Spectral goodness of fit for network models

Jesse Shore *, Benjamin Lubin

Boston University, Questrom School of Business, United States

ABSTRACT

We introduce a new statistic, 'spectral goodness of fit' (SGOF) to measure how well a network model explains the structure of the pattern of ties in an observed network. SGOF provides a measure of fit analogous to the standard R^2 in linear regression. Additionally, as it takes advantage of the properties of the spectrum of the graph Laplacian, it is suitable for comparing network models of diverse functional forms, including both fitted statistical models and algorithmic generative models of networks. After introducing, defining, and providing guidance for interpreting SGOF, we illustrate the properties of the statistic with a number of examples and comparisons to existing techniques. We show that such a spectral approach to assessing model fit fills gaps left by earlier methods and can be widely applied.

1. Introduction

Models of network structure play several important roles in contemporary science. Parametric statistical models of network structure and dynamics allow inferences to be made about dependencies among network ties, network position, and nodal and dyadic covariates (Frank and Strauss, 1986; Anderson et al., 1992; Snijders, 2001; Schweinberger and Snijders, 2003; Handcock, 2003; Doreian et al., 2005; Hunter and Handcock, 2006; Steglich et al., 2010). Algorithmic generative models illustrate how complex macroscopic structure can arise from simple and often local rules (Watts and Strogatz, 1998; Vázquez, 2003; Saramäki and Kaski, 2004). Despite the importance and diversity of research within both the model-based inference and generative algorithms categories, one aspect of network modeling research that has been relatively slow to develop is that of assessing goodness of fit, or how well a given model describes the empirical data being modeled. Moreover, the methods that are commonly used to assess fit within one type of model may be uncommon or unavailable in another, making it difficult to integrate research techniques and results across scholarly communities.

The purpose of this paper is therefore to define a new measure of goodness of fit that substantially fills the gaps left by current methods. In particular, leveraging the features of the spectrum of the graph Laplacian, we define a new goodness of fit statistic that measures the percent improvement a network model makes over a null model in explaining the structure in the observed data. As such, we provide a goodness of fit measure that can be applied across modeling techniques.

1.1. Existing methods

Commonly used existing methods for assessing goodness of fit can be roughly classified into two overlapping groups: one based on comparing structural statistics from networks simulated from a fitted model to structural statistics from the observed network (Hunter et al., 2008; Schweinberger, 2012), and the other based on a model's likelihood function, exemplified by the Akaike Information Criterion (Hunter et al., 2008). Some methods (for example, the score test) are both likelihood- and structural statistics-based.

1.1.1. Comparison of structural statistics

The most commonly used method of assessing goodness of fit is the structural statistics approach, which is implemented in software for estimating Exponential Random Graph Models (ERGMs) as well as dynamic actor-oriented models (also known as 'Siena' models). Although not done in a hypothesis testing framework, important algorithmic models (e.g. Watts and Strogatz, 1998) have also been described in terms of how well the algorithm reproduces structural statistics in observed networks.

In this approach, after fitting a model, it is necessary to generate a large number of simulated networks based on that model. At that point comparisons can be made between the observed and the simulated networks. The modeler might ask if the observed number of closed triads (or distribution of closed triads over the nodes) could have been drawn from the distribution defined by the simulated networks, or if the observed degree distribution could have been drawn from the distribution of degree in the simulated networks,

* Corresponding author. Tel.: +1 6173586390.
E-mail address: jccs@bu.edu (J. Shore).

or any number of other questions of fit between statistics describing the observed and simulated networks. If the structures in the observed network are very unlikely to have been generated by the fitted model, the modeler can reject the hypothesis that the model fits well. One such test—the score test—is frequently used in ERGMs and in certain cases can be interpreted as the uniformly most powerful unbiased test of the hypothesis that an omitted parameter for a given structural statistic is equal to zero (Holland and Leinhardt, 1981).

The structural-statistical approach has many advantages. By specifying different statistics to compare, the approach can be readily adapted to different specific questions of model fit. For example, one researcher may have a theoretical reason to emphasize the length of geodesics, while another may focus on triadic closure. Moreover, the statistics being examined need not be those used to fit the model; for example, one could use a community discovery algorithm (e.g. Newman and Girvan, 2004) to compare the distribution of community sizes in the observed network to those in simulated networks. The results of such an analysis are also easy to interpret and lend themselves to graphical representation and inspection (as in Hunter et al., 2008).

On the other hand, this method also has limitations. Even if the theoretical focus of a given researcher is on a single structural issue, say, modeling geodesics, the *overall* fit of the model to the whole network is still important. A model that accurately reproduces the distribution of geodesics but does not reproduce the overall structure of the network is probably inferior to one that captures the geodesic distribution and the overall structure simultaneously.

The difficulty in the subgraph-statistical approach is that it is not clear how to measure the overall structure of the network, except in terms of a list of its statistics. This approach necessarily decomposes the goodness of fit of a whole model into multiple goodness of fit tests on specific features of the model. Theoretically, this is problematic; practically, the validity of the goodness of fit assessment depends heavily on which statistics are specified by the researcher for examination. In a sense, in order to construct a valid goodness of fit test, the researcher is required to know a priori what the important statistics are for a given observed network; this is sometimes a nonsensical requirement, as goodness of fit tests are often undertaken exactly because the research does *not* know whether a given set of statistics (those described by the model parameters) are a good description of a network. The pragmatic solution is to use a commonly accepted set of statistics (Hunter et al. (2008) provides a good argument for one such set), but the possibility remains that important aspects of structure are not considered in such a goodness of fit test.

