Network Science (2025), **13**, e11, 1–19
doi:10.1017/nws.2025.10005

CAMBRIDGE UNIVERSITY PRESS

RESEARCH ARTICLE

Spectral goodness-of-fit tests for complete and partial network data

Shane Lubold[1], Bolun Liu[2] and Tyler McCormick[1]

[1]Department of Statistics, University of Washington, Seattle, WA, USA and [2]Department of Biostatistics, Johns Hopkins University, Baltimore, MD, USA
Corresponding author: Tyler McCormick; Email: tylermc@uw.edu

Abstract

Networks describe complex relationships between individual actors. In this work, we address the question of how to determine whether a parametric model, such as a stochastic block model or latent space model, fits a data set well, and will extrapolate to similar data. We use recent results in random matrix theory to derive a general goodness-of-fit (GoF) test for dyadic data. We show that our method, when applied to a specific model of interest, provides a straightforward, computationally fast way of selecting parameters in a number of commonly used network models. For example, we show how to select the dimension of the latent space in latent space models. Unlike other network GoF methods, our general approach does not require simulating from a candidate parametric model, which can be cumbersome with large graphs, and eliminates the need to choose a particular set of statistics on the graph for comparison. It also allows us to perform GoF tests on partial network data, such as Aggregated Relational Data. We show with simulations that our method performs well in many situations of interest. We analyze several empirically relevant networks and show that our method leads to improved community detection algorithms.

Keywords: random matrix theory; network analysis; latent space model

1. Introduction

Networks consist of connections, also known as edges or ties, between individual actors or nodes. Such data are common in a variety of settings in the social, economic, and health sciences. The simplest model is the Erdös-Rényi model (Erdős & Rényi, 1959), in which edges form independently with the same probability. More complex models have been developed, such as stochastic block models (SBM) and degree-corrected variants (Holland et al., 1983; Airoldi et al., 2006; Rohe et al., 2011; Yan et al., 2014), latent space models (Hoff et al., 2002; Hoff, 2005; Shalizi & Asta, 2017), exponential random graph models (ERGMs) (Holland & Leinhardt, 1981; Hunter et al., 2008), and many more. Among others, Goldenberg et al. (2009) provides a survey of common network models.

Given the multiple available models, a natural question in practice is how to choose a model that is appropriate for a particular dataset. Broadly speaking, there are two common approaches to this problem currently in the literature. A first approach leverages the fact that the problem has a "parametric null." Since many network models are also generative, one common strategy is to estimate parameters of the model in question, then simulate a series of graphs. Statistics from the fitted model should resemble the observed statistics (Ouadah et al., 2019; Chen & Onnela, 2019; Shore & Lubin, 2015; Gao et al., 2019). A potential issue with these methods is that, in many setting, there is limited information available to a practitioner to decide which statistics to use for comparison. In some cases, the researcher can choose statistics that are important to their application, but it might not always be possible to select *a priori* which statistics will be the

© The Author(s), 2025. Published by Cambridge University Press. This is an Open Access article, distributed under the terms of the Creative Commons Attribution-NonCommercial-NoDerivatives licence (https://creativecommons.org/licenses/by-nc-nd/4.0), which permits non-commercial re-use, distribution, and reproduction in any medium, provided that no alterations are made and the original article is properly cited. The written permission of Cambridge University Press must be obtained prior to any commercial use and/or adaptation of the article.

https://doi.org/10.1017/nws.2025.10005 Published online by Cambridge University Press

most important in future analyses. This method also requires taking multiple samples from the generative process, which can be cumbersome in high dimensional settings.

A second common strategy for assessing goodness-of-fit (GoF) involves using the penalized likelihood methods, such as the Bayesian Information Criterion (BIC) or Akaike Information Criterion (AIC). For example, AIC or BIC could be used to select the dimension of the latent space in latent space models, but as we show with simulations in Supplementary Material B, the BIC approach leads to poor dimension estimates. In fact, the manual for one of the most common software packages for fitting parametric models, `latentnet`, states "*It is not clear whether it is appropriate to use this BIC to select the dimension of latent space . . ..*" This issue has also been documented previously (Oh & Raftery, 2001; Handcock et al., 2007; Raftery et al., 2007; Lenk, 2009; Gormley & Murphy, 2010). By contrast, the GoF test we propose here does not use a penalized likelihood to select dimension. Instead, it uses the eigenvalues of a random matrix that measures how well an assumed model fits the data. This procedure, as we show in this work, outperforms BIC and similar metrics when applied to selecting the dimension of the latent space.

In this work, we present a novel GoF test to assess model fit when the network of interest is undirected or directed. Our method also accommodates partial network data, which is a vital part of modern network analysis (Bernard et al., 2010; McCormick et al., 2010; Breza et al., 2020, 2019; Alidaee et al., 2020), but are not easily handed in existing GoF tools. Our goal is to derive a testing framework for the hypotheses

$$H_0{:}G \sim F_\theta, \quad H_a{:}H_0 \text{ is false}, \tag{1}$$

where G is a random network of interest and F_θ is a parametric network model with an (unknown) parameter vector represented by θ. In words, we have an assumed parametric network model, F_θ, and our goal is to test whether G could be drawn from F_θ. Throughout this work, we will use graph and network interchangeably.

A critical aspect in the setup of the tests above is that they assess GoF for the entire model simultaneously. In many settings, this means that the contribution of individual parameters to the fitness measure may not be separately identified. Take, as an example, the latent distance model discussed above. This model has both individual effects for each respondent and latent distances for each pair. A common question, as described above, involves testing for the dimension of the latent space. To test exclusively for the dimension, we would need to either marginalize over or condition on potential values for the additional model parameters (see the discussion in Oh & Raftery, 2001, for example, which makes this point in the related setting of multidimensional scaling). In the latent space model this is particularly challenging since both the latent distances and individual effects impact overall graph properties, such as the density (see, e.g. Lubold et al., 2020 for further discussion). In our approach, we ask a related, but distinct question from the literature that tests for specific model parameters. In the case of the latent space model, for example, our test asks whether a model, overall, could have plausibly generated a given set of data, rather than attempting to identify a single "true" latent dimension. Despite this, in our simulations, we see however that this approach tends to find the true dimension with high probability.

The motivation for our test statistic is taken from a result that has been used before in community detection (Lei, 2016; Bickel & Sarkar, 2015) and two-sample tests for networks (Chen et al., 2020). The result, which goes back to Erdős et al. (2012), states that if A is a $n \times n$ random symmetric matrix with (i) $E(A_{ij}) = 0$ and (ii) $\sum_{j \neq i} \text{Var}(A_{ij}) = 1$ for each i, then $n^{2/3}(\lambda_{\max}(A) - 2)$ converges in distribution to a random variable with a Tracy-Widom distribution, where $\lambda_{\max}(A)$ is the largest eigenvalue of the matrix A. The same argument shows that the smallest eigenvalue of A, which we denote by $\lambda_{\min}(A)$, satisfies a similar central limit theorem. Before continuing, we want to highlight the fact that all eigenvalues can contain signal which we could use to do model selection, they are not equally useful. For example, we show in Supplementary Material Figure C.1 that model selection procedures using the largest and smallest eigenvalues are more powerful

https://doi.org/10.1017/nws.2025.10005 Published online by Cambridge University Press

than procedures based on other eigenvalues. This motivates us to only use the largest and smallest eigenvalues in our model selection procedure.

Leveraging this result, we propose a two-step procedure to test the hypothesis in (1). First, we compute an estimate $\hat{\theta}$ of θ and estimate $\hat{P}_{ij} := P(G_{ij} = 1|\hat{\theta})$, where $G_{ij} = 1$ if person i and person j are observed to be connected in the network, and P_{ij} is the probability of such a connection (as defined by the assumed parametric model). Second, we define the random matrix A by, for $i \neq j$,

$$A_{ij} = \frac{G_{ij} - \hat{P}_{ij}}{\sqrt{(n-1)\hat{P}_{ij}(1-\hat{P}_{ij})}}, \tag{2}$$

and $A_{ii} = 0$. Under H_0, we expect that a reasonable estimator for $\hat{P}_{ij}$ should approximate $P(G_{ij} = 1|\theta)$, so A from (2) should approximately satisfy conditions (i) and (ii). We can then compare the largest and smallest eigenvalues of A against the quantiles of the Tracy-Widom distribution to construct a test of H_0 in (1).

Our paper contributes to the literature on testing GoF for network models in three ways. First, we expand work by Lei (2016), which estimates the number of communities in a stochastic block model, to accommodate a variety of common parametric network models. Second, we develop a test for directed data by introducing a similar central limit theorem for eigenvalues of nonsymmetric matrices from Johnstone (2001) and Chafaï (2009). Third, we show how to test (1) when the researcher only has access to partial network data, such as Aggregated Relational Data (ARD) (see McCormick & Zheng, 2015 for further discussion). Along with asymptotic arguments we also present a bootstrap procedure which improves performance of our hypothesis tests in finite samples.

