BIOMETRIC METHODOLOGY

Testing for association in multiview network data

Lucy L. Gao | Daniela Witten | Jacob Bien

1 Department of Statistics and Actuarial Science, University of Waterloo, Waterloo, Ontario, Canada
2 Departments of Statistics and Biostatistics, University of Washington, Seattle, Washington, USA
3 Department of Data Sciences and Operations, University of Southern California, Los Angeles, California, USA

Correspondence
Lucy L. Gao, Department of Statistics and Actuarial Science, University of Waterloo, Waterloo, Ontario, Canada.
Email: lucy.gao@uwaterloo.ca

Funding information
Simons Foundation, Grant/Award Number: Simons Investigator Award No. 560585;
National Institutes of Health, Grant/Award Numbers: DP5OD009145, R01GM123993, National Science Foundation, Grant/Award Numbers: CAREER Award DMS-1252624, CAREER Award DMS-1653017

Abstract
In this paper, we consider data consisting of multiple networks, each composed of a different edge set on a common set of nodes. Many models have been proposed for the analysis of such multiview network data under the assumption that the data views are closely related. In this paper, we provide tools for evaluating this assumption. In particular, we ask: given two networks that each follow a stochastic block model, is there an association between the latent community memberships of the nodes in the two networks? To answer this question, we extend the stochastic block model for a single network view to the two-view setting, and develop a new hypothesis test for the null hypothesis that the latent community memberships in the two data views are independent. We apply our test to protein–protein interaction data from the HINT database. We find evidence of a weak association between the latent community memberships of proteins defined with respect to binary interaction data and the latent community memberships of proteins defined with respect to cocomplex association data. We also extend this proposal to the setting of a network with node covariates. The proposed methods extend readily to three or more network/multivariate data views.

KEYWORDS
community detection, data integration, multiview data, node covariates, stochastic block model

1 INTRODUCTION

A network consists of the pairwise relationships (edges) between objects of interest (nodes). For example, nodes could correspond to proteins, with edges representing physical interactions, or nodes could correspond to people, with edges representing social interactions. Of the many models for network data (Erdős and Rényi, 1960; Holland and Leinhardt, 1981; Hoff et al., 2002), one of the best known is the stochastic block model (SBM) (Holland et al., 1983), which assumes that nodes belong to latent communities, and that the probability of an edge between a pair of nodes is a function of their community memberships only.

Multiple sets of edges are often available on a common set of nodes, as shown in Figure 1A. Consider a pair of protein–protein interaction networks in which the nodes correspond to proteins. In one network, the edges represent physical interactions, and in the other, they represent cocomembership in a protein complex. Another often-encountered scenario involves a single network, with a set of covariates corresponding to each node, as shown in Figure 1B. For instance, we might have a social network along with p demographic covariates for each member of the network. Both Figures 1A and 1B are examples of the multiview data setting (Sun, 2013). We will refer to the two networks in Figure 1A, or the network and the covariates corresponding to the nodes in Figure 1B as two data views.

Extensions of network models to the multiview data setting (Fosdick and Hoff, 2015; Han et al., 2015; Gollini and Murphy, 2016; Binkiewicz et al., 2017; Salter-Townshend...
and McCormick, 2017; D’Angelo et al., 2019) often assume that the data views are closely related. For example, extensions of the SBM typically assume that the latent communities within each network view are closely related (Han et al., 2015; Peixoto, 2015; Stanley et al., 2016; Binkiewicz et al., 2017; Stanley et al., 2019).

In this paper, we propose a test of the assumption that the latent communities are related. Why is this important? First of all, we should check whether two data views are, in fact, associated before we fit a model that relies on this assumption. Second, the relationship between the views may itself be of interest, and the test that we propose will allow us to assess this relationship. For example, such a tool can help shed light on whether the two distinct definitions of protein interactions capture similar versus complementary latent structures. Likewise, it can provide insight about whether peoples’ social interactions and demographics are related. Gao et al. (2020) investigated a similar problem for two multivariate data views, but did not consider the case where one or both views are networks.

To this end, we extend the SBM to the multiview network setting (Figure 1A) without assuming that the network views are closely related. We then ask: are the latent communities within each network view associated? Similarly, for the case of a network view and a multivariate view (Figure 1B), we model the network view with a SBM and model the multivariate view with a finite mixture model (FMM), without assuming that the views are closely related. We then ask: are the latent communities within the network data view and the latent clusters within the multivariate data view associated?

The rest of the paper is organized as follows. We review the SBM in Section 2. We extend the SBM to two network data views in Section 3, and develop a test for association between the latent communities within each view in Section 4. We develop a related test for the case of a network view and a multivariate view in Section 5. We review related literature in Section 6, and explore the performance of our tests via simulation in Section 7. In Section 8, we apply the test from Section 4 to protein networks from the HINT database (Das and Yu, 2012b). Section 9 provides a discussion.

2 | THE STOCHASTIC BLOCK MODEL (Holland et al., 1983)

In this section, we briefly review the SBM proposed by Holland et al. (1983) for a single network; see Matias and Robin (2014) for a detailed review.

2.1 | Model and notation

Let \( X \in \{0, 1\}^{n \times n} \) be the adjacency matrix of an undirected, unweighted network with \( n \) nodes and no self-loops, so that \( X \) is symmetric and \( X_{ii} = 0 \) for \( i = 1, 2, \ldots, n \). We assume that the nodes are partitioned into \( K \) communities, with unobserved memberships given by a latent random vector \( Z = (Z_1, \ldots, Z_n) \) with independent and identically distributed (i.i.d.) elements and \( P(Z_i = k) = \pi_k \) for \( k \in \Delta^K \equiv \{k \in \mathbb{R}^K : 1/K \pi = 1, \pi_k > 0\} \). Conditional on \( Z \), the edges are independently drawn from a Bernoulli distribution, with \( P[X_{ij} = 1 \mid Z] = \theta_{Z_i, Z_j} \) for a symmetric matrix \( \theta \in [0, 1]^{K \times K} \). It follows that

\[
f(X \mid Z) = \prod_{i=1}^{n} \prod_{j=1}^{i-1} \left( \theta_{Z_i, Z_j} \right) X_{ij} \left( 1 - \theta_{Z_i, Z_j} \right)^{1-X_{ij}},
\]

\[
P(Z = z) = \prod_{i=1}^{n} \pi_{z_i},
\]

(1)

2.2 | Approximate pseudolikelihood function

As a result of (1), the log-likelihood function for the SBM is given by

\[
\ell(\theta, \pi; X) \equiv \log \left( \sum_{z_1=1}^{K} \cdots \sum_{z_n=1}^{K} \left( \prod_{i=1}^{n-1} \prod_{j=1}^{i} (\theta_{z_i, z_j})^{X_{ij}} \right) \left( 1 - \theta_{z_i, z_j} \right)^{1-X_{ij}} \right) \times \left( \prod_{i=1}^{n} \pi_{z_i} \right).
\]

(2)
Equation (2) sums over $K^n$ terms, and is thus computationally intractable. Therefore, Amini et al. (2013) developed an approximate pseudolikelihood function, in the sense of Besag (1975). We briefly review this approach; see Web Appendix A for a detailed review.