Additionally, assessing model fit in terms of subgraph statistics does not provide a means of selecting between two models that are both rejected or both not rejected. Finally, it is difficult to compare published results from different studies when they do not report the same subgraph statistical tests or analysis.

1.1.2. Akaike Information Criterion

Likelihood-based approaches, exemplified by the Akaike Information Criterion (AIC) and its cousin, the Bayesian Information Criterion (available for example, to users of the `ergm` package in R Handcock et al., 2014; Hunter et al., 2008), fills some of the gaps left by hypothesis tests on structural statistics.

There are several limitations of the AIC and related methods as well. First, many models do not have an AIC that can be easily computed, including ERGMs that are conditioned on having the exact number of edges present in the observed network, as well as models of networks that were not estimated from a statistical model at all (cases that we consider in more detail below).

Second, the likelihood from parametrized models is calculated in terms of the probability of the observed data conditional on the parameter estimates. This means that while the AIC can readily compare the fit of nested models, it is difficult to interpret when the models are not nested because the scale of the AIC depends on the parametrization of the model.

Third, like the structural-statistics approach to which it is related, one cannot know if there are omitted variables that would have improved the fit of the model.

1.2. Spectral goodness of fit

Given these tools already available to network modelers, a desirable measure of goodness of fit would have the following properties:

- it would be straightforward to compute for all models
- it would not require the modeler to know the true model or which structural statistics are important in the observed network
- It would allow comparison of a wide range of models, including non-nested models, those without easily computed likelihood functions or even without statistical parametrizations

Here, we propose such a statistic: spectral goodness of fit (SGOF). In many ways, SGOF is analogous to the R^2 used in standard linear regression. Both SGOF and R^2 have the properties just listed in part because they compare synthetic data (simulated networks for SGOF, predicted values for R^2) to observed data rather than working on the fitted models themselves. For SGOF these gains come at the expense of an important limitation (discussed at greater length in Section 5, which is that it is only a measure of goodness of fit for unlabeled (permutation-invariant) graph properties.

Throughout the rest of this article we make several assumptions. We consider only undirected networks explicitly, although we discuss extensions to directed networks in the final section, below. Additionally, in proposing to assess goodness of fit, we assume that a researcher has data on an observed network and has fit (or otherwise chosen) a model of network structure to that data. We do not make any assumptions about the functional form of that model or even whether the model is parametric at all, but we do assume that the researcher can generate simulated networks based on the 'fitted' model.

1.3. Computer code

We have made computer code for calculating SGOF and visualizing the results of the analysis available as an R package, `spectralGOF`.

2. The spectrum of the graph Laplacian

2.1. Definitions and notation

Networks are frequently represented as square adjacency matrices (which we will denote by A), such that if there is a link from node u to node v, then $A_{uv} > 0$. For the purposes of this article, we are considering only undirected networks, so $A_{uv} = A_{vu}, \forall u, v$. We also leave analysis of signed networks (in which edges can have a negative weight) for future work.

The Laplacian matrix is a transformation of the adjacency matrix given by $L = D - A$, where D is the 'degree matrix,' containing the row sums of A on its diagonal and zeros elsewhere. The spectrum of L is the ordered multiset of eigenvalues, $\boldsymbol{\lambda}$, such that $0 = \lambda_1 \leq \lambda_2 \cdots \leq \lambda_n$. There is one Laplacian eigenvalue (hereafter, for brevity, 'eigenvalues' and 'spectrum' always refer to the eigenvalues of the Laplacian) equal to zero for every connected component in the network (Brouwer and Haemers, 2011). Therefore, λ_1 is always 0. Further,

the sum of all eigenvalues is equal to the total weight of all edges in the network:

$$\sum_{i=1}^{n}\lambda_i = \sum_{u=1, v=1}^{n} A_{uv} \tag{1}$$

2.2. The spectrum of the Laplacian as a representation of network structure

The spectrum is a "graph invariant," meaning that if two networks are isomorphic,[2] then they have the same spectrum. The spectrum is also a compact representation of a great deal of structural information, and spectral techniques (sometimes including analysis of both the spectrum and its associated eigenvectors) have thus been used extensively to characterize the structure of complex networks (Pothen et al., 1990; Newman, 2006) and to compare and recognize complex objects in other applications such as facial recognition in computer vision (Turk and Pentland, 1991; Belkin and Niyogi, 2003). The properties of the Laplacian spectrum have been studied extensively (see Mohar and Alavi, 1991; Chung, 1997; Brouwer and Haemers, 2011 for relatively accessible mathematical overviews) and a full treatment is well beyond the scope of this article. However, to provide context for our definition of the spectral goodness of fit statistic, we here provide some basic intuition for the connection between the spectrum and network structure.