The paper is structured as follows. First, we review relevant literature in the remainder of this section. Next, we introduce the construction of the Tracy-Widom distribution and asymptotic arguments for undirected, directed, and partial network data in Section 2. In Section 3, we discuss a bootstrap correction algorithm to improve finite sample properties. We then present a series of network models that are compatible with our method in Section 4 along with simulation results. Lastly, in Section 5, we analyze several observed networks using a latent distance model (Hoff et al., 2002; Hoff, 2003, 2005), which assumes that relationships in the network depend on the positions of actors in latent "social space" of low but unknown dimension. Our goal with these data is to test for the minimal latent dimension.

1.1 Literature review

GoF methods for dyadic data generally address the question "Does the proposed model fit my network data well?" One reason this problem is challenging, among others, is that there is often only one network of interest. In other words, we cannot access more draws from the distribution that generated the observed network. Many GoF methods try to use Monte Carlo methods to evaluate auxiliary statistics from networks generated under the fitted model against those in the observed network. The canonical example is the ERGM GoF of Hunter et al. (2008), which contrasts degree, edgewise shared partner, and geodesic distance distributions, variants of which now extends to small-network ERGMs (Vega Yon, Slaughter, and Haye, 2021) and bipartite ERGMs (Stivala et al., 2025).

Such Monte Carlo ideas also underpin the development of asymptotic GoF test statistics beyond ERGMs. For example, Shore & Lubin (2015) proposes a general GoF method based on resampling the graph Laplacian's eigenvalues and constructing confidence intervals based on these values. Gao & Lafferty (2017) looks at using small graph statistics, such as the number of triangles or edges, to determine if there is community structure in a network. Ouadah et al.

https://doi.org/10.1017/nws.2025.10005 Published online by Cambridge University Press

(2019) which derives a test for an Erdös-Rényi network using the degree distribution. Each of these methods, generally speaking, requires a new derivation of a central limit theorem for the network statistic of interest under a suitable null hypothesis. Recent work has moved beyond that asymptotic regime. Xu & Reinert (2021) propose a kernel-Stein GoF test for ERGMs, and Fatima & Reinert (2025) propose a related test for general inhomogeneous random graphs. Karwa et al. (2024), on the other hand, developed a finite-sample, nonasymptotic GoF test for a class of SBMs using algebraic statistics.

Simulation-based GoF has also migrated to longitudinal network models. For stochastic actor-oriented models, Lospinoso & Snijders (2019) combine multiple auxiliary statistics into a Mahalanobis omnibus GoF test, and Amati et al. (2024) further adapt the strategy to relational-event data.

To the best of our knowledge, network GoF tests using partial data are not well-studied. Recent work on modeling partial network data, such as Bernard et al. (2010), Breza et al. (2020), Breza et al. (2019), Alidaee et al. (2020), and Chatterjee (2015), consider model adequacy using out of sample prediction, which may be appropriate in some circumstances but asks a fundamentally different question than gGoF.

Our methods draw on results from random matrix theory and its applications. See, among others, Tao (2012) and references therein, for an introduction, as well as Erdős et al. (2012) and Füredi & Komlós (1981) and their references for recent work on related spectral central limit theorems for eigenvalues. Our method builds on the method presented in Lei (2016) and Bickel & Sarkar (2015), which also use spectral properties to estimate the number of communities in a stochastic block model. Recent work by Wu & Hu (2023) and Chen & Hu (2025) proposes a different approach, using the trace of the normalized adjacency matrix instead.

Lastly, there are several well-known GoF methods fall outside the Monte Carlo and spectral families. Among others, Hu et al. (2021) and Wu et al. (2025) examine maximum entry-wise deviations; Jin et al. (2023) propose a stepwise GoF procedure for community detection in SBMs; Zhang & Amini (2020) propose an adjusted chi-square GoF test for degree-corrected SBMs; and Han et al. (2023) develop a universal rank estimation procedure using on the same normalized adjacency matrix used by Lei (2016) and Bickel and Sarkar (2015).

2. Methodology

We now outline the GoF problem. Let Y be an $n \times n$ matrix containing relationships between actors or nodes, which we label from 1 to n. In this work, we usually only consider Y to be binary-valued, so Y represents a network on n nodes. We suppose that Y is drawn from some distribution F. The network GoF question then asks whether a given set of observed data Y could plausibly have been generated by F. In many cases, it is possible to index F by some parameter θ. We are therefore interested in testing the GoF hypothesis in (1). When $F = F_\theta$, we write $P_{ij} = P(Y_{ij} = 1|\theta)$ to mean the probability that nodes i and j connect, given the parameter θ.

The methodology we present in this work to test (1) requires an estimate of θ. We can estimate θ via maximum likelihood estimation (MLE), for example.

Assuming we have estimated θ with $\hat{\theta}$, we can use the parametric form of F_θ to obtain a fitted distribution $F_{\hat{\theta}}$. From this distribution, we can estimate $\hat{P}_{ij} = P(Y_{ij} = 1|\hat{\theta})$, which is the probability that nodes i and j connect, given the parameter $\hat{\theta}$. In the next section, we derive a testing framework to test the hypothesis in (1). We discuss the cases of undirected, directed, and partially observed networks in their own sections, since each case requires a different approach.

2.1 Undirected networks

We first consider the case where Y corresponds to an undirected binary matrix. The following result, which motivates our test statistic for (1), states that the eigenvalues of a transformation

https://doi.org/10.1017/nws.2025.10005 Published online by Cambridge University Press

of the adjacency matrix satisfy a central limit theorem. See Erdős et al. (2012), Lee & Yin (2014), Füredi & Komlós (1981), Lei (2016), and Wigner (1958) for related results and discussion. Formally, this result combines results from Erdős et al. (2012) and Lee & Yin (2014) and is formulated in Lemma A.1 of Lei (2016), among other works.

Theorem 1 (Lemma A.1 of Lei, 2016). *Let Y be the adjacency matrix of a random graph on n nodes with edges drawn independently with probability P_{ij}. Define the $n \times n$ random matrix A with entries*

$$A_{ij} := \frac{Y_{ij} - P_{ij}}{\sqrt{(n-1)P_{ij}(1-P_{ij})}}, \quad A_{ii} = 0.$$

Then, as $n \to \infty$,

$$t_1 := n^{2/3}\left(\lambda_{\max}(A) - 2\right) \xrightarrow{d} TW_1, \tag{3}$$

$$t_2 := n^{2/3}\left(-\lambda_{\min}(A) - 2\right) \xrightarrow{d} TW_1 \tag{4}$$

where TW_1 is the Tracy-Widom distribution with parameter 1.

In words, this result states the largest and smallest eigenvalues of the matrix A satisfy a central limit theorem. In Figure 1 we plot the TW_1 distribution as well as $n^{2/3}(\lambda_{\max}(A) - 2)$ to illustrate this theorem.

In general when testing (1), we do not know θ, but as mentioned in the previous section, we do have an estimate $\hat{\theta}$. We therefore plug in $\hat{P}$ in place of P, where $\hat{P}_{ij} = P(G_{ij} = 1|\hat{\theta})$. We then define

$$\hat{A}_{ij} := \frac{Y_{ij} - \hat{P}_{ij}}{\sqrt{(n-1)\hat{P}_{ij}(1-\hat{P}_{ij})}}, \quad \hat{A}_{ii} = 0.$$

This suggests a test of (1) based on both $\lambda_{\max}(\hat{A})$ and $\lambda_{\min}(\hat{A})$, using the statistics $\hat{t}_1$ and $\hat{t}_2$, where the hat indicates that we replace the unknown P_{ij} with the estimate $\hat{P}_{ij}$.

We reject H_0 in (1) when

$$\max\{\hat{t}_1, \hat{t}_2\} > TW_1(1-\alpha/2) \text{ or } \min\{\hat{t}_1, \hat{t}_2\} < TW_1(\alpha/2), \tag{5}$$

where t_1 and t_2 are the test statistics from (3) and (4), $TW_1(\alpha/2)$ and $TW_1(1-\alpha/2)$ are the $(\alpha/2)$% and $(100-\alpha/2)$% quantile of the TW_1 distribution, respectively. If we instead use t_1 and t_2, this test has size α by a union bound argument. In practice, the test that uses $\hat{t}_1$ and $\hat{t}_2$ is size α if the eigenvalues of $\hat{A}$ converge quickly enough to the eigenvalues of A in probability. The next result, which we do not believe has previously been reported in the literature, gives a rate at which this happens.