Let $\hat{Z} \in \{1, \ldots, K\}^n$ be the results of applying spectral clustering with perturbations (Amini et al., 2013) to $X$. Define $\hat{b} \in \mathbb{R}^{n \times K}$ with rows $\hat{b}_i$ and $\hat{b}_{im} = \sum_{j=1}^n X_{ij} \mathbb{I} (\hat{Z}_j = m)$, and let $d = X_1 n$. Here, $\hat{b}_{im}$ is the number of edges connecting the $i$th node to the $m$th estimated community in $\hat{Z}$, and $d$ contains the degrees of the $n$ nodes. Let $R$ be the confusion matrix between $\hat{Z}$ and $Z$, and define the $K \times K$ matrix $\eta = (\text{diag}(\theta R_1))^{-1} \theta R$, with rows $\eta_1, \ldots, \eta_K \in \Delta_+^K$. Let $g(\cdot; N, \eta)$ denote the probability mass function of a Multinomial$(N, \eta_1, \ldots, \eta_K)$ random variable. Amini et al. (2013) treated $\hat{Z}$ and $\eta$ as fixed and showed that

$$\hat{b} \mid d, Z \sim \prod_{i=1}^n g (\hat{b}_i; d_i, \eta_{Z_i}) \quad \text{(3)}$$

where $\sim$ denotes “approximately distributed as.” Ignoring any dependence between $Z$ and $d$, and marginalizing over $Z$ in (3) to approximate the conditional distribution of $\hat{b}$ given $d$, yields the following log-pseudolikelihood function:

$$\ell_{PL}(\eta; \pi; \hat{b} \mid d) \equiv \sum_{i=1}^n \log \left( \sum_{k=1}^K \pi_k g (\hat{b}_i; d_i, \eta_k) \right) \quad \text{(4)}$$

This can be viewed as the log-likelihood function of an FMM (McLachlan and Peel, 2000) with $K$ components, of which the $k$th component has prior probability $\pi_k$ and density function $g (\hat{b}_i; d_i, \eta_k)$. (3)–(4) are i.i.d.

### 3 | A STOCHASTIC BLOCK MODEL FOR TWO NETWORK DATA VIEWS

In this section, we extend the SBM to the setting of two network data views, and derive approximate pseudolikelihood functions for the proposed multiview SBM.

#### 3.1 | Model and notation

Suppose that we have two network views on a common set of $n$ nodes, as in Figure 1A, for example, a binary network and a complex network on $n$ proteins. We assume that the networks are undirected, unweighted, and have no self-loops. Let $X^{(1)}, X^{(2)} \in \{0, 1\}^{n \times n}$ be the symmetric adjacency matrices of the two networks, where $X^{(l)}_{ii} = 0$ for $i = 1, 2, \ldots, n$ and $l = 1, 2$.

We model $X^{(1)}$ with an SBM (Section 2.1) with $K^{(1)}$ communities, and $X^{(2)}$ with an SBM with $K^{(2)}$ communities. It follows from (1) that for $l = 1, 2$,

$$f (X^{(l)} \mid Z^{(l)}) = \prod_{i=1}^n \prod_{j=1}^{n-l-1} \left( \frac{\theta^{(l)}_{i,j}^{(l)}}{Z^{(l)}_{i,j}^{(l)}} \right) \left( 1 - \frac{\theta^{(l)}_{i,j}^{(l)}}{Z^{(l)}_{i,j}^{(l)}} \right)^{1 - \gamma^{(l)}_{i,j}^{(l)}}$$

for a symmetric matrix $\theta^{(l)} \in [0, 1]^{K^{(l)} \times K^{(l)}}$ and $\pi^l \in \Delta_+^{K^{(l)}}$. Here, for $l = 1, 2$, $Z^{(l)}$ represents the latent community memberships for the $n$ nodes within the $l$th network data view. We assume that the $n$ pairs $(Z^{(1)}_i, Z^{(2)}_i)$ are i.i.d.

It follows from applying Proposition 1 to each of the $n$ pairs of categorical variables $(Z^{(1)}_i, Z^{(2)}_i)$, that there exists a unique $K^{(1)} \times K^{(2)}$ matrix $C \in C_{\Delta_+^{K^{(1)}}, \Delta_+^{K^{(2)}}}$ such that

$$P (Z^{(1)} = z^{(1)}_i, Z^{(2)} = z^{(2)}_i) = \prod_{i=1}^n P (z^{(1)}_i = z^{(1)}_i, z^{(2)}_i = z^{(2)}_i)$$

$$= \prod_{i=1}^n \pi^{(1)}_{z^{(1)}_i} \pi^{(2)}_{z^{(2)}_i} C_{z^{(1)}_i, z^{(2)}_i} \quad \text{(6)}$$

where the first equality follows from the independence of the $n$ pairs $(Z^{(1)}_i, Z^{(2)}_i)$. Here, $C_{kk'} = \frac{P (Z^{(1)}_i = k, Z^{(2)}_i = k')}{P (Z^{(1)}_i = k) P (Z^{(2)}_i = k')}$ describes the dependence between the $k$th community in the first view and the $k'$th community in the second view, with $C_{kk'} > 1$ indicating independence, $C_{kk'} < 1$ indicating negative dependence, and $C_{kk'} > 1$ indicating positive dependence.

#### 3.2 | Approximate pseudolikelihood function

The log-likelihood function of model (5)–(6) is given by
\( \ell(\theta^{(1)}, \theta^{(2)}, \pi^{(1)}, \pi^{(2)}, C; X^{(1)}, X^{(2)}) \)
\[
\equiv \log \left( \sum_{z_{i}^{(1)}=1}^{K^{(1)}} \cdots \sum_{z_{n}^{(1)}=1}^{K^{(1)}} \sum_{z_{i}^{(2)}=1}^{K^{(2)}} \sum_{z_{n}^{(2)}=1}^{K^{(2)}} \prod_{l=1}^{2} \prod_{l'}=1^{n-1} \left( \pi_{z_{i}^{(1)}, z_{l}^{(1)}} \right) \left( \pi_{z_{i}^{(2)}, z_{l}^{(2)}} \right) \right) 
\left( 1 - \theta_{z_{i}^{(1)} z_{l}^{(1)}} \right) \left( 1 - \theta_{z_{i}^{(2)} z_{l}^{(2)}} \right) \prod_{l=1}^{2} \prod_{l'=1}^{n-1} \left( \theta_{z_{i}^{(l)} z_{l'}^{(l)}} \right) \prod_{l=1}^{2} \prod_{l'=1}^{n-1} \left( \delta_{z_{i}^{(l)} z_{l'}^{(l)}} \right) \right). 
\]

Equation (7) is computationally intractable, because it involves summing over \((K^{(1)} K^{(2)})^n\) terms. Thus, we will derive an approximate pseudolikelihood function for model (5)–(6). For \(l = 1, 2\), let \(\hat{Z}^{(l)} \in \{1, \ldots, K^{(l)}\}^n\) be the results of applying spectral clustering with perturbations (Amini et al., 2013) to \(X^{(l)}\), let \(\hat{B}^{(l)}\) be the \(n \times K^{(l)}\) matrix defined by \(\hat{b}_{im}^{(l)} \equiv \sum_{i=1}^{n} X_{ij}^{(l)} \mathbb{1}[Z_{j}^{(l)} = m]\), and let \(d^{(l)} = X^{(l)T} 1_n\).