As we have already noted, the number of components is reflected in the spectrum by the number of zeros. The magnitude of the smallest non-zero eigenvalue is related to the minimum number of ties (how much total weight) that would have to be cut (i.e., removed from the network) to divide the network into two disconnected components and is known as the "algebraic connectivity" of a network (Fiedler, 1973). The magnitudes of the next smallest eigenvalues represent the relative modularity of the next-most macroscopic community structure of a network. Donetti et al. (2006) illustrate this logic as follows. Imagine a network comprising four totally disconnected components. Its spectrum would contain four eigenvalues equal to zero. If we perturb this network by connecting the components with a small number of ties (Cvetković et al., 1997), such that they are no longer disconnected, then rather than having one eigenvalue equal to zero for each component, we would have one *small* eigenvalue for each modular cluster (Donetti et al., 2006), one of which would be zero (as there would be one component, and thus one eigenvalue equal to zero). The more weight that was added between the components, the larger the eigenvalues would become.

The sizes of successively larger eigenvalues provide information on successively finer divisions of the network into smaller sub-communities. In general, a common interpretation of the magnitudes of eigenvalues of the Laplacian is one of correspondence to the relative weight removed by a series of minimum cuts of the network (for a more detail, see, e.g. Bollobás and Nikiforov, 2004). The largest eigenvalue therefore contains information about the number of ties incident to the single most highly connected node (Schur, 1923; Brouwer and Haemers, 2011). In general, the small eigenvalues contain information about the macroscopic structure of the network, while the large eigenvalues contain information about microscopic structure.

2.3. Normalizing the spectrum

The shape of the spectrum describes how the aggregate tie strength in a given network is organized relative to other networks with the same density (total tie weight). Given this, in the definition of the spectral goodness of fit (SGOF) statistic below, we normalize all spectra to sum to unity.

$$\hat{\lambda}_i = \frac{1}{\sum_j \lambda_j}\lambda_i \tag{2}$$

An increase in the density of A that is accomplished by multiplying all entries in A by a non-zero scalar constant also does not change $\hat{\lambda}$. In other words, such a change only alters the size and not the shape of the original spectrum, λ. On the other hand, an increase in the density of A accomplished by adding new ties or increasing the strength of certain ties and not others, both increases the sizes of λ and changes its shape: and it results in a changed $\hat{\lambda}$ as well.

3. Spectral goodness of fit

3.1. Spectral distance

Given the structural information contained in the spectrum, the Euclidean distance between two spectra is frequently used as a measure of the structural similarity of two matrices (Cvetković, 2012). The Euclidean spectral distance (ESD) can be written as $||\hat{\lambda}^A - \hat{\lambda}^B||$, where the normalized full spectra of graphs A and B are given by $\hat{\lambda}^A$ and $\hat{\lambda}^B$ respectively, and the double bars denote the vector norm.

We wish to apply this notion of distance to our network models, but such models do not have spectra themselves. However, if networks can be generated from the model (e.g. by simulation), spectra for these generated networks can then be calculated. It is the distance between these spectra and the observed spectrum that we will consider. If we have, say, $N_{sim} = 1000$ simulated networks, we can calculate the mean spectral distance between the simulated networks and the observed network, as well as other distributional statistics, such as the 5th and 95th percentiles of the spectral distance between simulations and the observed network.

Formally, after normalizing the spectra as above, let us call the absolute value of the difference between the ith observed eigenvalue and the ith eigenvalue from the kth simulated network an 'error.'

$$\epsilon_{k,i} = \left|\hat{\lambda}_i^{obs} - \hat{\lambda}_i^{sim_k}\right| \tag{3}$$

In this context then, ESD is the square root of the sum of squared errors.

$$ESD_{obs, sim_k} = ||\hat{\lambda}^{obs} - \hat{\lambda}^{sim_k}|| = \sqrt{\sum_i (\epsilon_{k,i})^2} \tag{4}$$

The mean Euclidean spectral distance, $E\bar{S}D$, is then defined as arithmetic mean of the ESDs from each of the individual simulated networks.

$$E\bar{S}D_{obs, sim} = \frac{1}{N_{sim}}\sum_{k=1}^{N_{sim}} ESD_{obs, sim_k} \tag{5}$$

3.2. Definition of null model

For network models we propose that goodness of fit be measured as an improvement in fit relative to a naive null model. It is therefore necessary to calculate the errors under the naive model and the fitted model for some number of simulated networks.

---

[2] Isomorphic networks have the same structure. They could be represented by the same adjacency matrix after permuting the rows and columns and disregarding any "labels" or names of the nodes.

The natural null model for dichotomous networks is the density-only model, also known as the Bernoulli model or Erdős-Rényi model, simulations from which are random networks with the same expected density as the observed network. For the remainder of this article, we adopt the density-only model as a null model.

In certain cases, other null models may be justified. The critical factor is specifying the null model such that it produces networks in the same "class" as the observed network. Directed networks should be have a directed null model (a directed Erdős-Rényi random graph); weighted networks should have a weighted null model (this could be done by randomly permuting the observed weights, for example, but also see Krivitsky (2012) for a parametric approach) and so on.

Another situation where the Erdős-Rényi model would not be appropriate as a null model is the case where the observed network has constant degree by construction. This could occur if measurement of the observed network was by means of a survey instrument that specified the number of alters each respondent was to nominate ('name five people you discuss important matters with'). In this case a degree-regular random graph (one in which each node has the same degree) would be the appropriate null model. For typical cases, however, the Erdős-Rényi model should be the default choice for the null model.