Theorem 2. *If $n^{2/3}(\lambda_{\max}(\hat{A}) - \lambda_{\max}(A)) = o_P(1)$, then $n^{2/3}(\lambda_{\max}(\hat{A}) - 2) \xrightarrow{d} TW_1$. Furthermore, the test in (5) has size α as $n \to \infty$.*

Theorem 2 states that if we want to show that the test based on $\hat{P}$, rather than the unknown P, has size α as $n \to \infty$, we must prove that $\lambda_{\max}(\hat{A})$ converges fast enough to $\lambda_{\max}(A)$. To see this, since $n^{2/3}(\lambda_{\max}(A) - 2) \xrightarrow{d} TW_1$, if $n^{2/3}(\lambda_{\max}(\hat{A}) - \lambda_{\max}(A)) = o_P(1)$ then

$$\begin{aligned} n^{2/3}(\lambda_{\max}(\hat{A}) - 2) &= n^{2/3}(\lambda_{\max}(A) - 2) + n^{2/3}(\lambda_{\max}(\hat{A}) - \lambda_{\max}(A)) \\ &= n^{2/3}(\lambda_{\max}(A) - 2) + o_P(1) \xrightarrow{d} TW_1. \end{aligned}$$

https://doi.org/10.1017/nws.2025.10005 Published online by Cambridge University Press

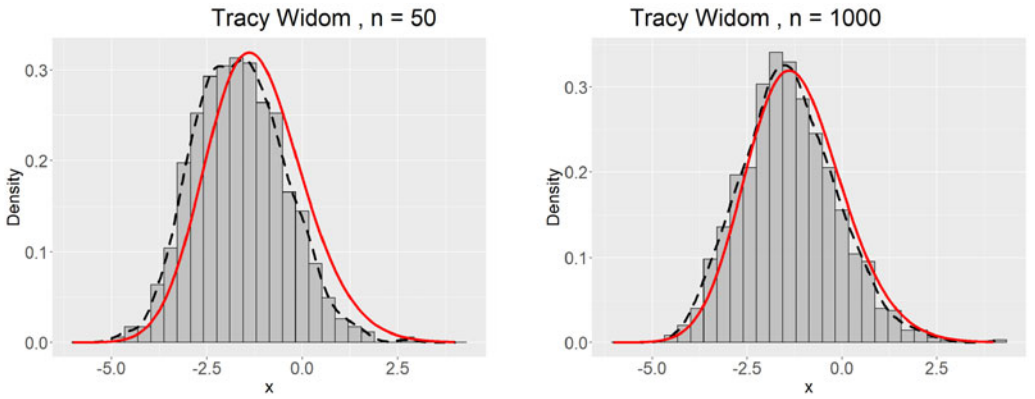


Figure 1. Distribution of statistic in Theorem 1 for $n = 50$ (left) and $n = 1000$ (right), where the solid curve in red corresponds to the Tracy-Widom distribution with $\beta = 1$. The difference in the distributions decreases as n increases, but the convergence is slow. This motivates the bootstrapping correction algorithm, given in Algorithm 1.

This condition is problem-specific and depends on the complexity of the graph distribution. Lei (2016) shows under certain constraints, the conditions of Theorem 2 hold in the case of a stochastic block model. To the best of our knowledge, there is no work that verifies this condition in other, more complicated models, such as the latent space model. In the simulations in this work, we assume that this condition is met and see that, in many models, we achieve an approximately size α test as $n \to \infty$. This suggests that in these models the condition of Theorem 2 holds.

Before continuing, we comment on the term $P(G_{ij} = 1|\hat{\theta})$. In many models, such as the SBM, this term is available in closed form in terms of $\hat{\theta}$. In other cases, such as ERGMs, this is not the case, since the graph model asserts a joint distribution over all pairs of edges. We are not aware of a formula for the marginal probability of a single edge in terms of the graph model. In these cases, we need to estimate the marginal probability matrix P. We present a simple method in Algorithm 3 to do this.

2.2 Directed networks

In the case where A is the adjacency matrix of a directed network, then Theorem 1 will not be applicable, since the eigenvalues of A are not guaranteed to be real. To test (1) in the case of directed networks, we therefore introduce two central limit theorems for the singular values of nonsymmetric random matrices, which always exist and are real. Both of these results assume a matrix with independent entries with (1) mean zero and (2) variance 1. Note that this differs slightly from the undirected case, where we required that the sum of the variance of entries in each row is 1. To satisfy conditions (1) and (2) in the directed case, we define the random matrix $\hat{A}$ with entries

$$\hat{A}_{ij} := \frac{Y_{ij} - \hat{P}_{ij}}{\sqrt{\hat{P}_{ij}(1 - \hat{P}_{ij})}}, \quad \hat{A}_{ii} = 0. \tag{6}$$

where again $\hat{P}_{ij} = P(G_{ij} = 1|\hat{\theta})$. Notice that there is no $(n-1)$ in the denominator of the expression for $\hat{A}$.

Theorem 3 (Theorem 1.1 of Johnstone, 2001). *Let A be a $m \times n$ standard Gaussian random matrix such that*

$$A_{ij} \sim_{iid} \mathcal{N}(0, 1) \quad \text{for all } 1 \leq i \leq m, 1 \leq j \leq n.$$

https://doi.org/10.1017/nws.2025.10005 Published online by Cambridge University Press

Let $s_{\max}(A)$ *be the largest singular value of* A*. Then, if* $m = m(n) \to \infty$*, with* $m \leq n$*, and* $\lim_{n\to\infty} m(n)/n = \gamma \in (0, 1]$*,*

$$\frac{s_{\max}(A)^2 - \mu_{1,n}}{\sigma_{1,n}} \xrightarrow{d} \mathrm{TW}_1,$$

where $\mu_{1,n} = (\sqrt{n-1} + \sqrt{m})^2$ *and* $\sigma_{1,n} = \sqrt{\mu_{1,n}}(1/\sqrt{n-1} + 1/\sqrt{m})^{1/3}$*. Moreover, if* $\gamma > 1$*, then the result remains true up to the swap of the roles of m and n in the formulas.*

Theorem 3 requires that the entries of A be Gaussian, which is not the case when A is the (binary) adjacency matrix of a random network. In Supplementary Material Figure D.3, we show that the convergence claim in Theorem 3 still holds reasonably well when the entries of A follow a Poisson binomial distribution. This suggests that we can use Theorem 3 to construct a test when the entries of A are not Gaussian. In fact, Pillai & Yin (2011) shows that the convergences of extreme eigenvalues to Tracy-Widom is universal, that is, as long as the entries are independent, have mean 0, variance 1, and have a sub-exponential tail decay.

For directed networks, Theorem 3 suggests that we take our test statistic to be $(s_{\max}(\hat{A})^2 - \mu_{1,n})/\sigma_{1,n}$ with $m = n$ and the rejection region to be $\{x : TW_1(\alpha/2) < x < TW_1(1-\alpha/2)\}$, which is identical to the undirected case. Moreover, it is not necessary to restrict $m = n$, which indicates the test statistic is also applicable on directed networks or networks for which we only have partial network data. We will elaborate on these ideas in a later section.

Our second result states that the scaled singular of a nonsymmetric random matrix, when suitably transformed as in (6), converge to an exponential-type distribution.

Theorem 4 (Theorem 2.4 of Chafaï, 2009). *Suppose that A is a random* $n \times n$ *nonsymmetric matrix whose entries have mean zero and variance 1. Then, if* $s_{\min}(A)$ *denotes the smallest singular value of A,*

$$\lim_{n\to\infty} \mathbb{P}\left(\sqrt{n}s_{\min}(A) \geqslant t\right) = \exp\left(-\frac{1}{2}t^2 - t\right).$$

Theorem 4 suggests that we take our test statistic to be $\sqrt{n}s_{\min}(A)$ and the rejection region to be $\{x : x > q_{\mathrm{E}}(1-\alpha)\}$ where $q_{\mathrm{E}}(1-\alpha)$ is the $(1-\alpha)100\%$ percentile of the distribution in Theorem 4.

To summarize, in this section we provided two central limit theorems for the singular values of random, nonsymmetric matrices, which require that the entries of the random matrix have mean zero and variance 1. We discussed how to use the observed adjacency matrix Y to construct such a matrix and to derive a test statistic for the GoF hypothesis in (1). In Section 4.4, we discuss the performance of these two test statistics.

2.3 Partial network data

Suppose our goal, as it was above, is to test whether a graph G is drawn from a particular model. That is, we want to test the hypothesis in (1). In many applications, complete network data are not available, is too expensive to collect, or cannot be collected for privacy-related reasons. A common form of partial network data, particularly in economics, is ARD (Breza et al., 2020, 2019; Alidaee et al., 2020). In this work, we focus on using ARD to test the GoF hypothesis in (1), but we believe our framework can be extended to other data types too.

To describe what ARD is, suppose that we can partition the nodes of the network into K categories $G_1, \ldots, G_K$. These categories correspond to different covariates, so, for example, all nodes in G_1 have black hair and all nodes in G_3 are left-handed. We then ask $m \leq n$ nodes how many people they know with trait j for $j = 1, \ldots, K$. In summary, by collecting ARD of a network with size n, we actually collect $\{Y_{ij} : i = 1, \ldots, m, j = 1, \ldots, K\}$, with

$$Y_{ij} := \sum_{k \in G_j} G_{ik}. \tag{7}$$

https://doi.org/10.1017/nws.2025.10005 Published online by Cambridge University Press

Recall that G_{ik} represents whether an edge exists between nodes i and k, so Y_{ij} represents how many people node i knows with trait j. We show in this section how to use ARD to test some of the network models previously mentioned. We assume, as is common in applications, that $|G_j|$ is known. Such information can come from census data or similar data sources.