Here, for the \(l\)th network, \(\hat{b}_{im}^{(l)}\) is the number of edges connecting the \(l\)th node to the \(m\)th estimated community, and \(d_{i}^{(l)}\) is the degree of the \(i\)th node. We write

\[
f(\hat{b}^{(1)}, \hat{b}^{(2)} | d^{(1)}, d^{(2)}, Z^{(1)}, Z^{(2)}) = \frac{f(\hat{b}^{(1)}, \hat{b}^{(2)}, d^{(1)}, d^{(2)} | Z^{(1)}, Z^{(2)})}{f(d^{(1)}, d^{(2)} | Z^{(1)}, Z^{(2)})} = \prod_{l=1}^{2} \frac{f(\hat{b}^{(l)}, d^{(l)} | Z^{(l)})}{f(d^{(l)} | Z^{(l)})} = \prod_{l=1}^{2} \frac{f(\hat{b}^{(l)}, d^{(l)} | Z^{(l)})}{f(d^{(l)} | Z^{(l)})}, \tag{8}
\]

where the first and third equalities follow from the definition of a conditional density, and the second equality follows from the fact that \(X^{(1)} \perp X^{(2)} | Z^{(1)}, Z^{(2)}\) and \(X^{(1)} \perp Z^{(2)} | Z^{(1)}\) and \(X^{(2)} \perp Z^{(1)} | Z^{(2)}\) (Section 3.1). Let \(\hat{R}^{(l)}\) be the confusion matrix between \(\hat{Z}^{(l)}\) and \(Z^{(l)}\) and let \(\eta^{(l)} = \text{diag}(\hat{R}^{(l)} \hat{R}^{(l)T} 1_{K^{(l)}})\). As in Amini et al. (2013), we treat \(\hat{Z}^{(l)}\) and \(\eta^{(l)}\) as fixed, and apply (3) in Section 2.2 to approximate \(f(\hat{b}^{(l)} | Z^{(l)}, d^{(l)})\) in (8), which yields

\[
f(\hat{b}^{(1)}, \hat{b}^{(2)} | d^{(1)}, d^{(2)}, Z^{(1)}, Z^{(2)}) \approx \prod_{l=1}^{2} \prod_{i=1}^{n} g(\hat{b}_{i}^{(l)}, d_{i}^{(l)}, \eta^{(l)}_{z_{i}^{(l)}},) \tag{9}
\]

Ignoring any dependence between \((d^{(1)}, d^{(2)})\) and \((Z^{(1)}, Z^{(2)})\) and marginalizing over the latent community memberships \(Z^{(1)}\) and \(Z^{(2)}\) in (9) to approximate the conditional distribution of \(\hat{b}^{(1)}\) and \(\hat{b}^{(2)}\) given \(d^{(1)}\) and \(d^{(2)}\) yields the following log-pseudolikelihood function:

\[
\ell_{PL}(\eta^{(1)}, \eta^{(2)}, \pi^{(1)}, \pi^{(2)}, C; \hat{b}^{(1)}, \hat{b}^{(2)} | d^{(1)}, d^{(2)})
\]
\[
\equiv \text{log} \sum_{k=1}^{K^{(1)}} \sum_{k'=1}^{K^{(2)}} \pi_{k}^{(1)} \pi_{k'}^{(2)} C_{kk'} g(\hat{b}_{k}^{(1)}, d_{k}^{(1)}, \eta_{k}^{(1)}) g(\hat{b}_{k'}^{(2)}, d_{k'}^{(2)}, \eta_{k'}^{(2)}), \tag{10}
\]

This closely resembles the log-likelihood function of the FMM for two multivariate data views from Gao et al. (2020).

## 4 ARE TWO NETWORK VIEWS’ COMMUNITY MEMBERSHIPS ASSOCIATED?

Recall from (6) that \(P(Z^{(1)} = z^{(1)}, Z^{(2)} = z^{(2)}) = \prod_{i=1}^{n} \pi_{z_{i}^{(1)}, z_{i}^{(2)}}, c_{z_{i}^{(1)}, z_{i}^{(2)}}, \) where \(C \in C_{\pi^{(1)}, \pi^{(2)}}, \) defined in Proposition 1. It follows from the definition of \(P(Z^{(l)} = z^{(l)})\) in (5) that

\[
P(Z^{(1)} = z^{(1)}, Z^{(2)} = z^{(2)}) = P(Z^{(1)} = z^{(1)})P(Z^{(2)} = z^{(2)})
\]

and only if \(C = 1_{K^{(1)}} 1_{K^{(2)}}\). Thus, testing the null hypothesis of independence between the latent community memberships \(Z^{(1)}\) and \(Z^{(2)}\) amounts to testing \(H_0 : C = 1_{K^{(1)}} 1_{K^{(2)}}\).

### 4.1 The \(P^2\)LRT statistic

To test \(H_0 : C = 1_{K^{(1)}} 1_{K^{(2)}}\), one might consider using a likelihood ratio test. The likelihood ratio test statistic is of the form

\[
\max_{\theta^{(1)}, \theta^{(2)}, \pi^{(1)}, \pi^{(2)}, C} \ell(\eta^{(1)}, \eta^{(2)}, \pi^{(1)}, \pi^{(2)}, C; X^{(1)}, X^{(2)})
\]

\[
- \max_{\eta^{(1)}, \eta^{(2)}, \pi^{(1)}, \pi^{(2)}} \ell(\theta^{(1)}, \theta^{(2)}, \pi^{(1)}, \pi^{(2)}, 1_{K^{(1)}} 1_{K^{(2)}}, X^{(1)}, X^{(2)}),
\]

where the log-likelihood function \(\ell\) is defined in (7). Unfortunately, recall from Section 3.2 that (7) is computationally intractable because it involves summing over \((K^{(1)} K^{(2)})^n\) terms. We could replace the log-likelihood functions \(\ell\) with log-pseudolikelihood functions \(\ell_{PL}\), defined in (10). This leads to a test statistic of the form

\[
\log \Lambda \equiv \max_{\eta^{(1)}, \eta^{(2)}, \pi^{(1)}, \pi^{(2)}, C} \ell_{PL}(\eta^{(1)}, \eta^{(2)}, \pi^{(1)}, \pi^{(2)}, C; \hat{b}^{(1)}, \hat{b}^{(2)} | d^{(1)}, d^{(2)})
\]

\[
- \max_{\eta^{(1)}, \eta^{(2)}, \pi^{(1)}, \pi^{(2)}, \pi^{(1)}, \pi^{(2)}, 1_{K^{(1)}} 1_{K^{(2)}}, \hat{b}^{(1)}, \hat{b}^{(2)} | d^{(1)}, d^{(2)}},
\]

(11)

However, \(\ell_{PL}\) is a nonconcave function of its arguments, and therefore, no algorithms are available to exactly
compute the two terms in (11)—they can at best be approximated via local maxima. Taking the difference between two local maxima can lead to undesirable behavior; for example, log Λ can be negative.