3.3. Definition of SGOF

To calculate the spectral goodness of fit (SGOF), we simply divide the mean Euclidean spectral distance under the fitted model by the mean Euclidean spectral distance under the null model, and subtract the result from one.

$$\text{SGOF} = 1 - \frac{\bar{ESD}_{obs,fitted}}{\bar{ESD}_{obs,null}} \tag{6}$$

3.4. Variability of SGOF

Although we use the mean of the *ESD* in the construction of our summary statistic, it is also important to consider (1) the standard error of the mean SGOF and (2) the variability in SGOF for simulated networks.

A small standard error provides evidence that sufficient simulated networks have been obtained to get a reliable estimate of the mean. As a rule of thumb, we recommend using 100 simulations for exploratory modeling and at least 1000 simulations for published results. However we also suggest that if the standard error of the mean is greater than 0.01, more simulated networks should be obtained to ensure reliability. By paying attention to such standard quality measures of our estimate of the mean, we can ensure that we are drawing appropriate conclusions from our SGOF statistic.

Additionally, given that almost all network models are stochastic in nature, they will produce a distribution of networks. Different models will have different variability in this distribution, and it is possible for a fitted model with a reasonably high mean SGOF to still produce simulated networks that are poor approximations of the observed data some of the time. Consequently, it is important to be able to get a quantitative sense of the range of networks a given model produces. We can use SGOF in service of this goal by simulating a large number of networks and then calculating SGOF with the 5th and 95th percentile results for $ESD_{obs,sim}$. Below, we report these in parentheses after the mean SGOF calculated using Eq. (6). This inter-quantile interval provides an indication of the dispersion of goodness of fit, a property inherent to the model. It is important to keep in mind that this variability in SGOF represents variability in the simulated networks generated by a model, and not uncertainty in the fit of the model to the observed data.

Concretely, if an observed network is not highly structured (that is, close to random), the inter-quantile range for the null model's SGOF will be very wide, extending, say, from −0.5 to 0.5, reflecting the fact that the observed network is not far from random. For observed networks with a great deal of structure, this interval for the null model's SGOF will be narrow, extending for example only from −0.001 to 0.001.

3.5. Interpretation of SGOF

The SGOF measures the amount of observed structure explained by a fitted model, expressed as a percent improvement over a null model, where structure means deviation from randomness. The observed spectrum will be distant from the spectrum of the null model in as much as the observed network has structure that is non-random. The SGOF is thus a summary measure of the percent of the observed structure that is explained by the fitted model.

3.5.1. Bounds for SGOF

Like R^2, SGOF is bounded above by one, when the fitted model exactly describes the structural data. Likewise, an SGOF of zero means no improvement over the null model. Finally, as with R^2, SGOF can be unboundedly negative[3] if the spectrum of the fitted model is more distant from the observed spectrum than is the spectrum of the null model. If the SGOF is negative, it is therefore evidence that the null model (an Erdős-Rényi random graph) is a better approximation of the observed network than the fitted model under consideration. This is likely to occur in cases where the observed network is not highly structured (and thus very similar to the null model), and the fitted model is (incorrectly) highly structured. If the observed network is not structured, then while $\bar{ESD}_{obs,fitted} \gg 0$, $\bar{ESD}_{obs,null}$ will be a small positive number. Thus by Eq. (6), SGOF will become a negative number of very large magnitude. However, for ordinary cases involving an observed network that contains structure, and this structure can be explained by a sensible model specifications, SGOF will fall between zero and one.

4. Applications and comparisons to existing methods

In this section, we illustrate the spectral goodness of fit method with several examples chosen to highlight its strengths and weaknesses with respect to existing methods.

4.1. Comparison with structural statistics: E. coli

It is frequently the case that a researcher does not ever discover the 'true' model underlying the formation of an observed network, but rather is only able to approximate the truth with several theoretically plausible candidate models. In such cases it is useful to have quantitative evidence about model goodness of fit to help adjudicate the decision. Structural statistical tests can sometimes play this role, but as mentioned above, it may also be the case that all models under consideration are rejected (or supported) by the test, and more information is therefore needed.

This example considers such a situation by comparing two specifications of a model of the degree distribution of the *E. coli* genetic regulatory network (Shen-Orr et al., 2002), both in the ERGM framework.

---

[3] In normal practice, however, the fitted model for R^2 is an ordinary least squares linear regression with a free intercept parameter; in this typical case, R^2 is bounded below by zero.

Using the `ergm` package in R, after fitting the models, we assessed their goodness of fit in the manner described by Hunter et al. (2008), using the `gof` function with its default settings. This goodness of fit routine assesses the probability that the distributions of degree, transitive closure and mean geodesic lengths over the nodes in the observed network could have been generated by the fitted model. Results from the `gof` analysis show that both of the proposed model specifications produce distributions of structural statistics that diverge from the observed values. Accordingly, the p-values for the goodness of fit diagnostics (not shown) indicate rejection of the models.