To illustrate ARD, we consider a simple example. Suppose that we consider $K = 2$ and $m = 4$ and we then collect the following data Y, written in matrix form as

$$Y = \begin{pmatrix} 3 & 10 \\ 1 & 7 \\ 0 & 3 \\ 1 & 5 \end{pmatrix}. \tag{8}$$

This means, for example, that the first person we surveyed knows 3 people with trait 1 and the fourth person we surveyed knows 5 people with trait 2. In the above example, m does not have to equal K (and usually does not in practice), so Y is often not square. This means that we cannot apply Theorem 1 to test (1). Instead, we test the hypothesis with Theorem 3, which is applicable for non-square matrices.

One challenge is to estimate θ, given just the ARD. In this work, we consider a simple test of whether there is degree heterogeneity, which is equivalent to testing if the underlying model is an Erdös-Rényi model. Other, more complicated methods exist for estimating the parameters using only ARD in more complex models, such as those given in Alidaee et al. (2020) or Breza et al. (2020).

Before continuing, we discuss whether the assumptions in Theorem 3 hold for ARD. We discuss three assumptions. First, recalling the notation from Theorem 3, this result requires that $m \leq n$ so the matrix is "long" rather than "tall." In practice, the number of traits is smaller than the number of nodes we survey, so Y is often "tall," as it is in (8). This does not pose a problem since the singular values for Y and Y^T are the same. Second, Theorem 3 requires that the $m/n \to \gamma \in (0, 1]$, which in the ARD context requires that the number of traits grows with m. Previous work on the large sample properties of ARD estimators has either taken the number of traits as fixed in Breza et al. (2019) or growing slowly, like $K = O(\sqrt{n})$ as in Alidaee et al. (2020). Despite the assumption that K grow with the sample size, our simulations in Section 4.3 show that this result of Theorem 3 still hold reasonably well. Lastly, in Theorem 3, A is required to be a Gaussian random matrix with continuous entries. ARD are, however, counts. If the number of ARD responses are relatively large, then the counts may appear reasonably normally distributed. In most cases, however, we expect that the counts for most categories will be small. Despite these potential violations of the assumptions, we show in Supplementary Material Figure D.3 that the approximation works well, at least visually, when the entries of the random matrix are not Gaussian and the underlying data come from a skewed distribution of counts, as would be the case in ARD. We give simulation evidence in Section 4.3 that the approximation is sufficiently accurate to achieve favorable performance in hypothesis tests.

3. Bootstrap correction

As we saw previously, our asymptotic results can require a large sample size in practice. We derive a bootstrapping correction method, similar to the one presented in Lei (2016). We note that our algorithm generalized the one in Lei (2016). Algorithm 1 contains the algorithm for the undirected case, and Algorithm A.1 in Supplementary Material contains the algorithm for the directed case.

Before continuing, we make a few remarks. First, the intuition behind this method is as follows: the distribution of $t' \equiv (\lambda_1(\hat{A}) - \mu_{\max})/s_{\max}$ is approximately TW_1 except that the mean and variance are incorrect, but by scaling t' by s_{TW} and then shifting t' by μ_{TW} we obtain a better

https://doi.org/10.1017/nws.2025.10005 Published online by Cambridge University Press

Algorithm 1 Bootstrap correction of Undirected Tracy-Widom statistic

Input: Observed sociomatrix: G; Estimated probability matrix $\hat{P}$; Bootstrap iterates: B; TW_1 mean: μ_{TW}; TW_1 standard deviation: s_{TW}; Significance Level: α.

Compute

$$\hat{A}_{ij} = (G_{ij} - \hat{P}_{ij})/\sqrt{(n-1)\hat{P}_{ij}(1-\hat{P}_{ij})};$$

for $b = 1$ **to** B **do**

  Sample $G_b^\star \sim F_{\hat{\theta}}$;

  Compute

$$[A_b^\star]_{ij} = ([G_b^\star]_{ij} - \hat{P}_{ij})/\sqrt{(n-1)\hat{P}_{ij}(1-\hat{P}_{ij})};$$

  Set $\lambda^\star_{b,\max} = \lambda_{\max}(A_b^\star)$ and $\lambda^\star_{b,\min} = \lambda_{\min}(A_b^\star)$;

end

Set $\mu_{\max}$ to be the sample mean of $\{\lambda^\star_{b,\max}\}_{b=1}^B$ and $s_{\max}$ to be the sample standard deviation of $\{\lambda^\star_{b,\max}\}_{b=1}^B$. Set $\mu_{\min}$ and $s_{\min}$ similarly;

Compute the test statistic t

$$t := \mu_{TW} + s_{TW} \cdot \max\left(\frac{\lambda_{\max}(\hat{A}) - \mu_{\max}}{s_{\max}}, -\frac{\lambda_{\min}(\hat{A}) - \mu_{\min}}{s_{\min}}\right);$$

if $TW_1(\alpha/2) < t < TW_1(1-\alpha/2)$ **then**

  Do not reject t and set Rej = FALSE;

else

  Reject t and set Rej = TRUE.

end

Output: Rejection of bootstrap statistic: Rej.

approximation of a TW_1 distribution. Second, we know that both $\lambda_{\max}(A)$ and $\lambda_{\min}(A)$ have TW_1 distributions, and since we are taking a max over these two quantities, we want to use the $\alpha/2$ quantile of TW_1. This follows from a simple application of Bonferroni and leads to an α-size test. Finally, in Supplementary Material A, we give a similar bootstrap correction algorithm for directed network data.

4. Models

In this section, we demonstrate how our method can be used to perform model selection on a broad class of network models. We consider the following problems: (1) Comparing latent space models with different dimensions (Section 4.1). (2) Comparing ERGMs with different forms (Section 4.2). (3) Testing degree heterogeneity using ARD (Section 4.3). (4) Testing Community Structure in Directed Networks (Section 4.4).

4.1 Testing latent space models with different dimensions

The latent space model, originally proposed in Hoff et al. (2002), is a generative network model that asserts that each node in a network has a position on some latent space. The closer two nodes in this latent space are, the more likely they connect. There is a large literature on latent space

https://doi.org/10.1017/nws.2025.10005 Published online by Cambridge University Press

Algorithm 2 Dimension prediction for Latent Space model

Input: Observed sociomatrix: G.
Set $d_{\text{fit}} = 0$ and $T = 0$;
while $T = 0$ **do**
  Update $d_{\text{fit}} = d_{\text{fit}} + 1$;
  Compute the estimate $\hat{\theta}$ via Projected Gradient Descent algorithm with d_{fit}. Use $\hat{\theta}$ and the model in (9) to compute $\hat{P}$;
  Use $\hat{P}$ in the bootstrap algorithm (Algorithm 1) to determine if the null hypothesis $H_0 : d_{\text{true}} = d_{\text{fit}}$ is rejected;
  if H_0 *is rejected* **then**
    Set $T = 1$;
  else
    Remain $T = 0$;
  end
end
Output: Predicted latent dimension: d_{fit}.

models. See, for example, Handcock et al. (2007), Hoff (2005), Asta & Shalizi (2014), Oh & Raftery (2001), Shalizi & Asta (2017) and their references. In many cases, the user is interested in testing the dimension of the latent space as well as the geometry type. Lubold et al. (2020) shows how to, among other things, estimate the dimension of the latent space by using the clique structure in the network, and Wilkins-Reeves & McCormick (2022) shows how to estimate the latent space curvature using triangle side lengths and midpoints. In this work, we take a different approach and use the entire network to estimate the fit of the model to a hypothesized latent space dimension.

Let G denote the adjacency matrix with observed covariate matrix X. One form of the latent space model (Hoff, 2005; Ma et al., 2020) asserts that conditioned on the network parameters, edges form independently with probability

$$\begin{aligned} P(G_{ij} = 1|\theta) &= P_{ij}, \quad \text{where} \\ \text{logodds}(P_{ij}) &= \alpha_i + \alpha_j + \beta X_{ij} + \langle z_i, z_j \rangle, \end{aligned} \tag{9}$$

where $\text{logodds}(x) = \log(x/1-x)$, $\{\alpha_i\}_{i=1}^n$ are the parameters modeling degree heterogeneity, β is the coefficient scaling the observed covariate X, $\langle z_i, z_j \rangle$ are the inner products between latent positions with $z_i \in \mathbb{R}^d$, and d is the dimension of the latent space model. We let $\theta = (\alpha_1, \ldots, \alpha_n, z_1, \ldots, z_n, \beta)$ denote the collection of all the model's parameters.