To overcome this problem, we take a different approach, motivated by the fact that each data view X^{(i)} marginally follows an SBM (Section 3.1). Rather than estimating the parameters η^{(1)}, η^{(2)}, π^{(1)}, π^{(2)}, and C by maximizing the log-pseudolikelihood function for the multiview SBM (10), we first estimate η^{(1)}, π^{(1)} and η^{(2)}, π^{(2)} by maximizing the log-pseudolikelihood function for the SBM (4) for each view separately. As (4) can be viewed as the log-likelihood function of an FMM (Section 3.2), it can be maximized using the expectation-maximization (EM; Dempster et al., 1977) algorithm for fitting FMMs (McLachlan and Krishnan, 2007). We then plug these estimates into (11), yielding the test statistic

$$\log \hat{\Lambda} \equiv \max_{C \in C_{\rho(1),\rho(2)}} \ell_{PL}(\hat{\eta}^{(1)}, \hat{\eta}^{(2)}, \hat{\pi}^{(1)}, \hat{\pi}^{(2)}, C; \hat{B}^{(1)}, \hat{B}^{(2)} \mid d^{(1)}, d^{(2)})$$

$$- \ell_{PL}(\hat{\eta}^{(1)}, \hat{\eta}^{(2)}, \hat{\pi}^{(1)}, \hat{\pi}^{(2)}, C; \hat{B}^{(1)}, \hat{B}^{(2)} \mid d^{(1)}, d^{(2)}).$$

(12)

Computing (12) requires maximizing the first term with respect to C, that is, to compute

$$\hat{C} \equiv \arg \max_{C \in C_{\rho(1),\rho(2)}} \ell_{PL}(\hat{\eta}^{(1)}, \hat{\eta}^{(2)}, \hat{\pi}^{(1)}, \hat{\pi}^{(2)}, C; \hat{B}^{(1)}, \hat{B}^{(2)} \mid d^{(1)}, d^{(2)}),$$

(13)

where C_{\rho} is defined in Proposition 1. Because the objective of (13) is a concave function of C, \hat{C} can be obtained using techniques from convex optimization. (In particular, we use an exponentiated gradient descent algorithm (Kivinen and Warmuth, 1997) developed in Gao et al. (2020) for maximizing concave functions of C under the constraint that C \in \rho(1,\rho(2)); the complexity of each iteration is \mathcal{O}(nk\rho(1)\rho(2)).) This means that (12) completely overcomes the challenges associated with the test statistic (11); for example, (12) cannot be negative. Furthermore, results from Liang and Self (1996) and Chen and Liang (2010) suggest that performing a partial maximization over the parameters (as in (12)) rather than a full maximization (as in (11)) does not lead to an appreciable loss in power when n is large.

We refer to log \hat{\Lambda} in (12) as a pseudo-pseudo-likelihood ratio test (P^2LRT) statistic. In the name P^2LRT, the term “pseudo” is used in two different senses: the first is because we use the pseudolikelihood function \ell_{PL} in place of the likelihood function, and the second is because we do not perform a full joint maximization over (\eta^{(1)}, \eta^{(2)}, \pi^{(1)}, \pi^{(2)}, C).

**Algorithm 1** Computing the P^2LRT statistic \log \hat{\Lambda} defined in (12).

1. For l = 1, 2, 
   i. Compute d^{(l)} = X^{(l)\rho(l)}. Apply spectral clustering with perturbations (Amini et al., 2013) to X^{(l)} to obtain \hat{Z}^{(l)}, and compute \hat{b}^{(l)} according to \hat{b}^{(l)}_{xy} = \sum_{j=1}^{n} x^{(l)j} t(X^{(l)j} = m).
   ii. Maximize \ell_{PL}(\eta^{(1)}, \eta^{(2)}, \hat{b}^{(1)} \mid d^{(1)}), where \ell_{PL} is defined in (4), and denote the maximizers by \hat{\eta}^{(1)} and \hat{\eta}^{(2)}. This can be done using the EM algorithm for fitting FMMs (McLachlan and Krishnan, 2007).

2. Compute \hat{C} according to (13):
   i. Define matrices \hat{g}^{(1)} \in \mathbb{R}^{nK(1)} and \hat{g}^{(2)} \in \mathbb{R}^{nK(2)} with elements \hat{g}^{(1)}_{ik} = g(\hat{\eta}^{(1)}; d^{(1)}, \hat{\pi}^{(1)}) and \hat{g}^{(2)}_{ik} = g(\hat{\eta}^{(2)}; d^{(2)}, \hat{\pi}^{(2)}).
   ii. Fix a step size s > 0, and let \hat{C}^{(1)} = 1_{K(1)}1_{K(2)}^T. For t = 1, 2, ..., until convergence:
      a. Define \hat{O}_{kk'} = \hat{C}^{(t)}_{kk'} \exp[s G_{kk'} - 1], where
         \[ G_{kk'} = \sum_{i=1}^{n} \hat{g}^{(1)}_{ki} \hat{g}^{(2)}_{k'i}. \]
      b. Let u^{(1)} = 1_{K(1)} and v^{(1)} = 1_{K(2)}. For t' = 1, 2, ..., until convergence:
         \[ u^{(t')} = \frac{1_{K(1)}}{\hat{O}^{(t')} \hat{\pi}^{(1)}}, \quad v^{(t')} = \frac{1_{K(2)}}{\hat{O}^{(t')} \hat{\pi}^{(2)}}, \]
         where the fractions denote element-wise vector division.
      c. Let u and v be the vectors to which u^{(t')} and v^{(t')} converge. Let \hat{C}^{(t+1)} = u \hat{O}_{kk'} v.\]
   iii. Let \hat{C} denote the matrix to which \hat{C}^{(t)} converges.
3. Compute log \hat{\Lambda} according to (12), where \ell_{PL} is defined in (10).

We summarize the procedure for computing the P^2LRT statistic in Algorithm 1.

### 4.2 Approximating the null distribution

Under the null hypothesis that the community memberships Z^{(1)} and Z^{(2)} are independent, that is, under H_0 : C = 1_{K(1)1^T_{K(2)}}, we can write the joint density of X^{(1)} and X^{(2)} as

\[
f(X^{(1)}, X^{(2)}) = E_{Z^{(1)}, Z^{(2)}}[f(X^{(1)} \mid Z^{(1)}, Z^{(2)})]
= E_{Z^{(1)}, Z^{(2)}}[f(X^{(1)} \mid Z^{(1)}) f(X^{(2)} \mid Z^{(2)})]
= E_{Z^{(1)}}[f(X^{(1)} \mid Z^{(1)})] E_{Z^{(2)}}[f(X^{(2)} \mid Z^{(2)})]
= f(X^{(1)}) f(X^{(2)}),
\]

where the second equality follows from the fact that X^{(1)} \perp X^{(2)} \mid Z^{(1)}, Z^{(2)} and X^{(1)} \perp X^{(2)} \mid Z^{(1)} and X^{(2)} \perp Z^{(2)} \mid Z^{(1)} (Section 3.1). Thus, under H_0 : C = 1_{K(1)1^T_{K(2)}}, the joint distribution of X^{(1)} and X^{(2)} is
Algorithm 2 $P^2$LRT for testing $H_0 : C = 1_{K_{(1)}}1_{K_{(2)}}^T$

1. Apply Algorithm 1 to compute $\hat{b}^{(1)}, \hat{b}^{(2)}, d^{(1)}, d^{(2)}$, and the $P^2$LRT statistic $\log \Lambda$ in (12).

2. For $m = 1, \ldots, M$, where $M$ is the number of random permutations:
   
   i. Apply the same permutation to the columns of $\hat{b}^{(2)}$ and the elements of $d^{(2)}$ to compute $\hat{b}^{(2,m)}$ and $d^{(2,m)}$.
   
   ii. Replace $\hat{b}^{(2)}, d^{(2)}$ with $\hat{b}^{(2,m)}, d^{(2,m)}$ in Step 2 of Algorithm 1 to compute $\tilde{C}^{(m)}$.
   
   iii. Replace $\tilde{b}^{(2)}, d^{(2)}, \tilde{C}$ with $\tilde{b}^{(2,m)}, d^{(2,m)}, \tilde{C}^{(m)}$ in (12) to compute $\log \tilde{\Lambda}^{(m)}$.

3. The $p$-value for testing $H_0 : C = 1_{K_{(1)}}1_{K_{(2)}}^T$ is given by

   \[
   \sum_{m=1}^{M} \frac{1}{M+1} \left[ \frac{1}{M+1} \right] \log \Lambda (\log \Lambda^{(m)})+1.
   \]

The $p$-value is never exactly zero (Belinda and Smyth, 2010).