Table 1 indicates this and gives values for the SGOF for these models, along with small network visualizations for reference. Although all the models are rejected by structural hypothesis tests, there are marked differences in how well these models fit. Specifically, the "curved exponential family" version of the model (for more detail, see Hunter and Handcock, 2006) provides a much better fit to the data than the other model without the curved exponential family specification. In fact, at −0.014, the SGOF of this model indicates that it is no better than the null model as an overall description of the structure of the observed data.

The simple lesson here is that goodness of fit based on structural statistics cannot easily distinguish between similar models when all of the models are either accepted or rejected. Visual inspection of the graphical output can often help in this regard, but it is not hard to come up with examples where it cannot. In these cases it would be good to have a measure of goodness of fit that can provide a principled means of model choice. The AIC is thus a more comparable measure of goodness of fit to the SGOF we propose here, and the following examples make the comparison explicit.

4.2. Comparison with AIC: star graph plus dyad

We next consider an example consisting of a 102-node graph constructed by hand to serve as an imaginary observed network. 100 of the 102 nodes form a star graph, and the remaining two are a complete dyad, disconnected from the rest of the network. In addition to the network ties, there is an observed attribute, indicated by the color of the nodes in the visualization. The attribute values have been measured by our hypothetical researcher, but they were not part of the process that generated the network ties. For this example, we compare the SGOF to AIC from fitted models in the `ergm` package (Table 2).

After the null model, the next model is one fitted with a term for homophily among red nodes in addition to the density term. The visualization shows that such a model produces a pattern of ties that is very similar to the null model, but a greater proportion of ties among red nodes, similar to the observed network. It is here that one major difference between SGOF and AIC can be seen. The SGOF indicates negligible improvement over the null model because the pattern of ties is only a negligible improvement over the null model. Meanwhile the AIC shows a substantial improvement, from 988.4 to 957.27, because the parameters of the fitted model, including a (spurious by construction) homophily effect, have a higher likelihood than the parameters of the null model, even after accounting for the number of parameters with Akaike's formula. The AIC is sensitive to how well the model's parameters fit the data as a whole, including non-structural data.

The third and fourth models are both ERGMs fit to the data with a k-star parameter (tendency toward nodes with degree k) in addition to the density parameter, but they differ in how the k-star parameter is specified. The first of the two parametrizes the network with a tendency toward 2-stars, while the second of the two parametrizes the network with a tendency toward 3-stars. Note that the k-stars are induced subgraphs, so although there are no

nodes with degree two, there are $\binom{99}{2} = 4851$ two-stars, each centered on the same node; likewise there are $\binom{99}{3} = 156,849$ three-stars in the observed network. Because of these differences in statistic counts, the estimated parameters are likewise different in the fitted models, with the parameter for 2-star tendency and 3-star tendency estimated at 0.1283 and 0.0026, respectively. Reductions in AIC differ because the parameter estimates are different and the underlying "data" (i.e. observations of 2-stars versus 3-stars) are even different.

Yet, both of these models reliably produce simulated networks that are very close to the observed network. Indeed, the simulated networks produced by the 2-star model are indistinguishable from those produced by the 3-star model. Those depicted in the visualization differ from the observed network only in that the two nodes that are not part of the star graph are not connected with a tie. Consequently, the SGOF for both of these models is very high at 0.978 (0.961, 0.99). According to the AIC, however, the two models are very different: the 3-star model is only a little better than the null model, with an AIC of 868.14, while the 2-star model is characterized as superior, with an AIC of 552.88, despite generating essentially the same network. Unlike the SGOF, the AIC depends on the model specification, so it can be extremely difficult to interpret when models are not nested.
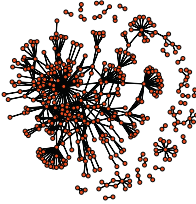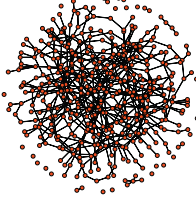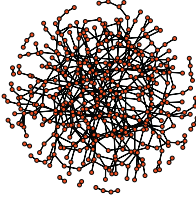
4.3. Second comparison to AIC: Faux Mesa High

The previous example was artificially constructed to illustrate the differences between AIC and SGOF. In this subsection, we give an example of a more typical social network using the "Faux Mesa High" data set of Hunter et al. (2008), adapted from the Add Health surveys (Harris and Udry, 2008). Similar to the star-graph example, above, after the null model we fit an ERGM model using only homophily effects on the observed covariates, which describe Race, Sex and Grade of the respondents. We go on to fit a model using only the "Geometrically Weighted Degree" (GWD) of Hunter and Handcock (2006) (which is a flexible approach to modeling degree distributions), followed by a model with both the GWD and homophily effects. The final model we consider differs in type: the preferential attachment model of Barabási and Albert (1999). Visualizations of the networks created by these models, as well as their AIC and SGOF statistics are shown in Table 3.

In this example, the homophily on the three covariates makes significant improvements in both SGOF and AIC, because unlike the star graph, there is almost certainly a real homophily effect in the original data. Likewise, both SGOF and AIC indicate that the model with both GWD and homophily is superior to the models with just one of those two types of effects. The lessons from Faux Mesa High are, however, otherwise consistent with those from the star graph. AIC indicates that the homophily-only model is superior to the GWD-only model. However, from the point of view of generating a pattern of ties alone, the SGOF indicates that the GWD-only model is superior to the homophily-only model. Again, the reductions in AIC due to modeling homophily are hard to compare to reductions in AIC due to modeling GWD, because the likelihood is calculated on different statistics. In contrast, because the SGOF compares whole simulated networks to the whole observed network, it does not depend on model specification. For models that are better than a null model (for unweighted networks, this is an Erdős-Rényi random graph), it will always fall between 0 and 1.