Let G be an observed network drawn from the model in (9), where $z_i \in \mathbb{R}^{d_{\text{true}}}$. We develop a GoF procedure in Algorithm 2 via the Tracy-Widom statistic and a fast MLE method via nonconvex projected gradient descent, described in Ma et al. (2020). The details can be found in Algorithms 1 and 3 in Ma et al. (2020) and also in the supplementary R code. We now give a motivation for our algorithm. For any hypothesized dimension d, we can fit the model in (9) and check if we reject the hypothesis that this dimension fits the network well. With no covariates or node effects, we expect to reject the null hypothesis for $d < d_{\text{true}}$, since a lower dimensional embedding should fail to accurately model the network structure, whereas dimensions equal to and higher than d_{true} will capture the structure well and so we expect to fail to reject the corresponding hypotheses. This suggests that we should take the predicted dimension to be the smallest dimension for which we fail to reject the corresponding hypothesis. As we mentioned in the introduction, the node effects have a confounding effect on the estimation procedure, so that model fits from two distinct

https://doi.org/10.1017/nws.2025.10005 Published online by Cambridge University Press

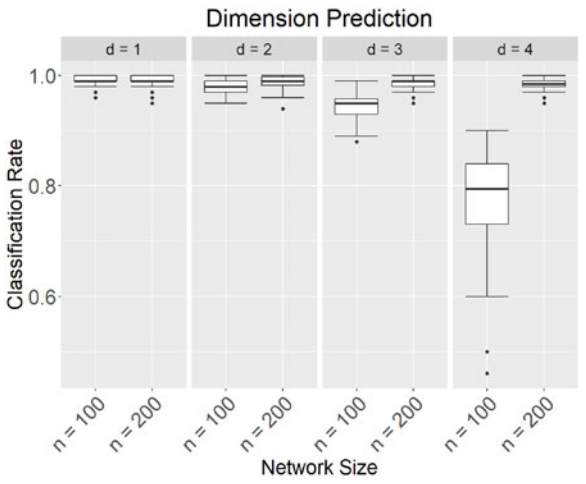


Figure 2. Correct classification rate for $n = 100, 200$ for the dimension of the latent space in Section 4.1. For a fixed n, increasing the dimension makes the problem harder and so the classification rate falls. However, the classification rate improves as we increase n.

dimensions, and their corresponding node effect estimates, might lead to equally good model fits. Even with the confounding issue, our simulations show that our procedure finds that the true dimension is often the smallest one that fits the model well.

In the following simulations, we will focus on the inner product model without covariate components. However, our algorithm can be generalized to any inner product models with "simple" covariates as described in Ma et al. (2020) by following an almost identical methodology. For $n = \{100, 200\}$, we generated 50 sets of $\{\alpha, z\}$ with $d = \{1, 2, 3, 4\}$ respectively, where $\alpha_i \sim_{iid} \text{Unif}(-2, -1) \times 10^{-2}$ and $z_i \sim_{iid} \text{N}(0, I_d)$. For each combination of $\{n, d, \alpha, z\}$, 100 networks are drawn from the corresponding generated models and predicted with Algorithm 2. The classification rates for each set of parameters are recorded and shown in Figure 2. We notice two trends. First, as n increases, the probability of correct dimension classification increases. Second, for a fixed n, larger dimensions are harder to classify correctly. This makes intuitive sense since higher dimensions often correspond to more complex latent space relationships, and so it takes more data to model these relationships well.

4.2 Comparing exponential random graph models with different forms

ERGMs are a common choice to model complex network data. To perform inference on these models, one must estimate an often intractable normalizing constant, which makes inference challenging. Some authors have presented maximum pseudo-likelihood Hunter et al. (2008) and Monte Carlo estimation methods. In this section, we show how to apply Theorem 1 to test the form of an ERGM.

We now briefly review the form of ERGMs. This model asserts that a random graph G arises through the model

$$P(G = g \mid \theta) = \frac{1}{c(\theta)} \exp\left(\sum_{i=1}^{K} \theta_i h_i(g)\right) \tag{10}$$

where $h_1, \ldots, h_K$ are functions of the graph g and $c(\theta)$ is the normalization constant. The user specifies the functions h as well as the value K. Some examples of h include $h(g) = \sum_{i<j} g_{ij}$, the

https://doi.org/10.1017/nws.2025.10005 Published online by Cambridge University Press

Algorithm 3 Given sociomatrix G, simulate $P\hat{o}$

Input: Observed sociomatrix: G; Bootstrap iterates: B.

Compute an estimate of θ, denoted by $\hat{\theta}$;
for $b = 1$ **to** B **do**
  Sample $G_k^* \sim F_{\hat{\theta}}$;
  Set A_k^* to be the $n \times n$ adjacency matrix for the graph G_k^*;
  Record A_k^*;
end
For all i, j, compute

$$\hat{P}_{ij} = \frac{1}{B}\sum_{k=1}^{B}[A_k^*]_{ij};$$

Output: Estimated probability matrix: $\hat{P}$.

number of edges in g, and

$$h(g) = \sum_{i,j,k} g_{ij}g_{jk}g_{kj},$$

the number of triangles in g. Except in simple cases, the MLE for θ, denoted by $\hat{\theta}$, is not available in closed form. We compute the MLE using the `ERGM` package in `R`.

Having estimated θ, we now need to estimate the $n \times n$ matrix P, where $P_{ij} = P(G_{ij} = 1|\theta)$. In most models, there is a clear correspondence between θ and P. For example, in a latent space model without covariates, once we estimate $\theta = (z_1, \ldots, z_n, \alpha_1, \ldots, \alpha_n, \beta)$, we can simply use the graph model in (10) to estimate P. But for ERGMs, the model in (10) asserts a model for the entire network G all at once, rather than specifying individual edge probabilities. To simulate P from $\hat{\theta}$, we therefore propose to simulate from the fitted model and record the number of edges between pairs of nodes across B simulations. We present this simple procedure in Algorithm 3.

Having now described how to estimate P from an estimate of the ERGM parameter, we now consider an ERGM model and show how to test the significance of its parameters. Consider the model

$$P(G = g) \propto \exp\left(\theta_1 \cdot \text{edges} + \theta_2 \cdot \text{triangle} + \theta_3 \cdot \text{kstar}(2)\right), \tag{11}$$

where edges counts the number of edges in g, triangles counts the number of triangles, and k-star(2) counts the number of 2-stars, which is a triangle with one edge missing.

Suppose that we are interested in testing whether $\theta_3 = 0$. In other words, we believe that the model above is correctly specified, with the exception that we do not know if $\theta_3 \neq 0$. Writing this as a hypothesis testing problem, we want to test the hypothesis

$$H_0 : \theta_3 = 0, \quad H_a : \theta_3 \neq 0. \tag{12}$$

To test this, we fit our data to the model in (11) with $\theta_3 = 0$. That is, we estimate (θ_1, θ_2) in the model $P(G = g) \propto \exp\left(\theta_1 \cdot \text{edges} + \theta_2 \cdot \text{triangle}\right)$. Let $(\hat{\theta}_1, \hat{\theta}_2)$ denote these estimates. We then simulate $\hat{P}$ using Algorithm 3. With these estimates, we can then form the matrix $\hat{A} = (G - \hat{P})/\sqrt{(n-1)\hat{P}(1-\hat{P})}$, with $\hat{A}_{ii} = 0$. We test H_0 using Algorithm 1. In Figure 3, we plot the

https://doi.org/10.1017/nws.2025.10005 Published online by Cambridge University Press

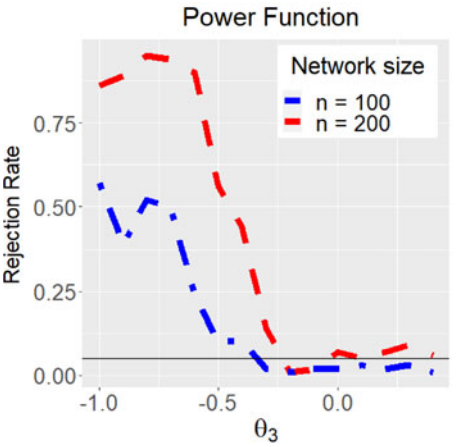


Figure 3. Power function for the hypothesis in (12). The dotted-dashed curve corresponds to $n = 100$ and the dashed curve corresponds to $n = 200$. The null hypothesis is $\theta_3 = 0$. The horizontal line represent the $\alpha = 0.05$ threshold.

power function for the hypothesis. We see that near $\theta_3 = 0$, the power is roughly equal to the Type I error $\alpha = 0.05$. As $|\theta_3|$ becomes larger, the power increases. We also see that the power increases for all $\theta_3 \neq 0$ as n increases.

4.3 Testing degree heterogeneity using ARD

In this section, we show an interesting application of our method to the case of partial network data. We focus on a particular type of partial network data known as ARD. This type of data is often cheaper to collect and can still be used to perform inference. For example, Breza et al. (2019) showed that the maximum likelihood estimate (MLE) for the latent space model, computed using only ARD instead of the entire network, is consistent as the graph size grows. Kunke et al. (2023) further compared the robustness of simple estimators of network scale-up method, which subsumes ARD.

Let G denote a network of interest on n nodes and suppose that we want to test if there is degree heterogeneity in the network. One way to model this question is through the following:

$$H_0 : g \sim \mathrm{ER}(p^\star) \text{ for some } p^\star, \quad H_a : \ H_0 \text{ is false.} \tag{13}$$

where $\mathrm{ER}(p^\star)$ denotes an Erdös-Rényi model with unknown parameter $p^\star$. Suppose that instead of observing the whole network g, we instead observe ARD.

Under the null hypothesis, each $Y_{ij} \sim \mathrm{Binomial}(n_j, p^\star)$, where $n_j = |G_j|$ is the size of group G_j. So if we define an $m \times K$ matrix A with

$$A_{ij} = \frac{Y_{ij} - n_j p^\star}{\sqrt{n_j p^\star (1 - p^\star)}},$$

then A is a $m \times K$ random matrix with mean zero and variance 1. Note that unlike in previous forms of A, in this case the diagonal of A is not set to be zero.