When we reject $H_0 : C = 1_{K_{(1)}}1_{K_{(2)}}^T$, it is often of interest to investigate the strength and location of the dependence between views. Recall from Section 3.1 that $C_{kk'}$ measures the dependence between the $k$th community in the first view and the $k'$th community in the second view. Thus, we can gain insight into the strength and location of the dependence between the communities in the two data views by examining $C_{kk'}$ defined in (13).

5 | EXTENSION TO A NETWORK VIEW AND A MULTIVARIATE VIEW

In this section, we develop a test of association between latent communities in a network view and latent clusters in a multivariate view.

5.1 | Model and notation

We now propose an extension of the SBM to an undirected network view, $X \in \{0, 1\}^{n \times p}$, and a multivariate view, $y \in \mathbb{R}^{n \times p}$. We assume that the network is undirected with no self-loops, so that $X$ is symmetric and $X_{ii} = 0$ for $i = 1, 2, \ldots, n$. We model $X$ with an SBM (Section 2.2) with $K^{(1)}$ communities and we model the rows of $Y$ with an FMM (McLachlan and Peel, 2000) with $K^{(2)}$ clusters, so that

\[
\begin{align*}
\log f(X | Z^{(1)}) &= \sum_{j=1}^{n} \sum_{l=1}^{n} \left( \gamma_{z_{l}^{(1)}z_{j}^{(1)}} X_{ij} (1 - \gamma_{z_{l}^{(1)}z_{j}^{(1)}}) \right)^{1-X_{ij}}, \\
\log f(Y | Z^{(2)}) &= \sum_{l=1}^{n} \phi(Y_{il} | \gamma_{z_{i}^{(2)}},) \tag{14}
\end{align*}
\]

where $\phi(\cdot | \gamma)$ is a density parameterized by $\gamma$, and for $l = 1, 2$, the latent random vector $Z^{(l)} = (Z_{i}^{(l)} = Z_{1}^{(l)}, \ldots, Z_{n}^{(l)})$ has i.i.d. elements with $P(Z_{i}^{(l)} = k) = \pi_{k}^{(l)}$ for $\pi^{(l)} \in \Delta_{+}^{K^{(l)}}$. Here, $Z^{(1)}$ represents the latent community memberships in the network view, and $Z^{(2)}$ represents the latent cluster memberships in the multivariate view. We assume that the $n$ pairs $\{(Z_{i}^{(1)}, Z_{i}^{(2)})\}_{i=1}^{n}$ are i.i.d., and that $X \perp Y | Z^{(1)}, Z^{(2)}$. Thus, as in Section 3.1, it follows from Proposition 1 that there exists $C \in C_{\pi^{(1)}, \pi^{(2)}}$ such that

\[
\begin{align*}
P(Z^{(1)} = z^{(1)}, Z^{(2)} = z^{(2)}) &= \prod_{l=1}^{n} \pi_{z_{i}^{(1)}}(1) \pi_{z_{i}^{(2)}}(1) C_{z^{(1)}z^{(2)}}, \tag{15}
\end{align*}
\]

where $C_{kk'}$ describes the dependence between the $k$th community in the network view and the $k'$th cluster in the multivariate view.

5.2 | Approximate pseudolikelihood function

The multiview log-likelihood function of model (14)–(15) is computationally intractable. Thus, we will derive a multiview log-pseudolikelihood function for model (14)–(15).

We begin by approximating the conditional density of $\hat{b}$ and $Y$ given $d$, where $\hat{b}$ contains the number of edges connecting each of the $n$ nodes in the network to each of the $K$ estimated communities in the network, and $d$ contains the node degrees:

\[
\begin{align*}
\hat{b}, Y | Z^{(1)}, Z^{(2)}, d &\sim \prod_{i=1}^{n} g(\hat{b}_{i}, d_{i}, \eta_{z_{i}^{(1)}}, \phi(Y_{il} | \gamma_{z_{i}^{(2)}},)) \tag{16}
\end{align*}
\]

The derivation of (16) is very similar to the derivation of (9) in Section 3.2. Ignoring any dependence between $d$ and $(Z^{(1)}, Z^{(2)})$, and marginalizing over $Z^{(1)}$ and $Z^{(2)}$ in (16) to approximate the conditional distribution of $\hat{b}$ and $Y$ given $d$, yields

\[
\ell_{PL}(\eta, \gamma, \pi^{(1)}, \pi^{(2)}, C; \hat{b}, Y | d)
\]
\[ \sum_{i=1}^{n} \log \left( \sum_{k,k'} \pi^{(1)}_{k} \pi^{(2)}_{k'} C_{kk'} g(\hat{\beta}_i; d, \eta_k) \phi(Y_i; \gamma_{k'}) \right). \]  

We observe that the log-pseudolikelihood function in (17) closely resembles (10).

### 5.3 Testing independence between \( Z^{(1)} \) and \( Z^{(2)} \)

We now propose a test for the null hypothesis that the latent community memberships \( Z^{(1)} \) and the latent cluster memberships \( Z^{(2)} \) in model (14)--(15) are independent. As in Section 4, this amounts to testing \( H_0 : C = 1_{K^{(1)}} 1_{K^{(2)}}^T \).

Recall that the network \( X \) marginally follows an SBM, and let \( \hat{\eta} \) and \( \hat{\pi} \) be the maximizers of \( \ell_{PL}(\eta, \pi^{(1)}; \hat{\beta} | d) \), where \( \ell_{PL} \) is the log-pseudolikelihood function for the SBM given by (4). As in Section 4.1, we can compute \( \hat{\eta} \) and \( \hat{\pi}^{(1)} \) by using the EM algorithm for fitting FMMs (McLachlan and Krishnan, 2007). Recall that the rows of the multivariate view \( Y \) marginally follow a FMM, and let \( \hat{\gamma} \) and \( \hat{\pi}^{(2)} \) be the maximizers of the log-likelihood function for the multivariate view, obtained via EM. We consider the \( P^2 \)LRT statistic given by

\[
\log \hat{\Lambda} = \arg \max_{C \in C_{g(1)} g(2)} \ell_{PL}(\hat{\eta}, \hat{\gamma}, \hat{\pi}^{(1)}, \hat{\pi}^{(2)}, C; \hat{\beta}, Y | d) \\
- \ell_{PL}(\hat{\eta}, \hat{\gamma}, \hat{\pi}^{(1)}, \hat{\pi}^{(2)}, 1_{K^{(1)}} 1_{K^{(2)}}^T; \hat{\beta}, Y | d),
\]

where \( \ell_{PL} \) is the log-pseudolikelihood function in (17), and \( C_{g} \) is defined in Proposition 1. Once again, we can perform the maximization over \( C \) using techniques from convex optimization; details of the exponentiated gradient descent algorithm that we use are similar to Step 2 of Algorithm 1. As in Section 4.2, we approximate the null distribution of \( \log \Lambda \) by taking \( M \) random permutations of the rows of \( X^{(2)} \), and comparing the observed value of \( \log \Lambda \) to its empirical distribution in the permuted data. Details are similar to Algorithm 2.