In Table 4 we compare the best ERGM from Table 3 and compare it to models that are algorithmic in nature rather than statistical. For all of these models we use the implementation in the `igraph` package (Csardi and Nepusz, 2006). We consider the

Table 1
Comparison of spectral goodness of Fit to structural hypothesis testing for the *E. coli* genetic regulatory network.

	SGOF	Struc. h-test	AIC	Simulated network
Observed network				
Null model	0 (−0.021, 0.024)	Reject	6355.11	
Geom. weighted degree	−0.014 (−0.033, 0.009)	Reject	6271.42	
Geom. weighted degree (curved exp. family)	0.247 (0.17, 0.336)	Reject	6013.68	

Note: All ERGMs include the terms specified above, plus an "edges" parameter.

Watts–Strogatz lattice model of small world networks (Watts and Strogatz, 1998), the Barabási-Albert preferential attachment model (Barabási and Albert, 1999) in its default specification and in a hand-tuned specification, and the 'configuration model' (Newman et al., 2001). These models are based on a generative algorithms, rather than parameters fitted to data that could be used to easily compute a likelihood function. Although in theory one could calculate a likelihood for algorithmic models by generating many simulated networks and counting the fraction of these that were identical to the observed network, this procedure is not feasible in practice because it could take years of computing time to carry out. Therefore, in the table, we have indicated that the AIC for these models is effectively incomputable with a question mark.

The SGOF is straightforward to calculate, however, as it is for *any model* that generates networks with the same number of nodes as the observed network, regardless of conditions put on the sample space or how (or whether) the model was estimated. As such, the SGOF makes it possible to compare models that cannot be compared on the basis of the AIC or other likelihood-based methods.

Examining the SGOFs from Table 4, we see that the Watts–Strogatz lattice model of small world networks, and the default preferential attachment model both produce networks that are much worse approximations of the pattern of ties in Faux Mesa High than the null model—an Erdős-Rényi random graph. This is not a surprising outcome, given that they were not "fitted" to the observed data in any way apart from having the same number of nodes. A hand-tuned version of the preferential attachment, however, can produce a fit that is only slightly inferior to the best ERGM

with respect to the unlabeled graph. The configuration model is slightly different in nature. This model creates random graphs with the exact degree distribution of the observed graph. Given that the degree distribution is guaranteed to be the same as the observed data it is not surprising that the SGOF is high, at 0.77. Since the degree distribution of the simulated networks is fixed, there is also less variability in SGOF: the 5th and 95th percentile values for SGOF are very close to the mean.
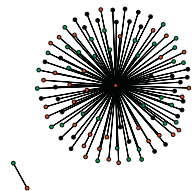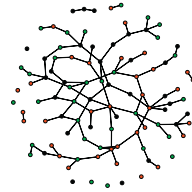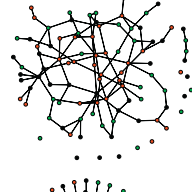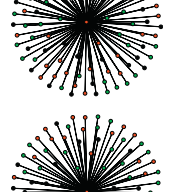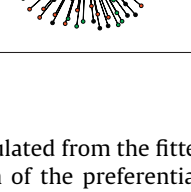
4.3.1. AIC versus SGOF

In practice the AIC and the SGOF are complementary in that they provide answers to different modeling questions. A researcher may wish to compare nested models, examine the fit of nodal or dyadic covariates in explaining the observed labelled graph, or assess the parsimony of the model. In these cases, the AIC is required. On the other hand, the researcher may wish to compare non-nested models, compare models for which the likelihood function is impractical to compute, or know how well a model that includes both structural effects and nodal and dyadic covariates explains the unlabeled pattern of ties alone. In these cases the SGOF is preferred.

4.4. Visualization of SGOF

As with other statistical methods, a fuller qualitative understanding of the SGOF can be gained through visualization. Fig. 1 plots spectral fits for the "GWD and Homophily" and the "Preferential attachment" models from Table 3, using the `plotSGOFerrors` function in the `spectralGOF` package.

Table 2
Comparison of spectral goodness of fit to AIC for a nearly star graph.

	SGOF	AIC	Simulated network
Observed network			
Null model	0 (−0.01, 0.016)	988.4	
Red node homophily	0.005 (−0.007, 0.023)	957.27	
2-Star tendency	0.978 (0.961, 0.99)	552.88	
3-Star tendency	0.978 (0.961, 0.99)	868.14	

Note: All ERGMs include the terms specified above, plus an "edges" parameter.

Each panel of the figure is a visualization of spectral error based on three spectra: the observed spectrum, the null model spectrum that is closest to the observed spectrum (by mean Euclidean distance), and the fitted model spectrum that is closest to the observed spectrum (again by mean Euclidean distance). The first and the second are the same in both panels and are plotted as grey and black points respectively. The fitted model spectrum differs between panels, and is plotted using orange points.