In general, we do not know $p^\star$ but given ARD, we can estimate $p^\star$ with

$$\hat{p} = \frac{1}{mK} \sum_{i=1}^{m} \sum_{j=1}^{K} \frac{Y_{ij}}{n_j}.$$

Under H_0, since $E(Y_{ij}/n_j) = p^\star$, it follows that $\hat{p} \xrightarrow{p} p^\star$ as $m \to \infty$. Here we consider K fixed; see the discussion at the end of Section 2.3. We can therefore define

https://doi.org/10.1017/nws.2025.10005 Published online by Cambridge University Press

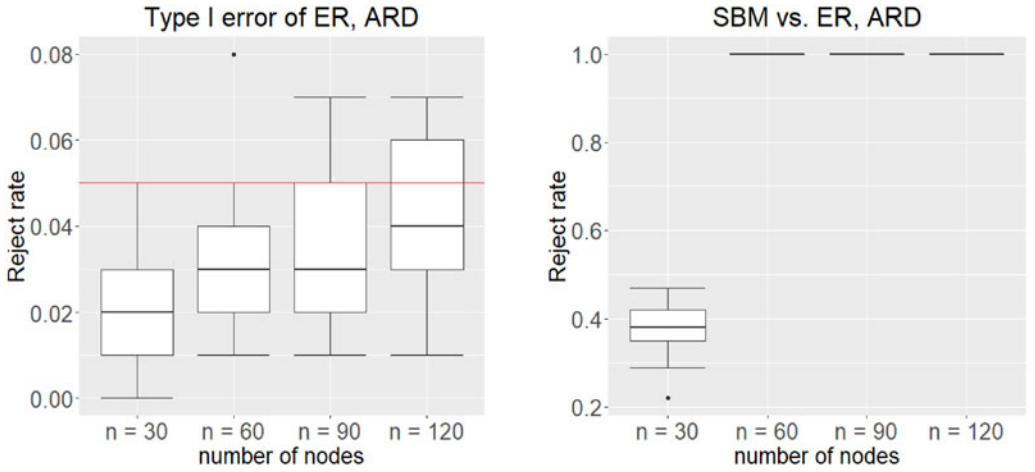


Figure 4. Left: Type I error of ER model via ARD. Right: Power of fitting SBM ARD to ER model. When the hypothesis model is correct, we observed a Type I error centered around the level of testing $\alpha = 0.05$. When the ARD of a more complex model is fitted to a simple hypothesis model (i.e. ER is a special case of SBM with one community), we will observe a very high power which grows with network size n.

$$\hat{A}_{ij} = \frac{Y_{ij} - n_j\hat{p}}{\sqrt{n_j\hat{p}(1-\hat{p})}},$$

We can use Theorem 3 to construct a test statistic for the null hypothesis. Our test statistic is the largest singular value of the matrix $\hat{A}$. Our rejection region for the null hypothesis is based on the quantiles of the Tracy-Widom distribution, as indicated in Theorem 3.

We first consider the Type I error of this method. For $n \in \{30, 60, 90, 120\}$, we draw an Erdös-Rényi graph with $m = \gamma_m n$ and $K = \gamma_K n$, where $\gamma_m = 1/3$ and $\gamma_K = 1/10$. We divide nodes equally into each of the K categories. Given a graph G, we define Y_{ij} as in (7). Our goal is to test whether G is drawn from an ER model. We plot our results in Figure 4.

Of course, more complicated testing problems can be used, but we leave that to future work. The goal of this section is to simply show how our method might be used to analyze network GoF in cases where only partial network data is available.

4.4 Directed network case

In this section, we show how to test (1) when the network is directed. Recall that Theorems 3 and 4 tell us the distribution of singular values and so they provide us with test statistics.

Suppose that we are given a directed graph g. We are interested in testing whether g is drawn from a directed Erdös-Rényi model. By this, we mean a directed graph whose directed edges form independently with probability $p^\star$. Our goal is to test

$$H_0 : G \sim \text{DER}(p^\star) \text{ for some } p^\star, \quad H_a : H_0 \text{ is false.} \tag{14}$$

where the notation DER(p) stands for a directed ER model. Theorems 3 and 4 give us test statistics to test this hypothesis. We start with the statistic from Theorem 3. This theorem states, informally, that the singular values of X, once rescale and re-centered, converge to a Tracy-Widom distribution. As in the undirected case, this convergence can be slow, so we use the bootstrap correction algorithm in Algorithm A.1 from the Supplementary Material. Theorem 4 also provides a test statistic to test (14). This theorem states, informally, that n times the largest singular value of a random matrix converges to an exponential random variable.

https://doi.org/10.1017/nws.2025.10005 Published online by Cambridge University Press

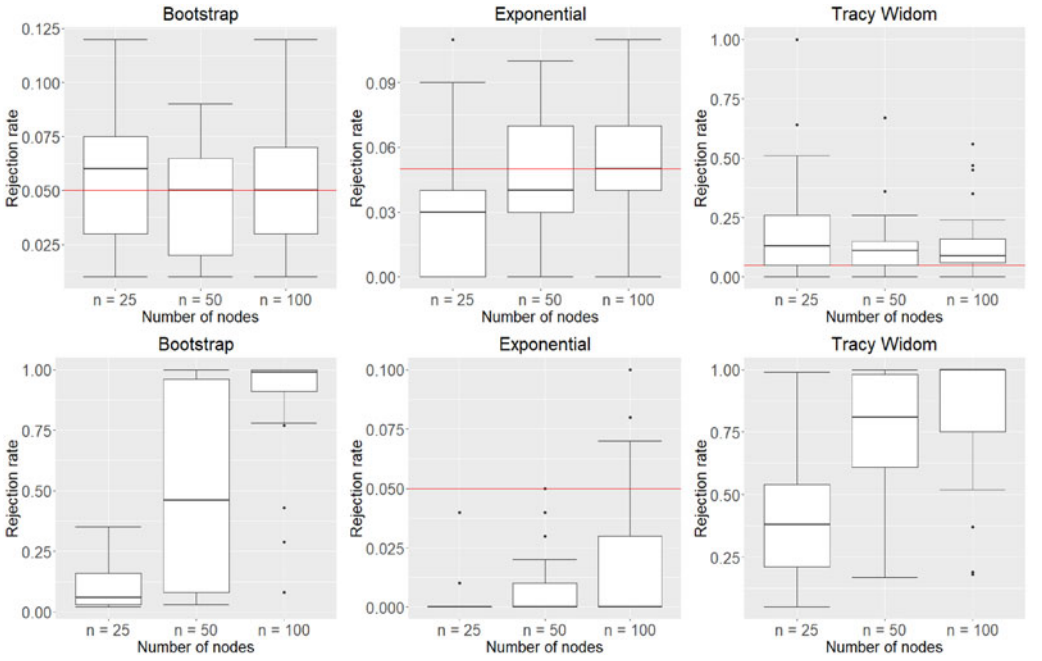


Figure 5. Type I error and rejection rates for directed network data. The first row corresponds to the case of a directed Erdös-Rényi model. In the top left figure, we plot the average rejection rate over 50 sets of simulations for $n = 25, 50, 100$ using the bootstrap test from Supplementary Material A. In the top middle, we plot the average rejection rate using the exponential test statistic in Theorem 4. In the top right, we plot the average rejection rate using Tracy-Widom test statistic in Theorem 3. In the second row, we plot the average rejection rates using a directed stochastic block model (DSBM) with 2 communities and distinct cross community probabilities. We see that bootstrap and exponential methods have good Type I error, yet that of Tracy-Widom statistics are relatively larger. In terms of power against DSBM, bootstrap and Tracy-Widom obtain good power, but the Exponential does not. Overall, the bootstrap statistic has a better performance in general.

Using these two theorems, we can test H_0 in (14). In Figure 5, we plot the Type I error in the first row for the "bootstrap" method from Algorithm A.1 and the "exponential" method. The second row plots the power of our method when g is drawn from a directed stochastic block model (DSBM) with two communities. We see that both methods have a good control on the Type I error at $\alpha = 0.05$, but only the "bootstrap" method is able to distinguish between a DER and a DSBM.

5. Community detection with latent space models

In this section, we analyze three data sets that are studied in Ma et al. (2020): the Political Blog data, the Simmons College data, and the Caltech data. Ma et al. (2020) fits these data sets to the latent space model in (9) without covariate components.