### 6 RELATED LITERATURE

Many papers have extended the SBM to the multiple network data view setting, under the assumption that a single set of communities is shared across all networks (Han et al., 2015; Peixoto, 2015; Paul and Chen, 2016) or a subset of networks (Stanley et al., 2016). The model proposed in Section 3.1 does not rely on this assumption. Most of the previous work that avoids the assumption of shared communities has focused on estimation of the community structure; Section 4 of Kim et al. (2018) reviews these papers in detail. By contrast, the primary goal of our paper is not estimation, but rather to develop a test of association between the communities underlying each network view (Section 4).

A related problem in functional neuroimaging is to test whether the communities underlying brain networks of two groups of healthy and diagnosed patients are the same; see Paul et al. (2020), and the references contained therein. However, the test statistics and/or \( p \)-values for these tests cannot be computed in the two network data view setting.

We proposed a test of the null hypothesis that the communities underlying two network views are independent. By contrast, Xiong et al. (2019) proposed a test of the null hypothesis that the networks are conditionally independent given their underlying communities.

In the case of a network view and a multivariate view, several papers have assumed that the communities underlying the network view and the clusters underlying the multivariate view are the same, and exploit this assumption to improve parameter estimation (Binkiewicz et al., 2017; Stanley et al., 2019; Yan and Sarkar, 2020). Our proposed model in Section 5.1 does not rely on this assumption. Another body of work estimates the relationship between community memberships and node covariates, but does not consider inference on this relationship (Yang et al., 2013; Newman and Clauset, 2016; Zhang et al., 2016).

In Section 5, we proposed testing for a specific type of relationship between the network view and the multivariate view: we test for association between the communities underlying the network view and the clusters underlying the multivariate view. Several papers have considered testing for other types of relationships between the network view and the multivariate view (Traud et al., 2011; Fosdick and Hoff, 2015; Peel et al., 2017). For example, Peel et al. (2017) test for association between the network view and a categorical node covariate.

### 7 SIMULATION RESULTS

In this section, we evaluate the power and Type I error of the tests proposed in Sections 4–5. Simulations in this paper were conducted using the simulator package (Bien, 2016).

#### 7.1 SBM for two network data views

We will evaluate the performance of four tests of \( H_0 : C = 1_{K^{(1)}} 1_{K^{(2)}}^T \).
FIGURE 2  Power of the $P^2$ LRT and the $G$-test with both views drawn from an SBM, as we vary the dependence between views ($\Delta$), the strength of the communities ($r$), the expected edge density ($s$), and how the number of communities is selected. Details are in Section 7.1. This figure appears in color in the electronic version of this article, and any mention of color refers to that version.

1. the $P^2$ LRT proposed in Section 4, using the true values of $K^{(1)}$ and $K^{(2)}$,
2. the $P^2$ LRT proposed in Section 4, using estimated values of $K^{(1)}$ and $K^{(2)}$,
3. the $G$-test for testing dependence between two categorical variables (Agresti, 2003, Chapter 3.2) applied to the estimated community memberships for each view, using the true values of $K^{(1)}$ and $K^{(2)}$, and
4. the $G$-test, using estimated values of $K^{(1)}$ and $K^{(2)}$.

We estimate $K^{(1)}$ and $K^{(2)}$ by applying the method of Le and Levina (2015) to $X^{(1)}$ and $X^{(2)}$, respectively. In all four tests, we approximate the null distribution with a permutation approach, as in Algorithm 2, using $M = 200$ permutation samples.

We generate data from model (5)–(6), with $n = 1000$, $K^{(1)} = K^{(2)} = K = 6$, and

$$C = (1 - \Delta)1_K1_K^T + \Delta \cdot \text{diag}(K1_K),$$

(18)

for $\Delta \in [0, 1]$. Here, $\Delta = 0$ corresponds to independent communities and $\Delta = 1$ corresponds to identical communities. We let $\pi^{(1)} = \pi^{(2)} = 1_K / K$, and $\theta^{(1)} = \theta^{(2)} = \theta$, with

$$\theta_{kk'} = \omega(1\{k \neq k'\} + 2r1\{k = k'\}),$$

(19)

for $r > 0$ describing the strength of the communities, and $\omega$ chosen so that the expected edge density of the network equals $s$, to be specified. We simulate 2000 data sets for a range of values of $s$, $\Delta$, and $r$, and evaluate the power of the four tests described above. Results are shown in Figure 2.

For all tests, power tends to increase as $\Delta$, which controls the dependence between views, increases. Power also tends to increase as the strength of the communities ($r$) increases, and as the expected edge density ($s$) increases. Estimating $K^{(1)}$ and $K^{(2)}$ tends to yield lower power than using the true values of $K^{(1)}$ and $K^{(2)}$. All tests control the Type I error, but the $P^2$ LRTs uniformly yield higher power than the $G$-tests. This is because the $P^2$ LRT can be interpreted as a version of the $G$-test that replaces the “hard” community assignments with “soft” community assignments (Gao et al., 2020, Section 5). Thus, the $P^2$ LRT outperforms the $G$-test when the communities are more difficult to detect.

We generate data with unbalanced community sizes in Web Appendix D, and investigate how the true values of $K^{(1)}$ and $K^{(2)}$ relate to power in Web Appendix E.

7.2 Degree-corrected SBM for two network data views

Under the SBM, nodes within the same community have the same expected degree. To investigate the performance of the test proposed in Section 4 in a setting where nodes can have different expected degrees, we generate each network view from the degree-corrected SBM (DCSBM, Karrer and Newman, 2011). We generate $n$ vec-
tors \((Z_i^{(1)}, Z_i^{(2)}, \delta^{(1)}, \delta^{(2)})\) i.i.d. for \(i = 1, 2, \ldots, n\), with \(Z_i^{(1)}\) and \(Z_i^{(2)}\) categorical with \(K^{(1)}\) and \(K^{(2)}\) levels, respectively, and \((Z_i^{(1)}, Z_i^{(2)}) \perp (\delta^{(1)}, \delta^{(2)})\). Here, \(\delta^{(1)}\) and \(\delta^{(2)}\) represent *popularities* for the nodes in the two views; more popular nodes have higher expected degrees. We generate each view with

\[
X^{(l)} \mid Z^{(l)}, \delta^{(l)} \sim \prod_{j=1}^{n} \prod_{i=1}^{j-1} \left( \delta^{(l)}_{i,j} \frac{1}{Z_i Z_j} \right) X^{(l)}_{i,j} \left( 1 - \delta^{(l)}_{i,j} \frac{1}{Z_i Z_j} \right)^{1-X^{(l)}_{i,j}}, \quad l = 1, 2. \tag{20}
\]

We set \(n, K^{(1)}, K^{(2)}, \pi^{(1)}, \pi^{(2)}, C, \phi^{(1)},\) and \(\phi^{(2)}\) as in Section 7.1 and take \(P(\delta^{(1)} = 2.5) = 0.2, P(\delta^{(2)} = 0.625) = 0.8,\) and \(\delta^{(1)} \perp \delta^{(2)}\). We simulate 2000 data sets, varying the dependence between views \((\Delta)\), the expected edge density \((s)\), and how the number of communities is selected. Details are in Section 7.2. This figure appears in color in the electronic version of this article.

In this subsection, we assumed that the node popularities \((\delta^{(1)}\) and \(\delta^{(2)}\)) are independent. This can sometimes be an unrealistic assumption in practice. If \(\delta^{(1)}\) and \(\delta^{(2)}\) are dependent, then \(X^{(1)}\) and \(X^{(2)}\) could be dependent even when the communities are independent, which could inflate the Type I error rate. To investigate this effect, in Web Appendix B.1, we generate data from a multiview DCSBM with \(\delta^{(1)}\) and \(\delta^{(2)}\) dependent, and apply the \(P^2\) LRT using a range of values of \(K^{(1)}\) and \(K^{(2)}\). We find that the Type I error rate is controlled, both when we estimate the number of communities and when we choose a fixed number of communities (as long as the number of communities is not grossly overspecified); Web Appendix B.2 gives intuition for why this is the case.