The spectral errors (differences among the observed, null and fitted model spectra) are indicated by colored bars as follows. When the fitted model's spectrum lies between the null and the observed spectra, the fitted model has improved the fit. The distance between the null and the fitted spectrum is error that has been "explained" and is indicated in light green. The error that still remains unexplained (error that is present under the null and the fitted models) is indicated in blue.

There are also parts of the plots where the fitted and null spectra are on opposite sides of the observed spectrum. In these cases, the fitted model has "explained" the error between the null and the observed, but introduced new error on the other side of the observed spectrum. This new error is indicated in red.

Turning to the specific models in Fig. 1, we see that the two fits differ considerably. In general, the spectrum of the fitted ERGM (top) lies between the observed spectrum and the null spectrum, indicating that the observed network is more structured (farther from random) than are networks simulated from the fitted ERGM. In contrast, portions of the spectrum of the preferential attachment model (bottom) are more distant from the null spectrum than is the observed spectrum. The preferential attachment model has explained more error than the ERGM (represented by more green area in its visualization), but it has also introduced structure not present in the observed network, producing more new error (more red area in the visualization), and resulting in a lower net SGOF.

4.5. SGOF as an objective function: collaborations among jazz musicians

There are cases where one wishes to implement algorithmic models that do not have an intrinsic method for fitting to observed data. In this case, SGOF can be useful as an objective function in an exploration of the algorithm's parameter space. To illustrate this type of application, we consider the network of jazz collaborations described by Gleiser and Danon (2003).

One theoretically plausible algorithmic model of how collaboration networks are formed is that of Saramäki and Kaski (2004). In this model, one assumes some network exists at t_0 to initialize the model. In subsequent time points, new individuals arrive and form ties to those already present by means of short random walks from a randomly selected node serving as the point of entry into the network.

Table 3
Comparison of spectral goodness of fit to AIC for Faux Mesa High.

	SGOF	AIC	Simulated network
Observed network			
Null model	0 (−0.197, 0.202)	2287.742	
Homophily on race, sex, grade only	0.223 (0.008, 0.445)	1890.922	
GWD Only	0.267 (−0.041, 0.556)	2244.7	
GWD and homophily	0.494 (0.236, 0.696)	1853.849	

Note: All ERGMs include the terms specified above, plus an "edges" parameter.

For musicians, the idea would be that after collaborating with some initial partner, one is likely to get to know one's partner's partners, and so on. In addition to being theoretically plausible, this algorithm generates networks with skewed degree distributions and local clustering, as we observe in the jazz collaborations data set.

To assess the fit of this model, one must first find the best values for the model's parameters, which we will do by appeal to SGOF. When implementing the algorithm, we identified two key parameters for fitting. The first is the mean number of edges to add with each new node added to the network. The second is how many steps in a random walk a new node takes before forming new relationships to existing members of the network. We generated 100 simulated networks for each of the using various combinations of values for these parameters, and then used these to compute an SGOF for each such parameter specification.

The results of this process are shown in Fig. 2, and indicate that the best fit occurs when the average number of edges added per node is 9, and the random walk distance is a single step. Thus we can not only use SGOF as a diagnostic tool, but also as a means

for identifying the parametric model settings that will be optimal under our spectral criterion.
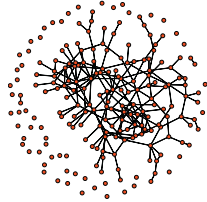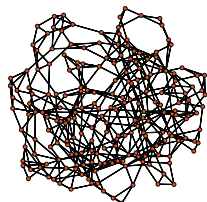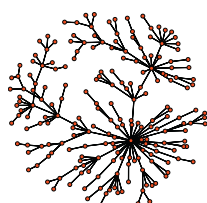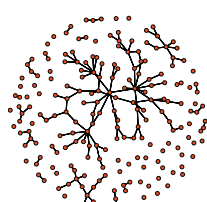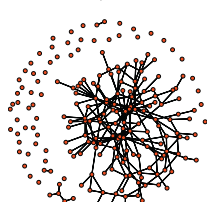
5. Future Extensions

5.1. Hypothesis testing

We have presented SGOF as a goodness of fit statistic, analogous to R^2. Using spectral distances, it is also possible to construct one and two-sample hypothesis tests for the purposes of formal rejection of certain models in favor of others. Space does not permit a full discussion of how such tests would be constructed; however, the authors will present this material in a separate manuscript.

5.2. Directed graphs

While the properties of the Laplacian spectrum of undirected graphs have been widely studied and applied, the spectral properties of directed graphs are less well-established. The present paper has therefore focused on undirected, possibly weighted, networks

Table 4

Faux Mesa High part 2: Comparing an ERGM to algorithmic models of structure.

	SGOF	AIC	Simulated network
Observed network			
GWD and homophily	0.494 (0.236, 0.696)	1853.849	
Watts–Strogatz model	−1.323 (−1.374, −1.273)	?	
Preferential attachment (default algorithm)	−1.124 (−2.572, −0.171)	?	
Preferential attachment (hand-tuned)	0.463 (0.132, 0.67)	?	
Configuration model	0.77 (0.722, 0.809)	?	