The authors computed the estimates of the latent space positions $\{z\}$ with the projected gradient descent methods, then applied a simple k-means clustering on the estimated latent positions for community detection. In Table 1 of their work, Ma et al. (2020) compares the clustering results with the community membership provided in the original data set, and reported the misclassification rate between the estimated clustering and the true network clustering. In this analysis, the fitted latent dimensions are set to either K or $K + 1$, where K is the known number of clusters in the data. We observed that fitting these datasets to different dimensions changed the misclassification rate, which suggests that choosing an optimal latent dimension is crucial for community detection.

https://doi.org/10.1017/nws.2025.10005 Published online by Cambridge University Press

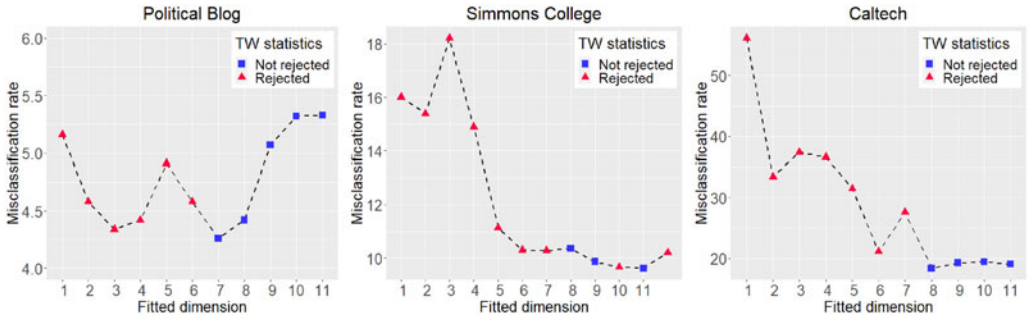


Figure 6. Misclassification rates of political blog, simmons college, and caltech data.

We made three major adjustments based on their evaluating procedure. First, instead of directly setting the latent dimension as K or $K+1$, we fit the data sets with Algorithm 2 and used the resulting d_{fit} as the fitted dimension. Second, the k-means method produces different clustering results even with the `replicate` command, so to avoid bias, we run the k-means clustering function 200 times in MATLAB and select the set of positions with the best fit. We then repeat this process 100 times and return the average misclassification rate across the 100 simulations. Lastly, instead of using only the first k eigenvectors of $\hat{z}$ as in Ma et al. (2020), we simply use the estimated positions $\hat{z}$ in the k-means algorithm. Our approach is intuitive, simple, and yields good performance on these three data sets.

We present our results in Supplementary Material Table B.2 and Figure 6. For Table B.2, in column t_{TW}'s, text labeled with star indicates the Tracy-Widom test statistics is not rejected. In column R_{mis}, bold text indicates the optimal classification rate. Figure 6 gives a visual representation of the misclassification rate over different latent dimensions.

In the Political Blog data and Caltech data, the optimal dimension chosen by our method are 7 and 8 respectively. The test statistics for $d_{\text{fit}} > d_{\text{opt}}$ are also not rejected. This behavior is similar to the behavior we saw in the latent space simulations. The optimal misclassification rates are also achieved at d_{opt}. Compared to the results in Table 1 of Ma et al. (2020), for the Political Blog data, we obtained a better misclassification rate, from 4.513% (latentnet) to 4.26% (Latent Space-based Community Detection (LSCD), $d_{\text{fit}} = 7$). For the Caltech data, we obtained the same optimal rate 18.35% (LSCD, $d_{\text{fit}} = 8$), as our predicted dimension coincides with the number of clusters. These results shows that the latent space model is a good fit of the two data sets, and our method performs well in achieving the optimal misclassification rate.

In the Simmons College data, the predicted dimension is $d_{\text{opt}} = 8$, with misclassification rate 10.37%. However, for $d_{\text{fit}} > d_{\text{opt}}$, we still observe that some fitted dimensions, namely $d = 10, 12$, are rejected. Moreover, the result for the Tracy-Widom statistics is not as robust as in previous two cases: our algorithm provides different predicted dimensions in different trials, whereas the results are consistent in the previous two data sets. This potentially suggests that the latent space model might not be a good fit for the Simmons College data. Nevertheless, our method still reveals certain natures of the network. The optimal rate is achieved at $d_{\text{fit}} = 11$, which is also substantially larger than the fitted dimension $d_{\text{fit}} = 4$ in Ma et al. (2020), at which our test statistic is not rejected. The misclassification rate is improved from 11.17% (LSCD, $d_{\text{fit}} = K+1$) to 9.62% (LSCD, $d_{\text{fit}} = 11$).

Our result shows that, based on the behavior of the test statistics with $d_{\text{fit}} > d_{\text{opt}}$, our Algorithm 2 potentially suggests whether the latent space model can be a good fit for the observed network. For networks that fit the latent space model well, our method will choose the optimal latent dimension that minimizes the community detection misclassification rate.

https://doi.org/10.1017/nws.2025.10005 Published online by Cambridge University Press

6. Conclusion

In this work we proposed a network GoF test that uses the eigenvalues of the centered, scaled adjacency matrix. We used recent work in random matrix theory to derive a test statistic that can test if an observed network is a good fit for common network models. This framework can handle undirected and directed networks, and can also handle cases where the researcher only has access to partial network data. We discussed the performance of this method on several common network models, like the latent space model, and showed that the test has favorable properties in terms of Type I error and power.

Theoretically, the developed GoF test depends on the assumption that Theorem 2 will hold under a given network model. Lei (2016) shows that this is the case for the stochastic block model, but it remains an open problem to show theoretically for other models. We demonstrate empirically that the result holds and show by simulations that the MLEs of several commonly used models are precise enough for our GoF test.

Along with the theoretical work suggested above, there are also many other avenues of future work. First, we would like to answer more general GoF questions when the researcher only has access to ARD. We believe that the estimation methods presented in Alidaee et al. (2020) can be used to estimate the $m \times K$ matrix P under a variety of realistic null hypotheses, which means that we can test the null hypothesis in (1) in a variety of more realistic settings. Second, we would like to extend this method to time-varying networks, such as those considered in Salter-Townshend & McCormick (2017). Finally, we would like to determine whether other random matrix theory results, such as Theorem 1 in Füredi & Komlós (1981), will lead to a test of (1) with better properties, like higher power.

Supplementary material. The supplementary material for this article can be found at https://doi.org/10.1017/nws.2025.10005.

Data availability statement. Code and datasets to replicate the simulations and figures in the paper are uploaded with submission as supplementary information. The realworld datasets analyzed in the paper are openly available at https://doi.org/10.1145/1134271. 1134277 (The political blog network) and https://doi.org/10.1016/j.physa.2011.12.021 (Two Facebook friendship networks of Simmons College and Caltech).

Funding statement. This work was supported by the National Institute Of Mental Health of the National Institutes of Health under Award Number DP2MH122405. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

Competing interests. None.

References

Airoldi, E. M., Blei, D. M., Fienberg, S. E., Xing, E. P., & Jaakkola, T. (2006). Mixed membership stochastic block models for relational data with application to protein-protein interactions. In *Proceedings of the International Biometrics Society Annual Meeting*, Vol. *15*.

Alidaee, H., Auerbach, E., & Leung, M. P. (2020).Recovering network structure from aggregated relational data using penalized regression. arXiv preprint arXiv: 2001.06052.

Amati, V., Lomi, A., & Snijders, T. A. B. (2024). A goodness of fit framework for relational event models [in en]. *Journal of the Royal Statistical Society Series A: Statistics in Society*, *187*(4), 967–988.

Asta, D., & Shalizi, C. (2014). Geometric network comparison. In *Uncertainty in Artificial Intelligence - Proceedings of the 31st Conference, UAI. 2015 (November)*.

Bernard, H. R., Hallett, T., Iovita, A., Johnsen, E. C., Lyerla, R., McCarty, C. et al. (2010). Counting hard-to-count populations: the network scale-up method for public health. *Sexually Transmitted Infections*, *86*(Suppl 2), 11–15.

Bickel, P. J., & Sarkar, P. (2015). Hypothesis testing for automated community detection in networks. *Journal of the Royal Statistical Society B*, *78*(1), 253–273.

Breza, E., Chandrasekhar, A., Lubold, S., McCormick, T., & Pan, M. (2019). Consistently estimating graph statistics using aggregated relational data. arXiv preprint arXiv: 1908.09881 (August).

Breza, E., Chandrasekhar, A. G., McCormick, T. H., & Pan, M. (2020). Using aggregated relational data to feasibly identify network structure without network data. *American Economic Review*, *110*(8), 2454–2484.

Chafaï, D. (2009). Singular value of random matrices. https://djalil.chafai.net/docs/sing.pdf.

Chatterjee, S. (2015). Matrix estimation by universal singular thresholdng. *Annals of Statistics*, *43*(1), 177–214.

Chen, D., & Hu, J. (2025). A new spectral-based goodness-of-fit for stochastic block models with strong power [in en]. *Stat (International Statistical Institute)*, *14*(2), e70065.

Chen, L., Lin, L., & Zhou, J. (2020). A hypothesis testing for large weighted networks with applications to functional neuroimaging data. *IEEE Access*, *8*, 191815–191825.

Chen, S., & Onnela, J.-P. (2019). A Bootstrap Method for Goodness of Fit and Model Selection with a Single Observed Network. *Scientific Reports*, *9*, 16674.

Erdos, L., Yau, H.-T., & Yin, J. (2012). Rigidity of eigenvalues of generalized wigner matrices. *Advances in Mathematics*, *229*(3), 1435–1515.

Erdős, P., & Rényi, A. (1959). On random graphs i. *Publicationes Mathematicae*, *6*, 290–297.