7.3 | SBM for a network view and a multivariate view

We will evaluate the performance of six tests of \(H_0 : C = 1_{K^{(1)}}1_{K^{(2)}T}\):

1. the \(P^2\) LRT proposed in Section 5, using the true values of \(K^{(1)}\) and \(K^{(2)}\),
2. the \(P^2\) LRT, using estimated values of \(K^{(1)}\) and \(K^{(2)}\),
3. the \(G\) test applied to the estimated community/cluster memberships in the network/multivariate view, using the true values of \(K^{(1)}\) and \(K^{(2)}\),
4. the \(G\) test applied to the estimated community/cluster memberships in the network/multivariate view, using the estimated values of \(K^{(1)}\) and \(K^{(2)}\),
5. the \(G\) test applied to the estimated community/cluster memberships in the network/multivariate view, using the true values of \(K^{(1)}\) and \(K^{(2)}\),
6. the \(G\) test applied to the estimated community/cluster memberships in the network/multivariate view, using the estimated values of \(K^{(1)}\) and \(K^{(2)}\).
4. the G-test, using estimated values of \( K^{(1)} \) and \( K^{(2)} \).
5. the BESTest (Peel et al., 2017) applied to the network view and the estimated cluster memberships in the multivariate view, using the true values of \( K^{(1)} \) and \( K^{(2)} \).
6. the BESTest, using estimated values of \( K^{(1)} \) and \( K^{(2)} \).

We estimate \( K^{(1)} \) by applying the method of Le and Levina (2015), we estimate \( K^{(2)} \) using BIC, and we approximate the null distributions using \( M = 200 \) permutation samples.

We generate data from model (14)–(15); we generate data from a degree-corrected version of model (14)–(15) in Web Appendix C. We set \( n = 500 \), and \( K^{(1)} = K^{(2)} = K = 3 \). Let \( \pi^{(1)} = \pi^{(2)} = 1_k / K \), and let \( C \) be given by (18). Let \( \theta \) be given by (19), so that the expected edge density is \( s = 0.015 \). We draw the multivariate data view from a Gaussian mixture model, for which the \( k \)th mixture component is an \( N_{10}(\mu_k, \sigma^2 I_{10}) \) distribution. The \( p \times K \) mean matrix for the multivariate data view is given by

\[
\mu = \begin{bmatrix}
0 \cdot 1_5 & 0 \cdot 1_5 & \sqrt{12} \cdot 1_5 \\
2 \cdot 1_5 & 2 \cdot 1_5 & 0 \cdot 1_5
\end{bmatrix}.
\]

We simulate 2000 data sets for a range of values of \( \Delta, r \), and \( \sigma \). Results are shown in Figure 4.

All tests control the Type I error rate. Power tends to increase as the dependence between views (\( \Delta \)) increases. Power also tends to increase as the strength of the communities (\( r \)) increases and the variance of the clusters (\( \sigma \)) decreases. The \( P^2 \)LRTs uniformly yield higher power than the G-tests and the BESTests.

8 | APPLICATION TO PROTEIN–PROTEIN INTERACTION DATA

In this section, we focus on two types of protein–protein interaction data. A binary interaction is a physical interaction between proteins, and a cocomplex association is a pair of proteins that are part of the same complex. These two data views represent distinct biological concepts; physical interactions can occur between a pair of proteins that are not in the same complex, and not all proteins in complexes physically interact.

To investigate whether the latent communities of proteins defined with respect to binary interactions and cocomplex associations are related, we consider Homo sapiens protein–protein interaction data from the HINT (High-quality INteractomes; Das and Yu (2012b)) database, and ask: are the communities within the binary network and the communities within the cocomplex network associated?

We remove self-interactions from both networks, and consider only those proteins that appear in both networks. This yields 43,874 binary interactions and 88,960 cocomplex associations among a common set of \( n = 9037 \) pro-
teins. We apply the $P^2$ LRT of $H_0 : C = 1_{K(1)}1_{K(2)}^T$ developed in Section 4, using $M = 10^4$ in Step 3 of Algorithm 2. As in Section 7, we estimate the number of communities in each view by applying the method of Le and Levina (2015) to each view separately, which (coincidentally) estimates 14 communities in both data views. Figure 5 displays $\hat{\pi}^{(1)}$ and $\hat{\pi}^{(2)}$ (defined in Section 4.1), and $\hat{C}$ (defined in Equation (13)). Our test yields a $p$-value of 0.013, and thus provides some evidence against the null hypothesis that communities of proteins defined with respect to binary interactions and communities of proteins defined with respect to cocomplex associations are independent.

Our test of $H_0 : C = 1_{K(1)}1_{K(2)}^T$ allows us to provide an answer to the high-level scientific question of whether there is a relationship between communities defined with respect to different types of protein interactions. However, it may also be of scientific interest to determine whether there is a relationship between the $k$th community in the binary view and the $k'$th community in the cocomplex view. Recall from Section 3.1 that $C_{kk'} = 1$ indicates that the $k$th community in the binary view and the $k'$th community in the cocomplex view are independent. In Figure 5, most values of $\hat{C}_{kk'}$ are close to 1. Thus, it may be of future interest to develop tests of $H_0 : C_{kk'} = 1$.

9 | DISCUSSION

In this paper, we considered testing whether communities defined with respect to two networks on a common set of nodes are related. We extended this test to the setting of one network and one multivariate data set on a common set of nodes. The proposed tests control the Type I error rate, and yield higher power than applying the $G$-test to the estimated community/cluster memberships in each data view.

We focused on testing the association between communities/clusters in two data views. If three or more data views are available, we may be interested in testing mutual independence between all data views. The models proposed in Sections 3.1 and 5.1 extend readily to $L > 2$ data views, and we can test for mutual independence by testing the null hypothesis that all entries of an $L$th-order tensor $C$ are equal to 1. We can construct a $P^2$ LRT statistic along the lines of (12), and we can approximate the null distribution by permuting the node labels in the second through $L$th views. If we are instead interested in pairwise independence between the data views, we could simply apply the tests developed in this paper to each pair of views.

In this paper, we considered only undirected, unweighted network views. There is a body of work that extends the single-view SBM to directed and/or weighted networks; see, for example, Wang and Wong (1987) and Aicher et al. (2014). It may be of future interest to extend the methodology developed in this paper to allow for directed and/or weighted networks.

ACKNOWLEDGMENTS

Lucy L. Gao received funding from the Natural Sciences and Engineering Research Council of Canada. Daniela Witten and Jacob Bien were supported by NIH Grant R01GM123993. Jacob Bien was supported by NSF CAREER Award DMS-1653017. Daniela Witten was supported by NIH Grant DP5OD009145, NSF CAREER Award DMS-1252624, and Simons Investigator Award.
No. 560585. We thank Haiyuan Yu for useful input on protein interaction data. Conflict of Interest: None declared.

DATA AVAILABILITY STATEMENT
The data that support the findings of this paper are openly available in the HINT (High-quality INTeractions) database at http://hint.yulab.org (Das and Yu, 2012a).

