dense case. *Electron. J. Probab.* **23** 485–512.

[3] BANERJEE, D. and MA, Z. (2017). Optimal hypothesis testing for stochastic block models with growing degrees. *arXiv:1705.05305*.

[4] BANKS, J., MOORE, C., NEEMAN, J. and NETRAPALLI, P. (2016). Information-theoretic thresholds for community detection in sparse networks. In *COLT* 383–416.

[5] BÉJAR, J., ÁLVAREZ, S., GARCÍA, D., GÓMEZ, I., OLIVA, L., TEJEDA, A. and VÁZQUEZ-SALCEDA, J. (2016). Discovery of spatio-temporal patterns from location-based social networks. *J. Exp. Theor. Artif. Intell.* **28** 313-329.

[6] BHATTACHARYYA, S. and BICKEL, P. J. (2015). Subsampling bootstrap of count features of networks. *Ann. Statist.* **43** 2384–2411.

[7] BICKEL, P. J., CHEN, A. and LEVINA, E. (2011). The method of moments and degree distributions for network models. *Ann. Statist.* **39** 2280–2301.

[8] BICKEL, P. J. and SARKAR, P. (2016). Hypothesis testing for automated community detection in networks. *J. R. Statist. Soc. B* **78** 253–273.

[9] BRUALDI, R. (1974). The DAD theorem for arbitrary row sums. *Proc. Amer. Math. Soc* **45** 189–194.

[10] BUBECK, S., DING, J., ELDAN, R. and RÁCZ, M. Z. (2016). Testing for high-dimensional geometry in random graphs. *Random Struct. Algor.* **49** 503–532.

[11] CHATTERJEE, S. (2015). Matrix estimation by Universal Singular Value Thresholding. *Ann. Statist.* **43** 177–214.

[12] DALL'AMICO, L., COUILLET, R. and TREMBLAY, N. (2019). Revisiting the Bethe-Hessian: Improved Community Detection in Sparse Heterogeneous Graphs. In *Adv. Neural. Inf. Process. Syst.* 4037–4047. Curran Associates, Inc.

[13] DONOHO, D. and JIN, J. (2004). Higher criticism for detecting sparse heterogeneous mixtures. *Ann. Statist.* **32** 962–994.

[14] DU, H. and YANG, S. J. (2011). Discovering collaborative cyber attack patterns using social network analysis. In *Proceedings of the 4th International Conference on Social Computing, Behavioral-cultural Modeling and Prediction*. *SBP'11* 129–136. Springer-Verlag.

[15] FU, Y.-H., HUANG, C.-Y. and SUN, C.-T. (2015). Using global diversity and local topology features to identify influential network spreaders. *Physica A.* **433** 344 - 355.

[16] GAO, C. and LAFFERTY, J. (2017). Testing for global network structure using small subgraph statistics. *arXiv:1710.00862*.

[17] GIRVAN, M. and NEWMAN, M. (2002). Community structure in social and biological networks. *Proc. Natl. Acad. Sci.* **99** 7821–7826.

[18] GULIKERS, L., LELARGE, M. and MASSOULIÉ, L. (2017). Non-Backtracking Spectrum of Degree-Corrected Stochastic Block Models. In *8th Innovations in Theoretical Computer Science Conference (ITCS 2017)*. Schloss Dagstuhl-Leibniz-Zentrum fuer Informatik.

[19] GULIKERS, L., LELARGE, M. and MASSOULIÉ, L. (2018). An impossibility result for reconstruction in the degree-corrected stochastic block model. *Ann. Appl. Probab.* **28** 3002–3027.

[20] HOLLAND, P. W., LASKEY, K. B. and LEINHARDT, S. (1983). Stochastic blockmodels: First steps. *Social networks* **5** 109–137.

[21] HU, J., QIN, H., YAN, T., ZHANG, J. and ZHU, J. (2017). Using maximum entry-wise deviation to test the goodness-of-fit for stochastic block models. *arXiv:1703.06558*.

[22] INGSTER, Y. I., TSYBAKOV, A. B. and VERZELEN, N. (2010). Detection boundary in sparse regression. *Electron. J. Statist.* **4** 1476–1526.

[23] JIN, J. (2015). Fast community detection by SCORE. *Ann. Statist.* **43** 57–89.

[24] JIN, J., KE, Z. T. and LUO, S. (2017). Estimating network memberships by simplex vertices hunting. *arXiv:1708.07852*.

[25] JIN, J., KE, Z. T. and LUO, S. (2018). Network global testing by counting graphlets. In *Proceedings of the 35th International Conference on Machine Learning, ICML 2018, Stockholmsmässan, Stockholm, Sweden, July 10-15, 2018* 2338–2346.

[26] JIN, J., LI, W. and WANG, T. (2019). A spectral approach for network global testing. *Manuscript*.

[27] JOHNSON, C. and REAMS, R. (2009). Scaling of symmetric matrices by positive diagonal congruence. *Linear Multilinear A* **57** 123-140.

[28] KARRER, B. and NEWMAN, M. (2011). Stochastic blockmodels and community structure in networks. *Phys. Rev. E* **83** 016107.

[29] KE, Z. T. (2019). Comparison of different network testing procedures. *Manuscript*.

[30] LE, C. M. and LEVINA, E. (2015). Estimating the number of communities in networks by spectral methods. *arXiv:1507.00827*.

[31] LEI, J. (2016). A goodness-of-fit test for stochastic block models. *Ann. Statist.* **44** 401–424.

[32] LI, T., LEI, L., BHATTACHARYYA, S., SARKAR, P., BICKEL, P. J. and LEVINA, E. (2018). Hierarchical community detection by recursive partitioning. *arXiv:1810.01509*.

[33] MA, Z. and WU, Y. (2015). Computational barriers in minimax submatrix detection. *Ann. Statist.* **43** 1089–1116.

[34] MARSHALL, A. and OLKIN, I. (1968). Scaling of matrices to achieve specified row and column sums. *Numer. Math.* **12** 83–90.

[35] MAUGIS, P.-A. G., PRIEBE, C. E., OLHEDE, S. C. and WOLFE, P. J. (2017). Statistical inference for network samples using subgraph counts. *arXiv:1701.00505*.

[36] MOSSEL, E., NEEMAN, J. and SLY, A. (2015). Reconstruction and estimation in the planted partition model. *Probab. Theory Relat. Fields* **162** 431–461.

[37] PAUL, D. (2007). Asymptotics of sample eigenstructure for a large dimensional spiked covariance model. *Stat. Sin.* **17** 1617–1642.

[38] RADICCHI, F., CASTELLANO, C., CECCONI, F., LORETO, V. and PARISI, D. (2004). Defining and identifying communities in networks. *Proc. Natl. Acad. Sci.* **101** 2658–2663.

[39] SALDANA, D., YU, Y. and FENG, Y. (2017). How many communities are there? *J. Comput. Graph Stat.* **26** 171-181.

[40] SINKHORN, R. (1974). Diagonal equivalence to matrices with prescribed row and column sums. II. *Proc. Amer. Math. Soc* **45** 195–198.

[41] WANG, Y. R. and BICKEL, P. J. (2017). Likelihood-based model selection for stochastic block models. *Ann. Statist.* **45** 500–528.

[42] ZHANG, Y., LEVINA, E. and ZHU, J. (2014). Detecting overlapping communities in networks with spectral methods. *arXiv:1412.3432*.

[43] ZHAO, Y., LEVINA, E. and ZHU, J. (2011). Community extraction for social networks. *Proc. Natl. Acad. Sci.* **108** 7321–7326.