Fatima, A., & Reinert, G. (2025). A kernelised stein discrepancy for assessing the fit of inhomogeneous random graph models, eprint: 2505.21580 (stat.ML).

Füredi, Z., & Komlós, J. (1981). The eigenvalues of random symmetric matrices. *Combinatorica*, *1*(3), 233–241.

Gao, L. L., & DanielaWitten, J. B. (2019). Testing for association in multi-view network data. arXiv preprint arXiv: 1909.11640.

Gao, C., & Lafferty, J. (2017). Testing for global network structure using small subgraph statistics. arXiv preprint arXiv: 1710.00862.

Goldenberg, A., Zheng, A. X., Fienberg, S. E., & Airoldi, E. M. (2009). A survey of statistical network models. arXiv preprint arXiv: 0912.5410.

Gormley, I. C., & Murphy, T. B. (2010). A mixture of experts latent position cluster model for social network data. *Statistical Methodology*, *7*(3), 385–405.

Han, X., Yang, Q., & Fan, Y. (2023). Universal rank inference via residual subsampling with application to large networks. *The Annals of Statistics*, *51*(3), 1109–1133. doi: 10.1214/23-AOS2282.

Handcock, M. S., Raftery, A. E., & Tantrum, J. M. (2007). Model-based clustering for social networks. *Journal of the Royal Statistical Society A*, *170*(2), 301–354.

Hoff, P. (2003). Random effects models for network data. Working Paper no. 28, Center for Statistics and the Social Sciences, University of Washington (January).

Hoff, P. (2005). Bilinear mixed-effects models for dyadic data. *Journal of the American Statistical Association*, *100*, 286–295.

Hoff, P. D., Raftery, A. E., & Handcock, M. S. (2002). Latent space approaches to social network analysis. *Journal of the American Statistical Association*, *97*(460), 1090–1098.

Holland, P. W., Laskey, K. B., & Leinhardt, S. (1983). Stochastic blockmodels: first steps. *Social Networks*, *5*(2), 109–137.

Holland, P. W., & Leinhardt, S. (1981). An exponential family of probability distributions for directed graphs. *Journal of the American Statistical Association*, *76*(373), 33–50.

Hu, J., Zhang, J., Qin, H., Yan, T., & Zhu, J. (2021). Using maximum entry-wise deviation to test the goodness of fit for stochastic block models. *Journal of the American Statistical Association*, *116*(535), 1373–1382. doi: 10.1080/01621459.2020.1722676.

Hunter, D. R., Goodreau, S. M., & Handcock, M. S. (2008). Goodness of fit of social network models [in en]. *Journal of the American Statistical Association*, *103*(48), 248–258.

Hunter, D. R., Handcock, M. S., Butts, C. T., Goodreau, S. M., & Morris, M. (2008). Ergm: a package to fit, simulate and diagnose exponential-family models for networks [in English (US)]. Copyright: Copyright 2018 Elsevier B.V., All rights reserved. *Journal of Statistical Software*, *24*(3), 1–29. issn: 1548-7660. https://doi.org/10.18637/jss.v024.i03.

Jin, J., Ke, Z. T., Luo, S., & Wang, M. (2023). Optimal estimation of the number of network communities. *Journal of the American Statistical Association*, *118*(543), 2101–2116. doi: 10.1080/01621459.2022.2035736.

Johnstone, I. M. (2001). On the distribution of the largest eigenvalue in principalcomponents analysis. *The Annals of Statistics*, *29*(2), 295–327. doi: 10.1214/aos/1009210544.

Karwa, V., Pati, D., Petrovic, S., Solus, L., Alexeev, N., Raic, M. . . . Yan, B. (2024). Monte carlo goodness-of-fit tests for degree corrected and related stochastic blockmodels [in en]. *Journal of the Royal Statistical Society. Series B, Statistical Methodology*, *86*(1), 90–121.

Kunke, J. P., Laga, I., Niu, X., & McCormick, T. H. (2023). Comparing the robustness of simple network scale-up method (nsum) estimators. arXiv preprint arXiv: 2303.07490.

Lee, J. O., & Yin, J. (2014). A necessary and sufficient condition for edge universality of wigner matrices. *Duke Mathematical Journal*, *163*(1), 117–173.

https://doi.org/10.1017/nws.2025.10005 Published online by Cambridge University Press

Lei, J. (2016). A goodness-of-fit test for stochastic block models. *The Annals of Statistics*, *44*(1), 401–424. doi: 10.1214/15-AOS1370.
Lenk, P. (2009). Simulation pseudo-bias correction to the harmonic mean estimator of integrated likelihoods. *Journal of Computational and Graphical Statistics*, *18*(4), 941–960.
Lospinoso, J., & Snijders, T. A. B. (2019). Goodness of fit for stochastic actor-oriented models [in en]. *Methodological Innovations*, *12*(3), 205979911988428.
Lubold, S., Chandrasekhar, A., & McCormick, T. (2020). Identifying the latent space geometry of network models through analysis of curvature. arXiv preprint arXiv: 2012.10559.
Ma, Z., Ma, Z., & Yuan, H. (2020). Universal latent space model fitting for large networks with edge covariates. *Journal of Machine Learning Research*, *21*(4), 1–67. http://jmlr.org/papers/v21/17-470.html.
McCormick, T. H., Salganik, M. J., & Zheng, T. (2010). How many people do you know?: efficiently estimating personal network size. *Journal of the American Statistical Association*, *105*(489), 59–70.
McCormick, T. H., & Zheng, T. (2015). Latent surface models for networks using aggregated relational data. *Journal of the American Statistical Association*, *110*(512), 1684–1695.
Oh, M.-S., & Raftery, A. E. (2001). Bayesian multidimensional scaling and choice of dimension. *Journal of the American Statistical Association*, *96*(455), 1031–1044.
Ouadah, S., Robin, S., & Latouche, P. (2019). Degree-based goodness-of-fit tests for heterogeneous random graph models: independent and exchangeable cases. arXiv preprint arXiv: 1507.08140.
Pillai, N., & Yin, J. (2011). Universality of covariance matrices. *Annals of Applied Probability*, *24*(3), 935–1001. https://doi.org/10.1214/13-AAP939.
Raftery, A. E., Newton, M. A., Satagopan, J. M., & Krivitsky, P. N. (2007). Estimating the integrated likelihood via posterior simulation using the harmonic mean identity. *Bayesian Analysis*, *8*, 1–45.
Rohe, K., Chatterjee, S., Bin, Y., et al. (2011). Spectral clustering and the high-dimensional stochastic blockmodel. *Annals of Statistics*, *39*(4), 1878–1915.
Salter-Townshend, M., & McCormick, T. H. (2017). Latent space models for multiview network data. *Annals of Applied Statistics*, *11*(3), 1217–1244.
Shalizi, C. R., & Asta, D. (2017). Consistency of maximum likelihood for continuous-space network models. arXiv preprint arXiv: 1711.02123.
Shore, J., & Lubin, B. (2015). Spectral goodness of fit for network models. *Social Networks*, *43*, 16–27.
Stivala, A., & PengWang, A. L. (2025). Improving exponential-family random graph models for bipartite networks. eprint: 2502.01892 (stat.ME).
Tao, T. (2012). *Topics in Random Matrix Theory (Graduate Studies in Mathematics)*. Vol. 132. American Mathematical Society.
Wigner, E. P. (1958). On the distribution of the roots of certain symmetric matrices. *Annals of Mathematics*.
Wilkins-Reeves, S., & McCormick, T. (2022). Asymptotically normal estimation of local latent network curvature. arXiv preprint arXiv: 2211.11673.
Wu, Q., & Hu, J. (2023). A spectral based goodness-of-fit test for stochastic block models. eprint: 2303.14508 (stat.ME).
Wu, Y., Lan, W., Feng, L., & Tsai, C.-L. (2025). Testing stochastic block models based on maximum sampling entry-wise deviations. eprint: 2503.11990 (stat.ME).
Xu, W., & Reinert, G. (2021). A stein goodness-of-test for exponential random graph models. In *Proceedings of the 24th International Conference on Artificial Intelligence and Statistics*, Proceedings of Machine Learning Research, Vol. 130, pp. 415–423. PMLR. https://proceedings.mlr.press/v130/xu21b.html.
Yan, X., Shalizi, C., Jensen, J. E., Krzakala, F., Moore, C., Zdeborová, L.,. . .Zhu, Y. (2014). Model selection for degree-corrected block models. *Journal of Statistical Mechanics: Theory and Experiment*, *5*, P05007.
Yon, V., George, G., Slaughter, A., & de la Haye, K. (2021). Exponential random graph models for little networks [in en]. *Social Networks*, *64*, 225–238.
Zhang, L., & Amini, A. A. (2020). Adjusted chi-square test for degree-corrected block models, eprint: 2012.15047 (math.ST).

Cite this article: Lubold S., Liu B. and McCormick T. (2025). Spectral goodness-of-fit tests for complete and partial network data. *Network Science* **13**, 1–19. https://doi.org/10.1017/nws.2025.10005

https://doi.org/10.1017/nws.2025.10005 Published online by Cambridge University Press