ORCID
Lucy L. Gao https://orcid.org/0000-0002-6811-0746

REFERENCES
Agresti, A. (2003) Categorical Data Analysis, Vol. 482. Hoboken: John Wiley & Sons.
Aicher, C., Jacobs, A.Z. and Clauset, A. (2014) Learning latent block structure in weighted networks. Journal of Complex Networks, 3, 221–248.
Amini, A.A., Chen, A., Bickel, P.J. and Levina, E. (2013) Pseudo-likelihood methods for community detection in large sparse networks. The Annals of Statistics, 3, 2097–2122.
Belinda, P. and Smyth, G.K. (2010) Permutation p-values should never be zero: calculating exact p-values when permutations are randomly drawn. Statistical Applications in Genetics and Molecular Biology, 3, 1–16.
Besag, J. (1975) Statistical analysis of non-lattice data. Journal of the Royal Statistical Society: Series D (The Statistician), 3, 179–195.
Bien, J. (2016) The simulator: an engine to streamline simulations. arXiv preprint arXiv:1607.00021.
Binkiewicz, N., Vogelstein, J.T. and Rohe, K. (2017) Covariate-assisted spectral clustering. Biometrika, 3, 361–377.
Chen, Y. and Liang, K.-Y. (2010) On the asymptotic behaviour of the pseudolikelihood ratio test statistic with boundary problems. Biometrika, 3, 603–620.
D’Angelo, S., Murphy, T.B. and Alfò, M. (2019) Latent space modelling of multidimensional networks with application to the exchange of votes in Eurovision song contest. The Annals of Applied Statistics, 3, 900–930.
Das, J. and Yu, H. (2012a) High-quality interactomes (HINT). http://hint.yulab.org. Accessed 22 January 2019.
Das, J. and Yu, H. (2012b) HINT: high-quality interactomes and their applications in understanding human disease. BMC Systems Biology, 3, 92.
Dempster, A.P., Laird, N.M. and Rubin, D.B. (1977) Maximum likelihood from incomplete data via the EM algorithm. Journal of the Royal Statistical Society: Series B (Methodological), 3, 1–22.
Erdoğ, P. and Rényi, A. (1960) On the evolution of random graphs. Proceedings of the Hungarian Academy of Sciences, 5, 17–61.
Fosdick, B.K. and Hoff, P.D. (2015) Testing and modeling dependencies between a network and nodal attributes. Journal of the American Statistical Association, 3, 1047–1056.
Gao, L.L., Bien, J. and Witten, D. (2020) Are clusterings of multiple data views independent? Biostatistics, 3, 692–708.
Gollin, I. and Murphy, T.B. (2016) Joint modeling of multiple network views. Journal of Computational and Graphical Statistics, 3, 246–265.
Han, Q., Xu, K. and Airoldi, E. (2015) Consistent estimation of dynamic and multi-layer block models. In: Proceedings of the 32nd International Conference on Machine Learning - Volume 37, pp. 1511–1520.
Hoff, P.D., Raftery, A.E. and Handcock, M.S. (2002) Latent space approaches to social network analysis. Journal of the American Statistical Association, 3, 1090–1098.
Holland, P.W., Laskey, K.B. and Leinhardt, S. (1983) Stochastic blockmodels: first steps. Social Networks, 3, 109–137.
Holland, P.W. and Leinhardt, S. (1981) An exponential family of probability distributions for directed graphs. Journal of the American Statistical Association, 3, 33–50.
Karrer, B. and Newman, M.E. (2011) Stochastic blockmodels and community structure in networks. Physical Review E, 3, 016107.
Kim, B., Lee, K.H., Xue, L. and Niu, X. (2018) A review of dynamic network models with latent variables. Statistics Surveys, 3, 105–135.
Kivinen, J. and Warmuth, M.K. (1997) Exponentiated gradient versus gradient descent for linear predictors. Information and Computation, 3, 1–63.
Le, C.M. and Levina, E. (2015) Estimating the number of communities in networks by spectral methods. arXiv preprint arXiv:1507.00827.
Liang, K.-Y. and Self, S.G. (1996) On the asymptotic behaviour of the pseudolikelihood ratio test statistic. Journal of the Royal Statistical Society. Series B (Methodological), 58, 785–796.
Matias, C. and Robin, S. (2014) Modeling heterogeneity in random graphs through latent space models: a selective review. ESAIM: Proceedings and Surveys, 3, 55–74.
McLachlan, G. and Krishnan, T. (2007). The EM Algorithm and Extensions, Vol. 382. New York, NY: John Wiley & Sons.
McLachlan, G. and Peel, D. (2000). Finite Mixture Models. New York, NY: John Wiley & Sons.
Newman, M.E. and Clauset, A. (2016) Structure and inference in annotated networks. Nature Communications, 3, 11863.
Paul, S. and Chen, Y. (2016) Consistent community detection in multi-relational data through restricted multi-layer stochastic blockmodel. Electronic Journal of Statistics, 3, 3807–3870.
Paul, S. and Chen, Y. (2020) A random effects stochastic block model for joint community detection in multiple networks with applications to neuroimaging. Annals of Applied Statistics, 3, 993–1029.
Peel, L., Larremore, D.B. and Clauset, A. (2017) The ground truth about metadata and community detection in networks. Science Advances, 3, e1602548.
Peixoto, T.P. (2015) Inferring the mesoscale structure of layered, edge-valued, and time-varying networks. Physical Review E, 3, 042807.
Sailer-Townshend, M. and McCormick, T.H. (2017) Latent space models for multiview network data. The Annals of Applied Statistics, 11, 1217–1244.
Stanley, N., Bonacci, T., Kwitt, R., Nietheram, M. and Mucha, P.J. (2019) Stochastic block models with multiple continuous attributes. Applied Network Science, 3, 1–22.
Stanley, N., Shi, S., Taylor, D. and Mucha, P.J. (2016) Clustering network layers with the strata multilayer stochastic block model. IEEE Transactions on Network Science and Engineering, 3, 95–105.
Sun, S. (2013) A survey of multi-view machine learning. Neural Computing and Applications, 3, 2031–2038.
Traud, A.L., Kelsic, E.D., Mucha, P.J. and Porter, M.A. (2011) Comparing community structure to characteristics in online collegiate social networks. SIAM Review, 3, 526–543.
Wang, Y.J. and Wong, G.Y. (1987) Stochastic blockmodels for directed graphs. Journal of the American Statistical Association, 3, 8–19.
Xiong, J., Shen, C., Arroyo, J. and Vogelstein, J.T. (2019) Graph independence testing. *arXiv preprint arXiv:1906.03661.*

Yan, B. and Sarkar, P. (2020) Covariate regularized community detection in sparse graphs. *Journal of the American Statistical Association.* Advance online publication.

Yang, J., McAuley, J. and Leskovec, J. (2013) Community detection in networks with node attributes. In: *Proceedings of the IEEE 13th International Conference on Data Mining,* pp. 1151–1156.

Zhang, Y., Levina, E. and Zhu, J. (2016) Community detection in networks with node features. *Electronic Journal of Statistics,* 3, 3153–3178.

**SUPPORTING INFORMATION**

Web Appendices referenced in Sections 3 and 7 are available with this paper at the *Biometrics* website on Wiley Online Library. The tests developed in this paper are implemented in the R package *multiviewtest,* which is available on CRAN. Code to reproduce the results in this paper is available at https://github.com/lucylgao/mv-network-test-code, and also available with this paper at the *Biometrics* website on Wiley Online Library.

**How to cite this article:** Gao LL, Witten D, Bien J. (2022) Testing for association in multiview network data. *Biometrics,* 78, 1018–1030. https://doi.org/10.1111/biom.13464