Notes: all ERGMs include the terms specified above, plus an "edges" parameter. A '?' indicates that the AIC is unknown because the likelihood is impractical to compute.

to establish the SGOF, but further work should consider the different properties of directed graphs. For now, we limit ourselves to the following remarks.

The Laplacian matrix for directed networks has been defined differently from that of undirected networks. In particular, Chung (2005) defines the Laplacian of directed networks as follows. First, given adjacency matrix, A, calculate a matrix, P, such that

$$P(i, j) = \frac{A_{ij}}{\sum_k A_{ik}}. \tag{7}$$

Then, treating P as the transition matrix of a Markov chain, calculate the Perron vector, ϕ, which is the all-positive left eigenvector of P

corresponding to the stationary distribution of the Markov chain (for strongly connected graphs). Define Φ as the matrix with ϕ on the diagonal and zeros elsewhere, and I in the standard way as the identity matrix. Finally, the Laplacian for directed graphs is defined as

$$L = I - \frac{\Phi^{1/2} P \Phi^{-1/2} \Phi^{-1/2} P^T \Phi^{1/2}}{2}. \tag{8}$$

One feature of this definition is that L is undirected and therefore has real-valued eigenvalues. Future work should consider the properties of this matrix from the point of view of goodness of fit, but also consider alternative transformations of the adjacency matrix for spectral analysis.
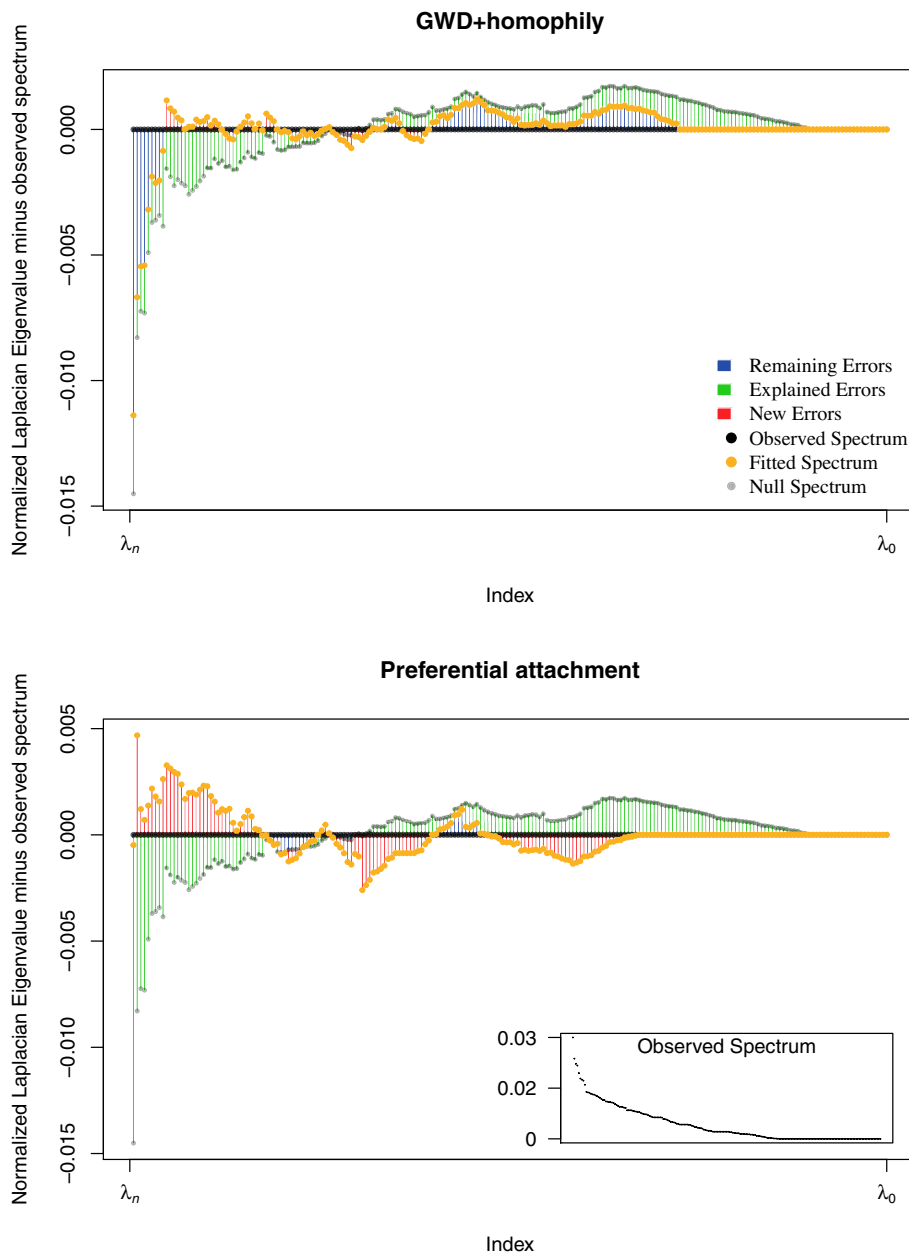
Fig. 1. Illustration of spectral qualities of two models of the Faux Mesa High data. The green and red indicate improvements and worsening of model fit, respectively, by moving from the null to the fitted model. Blue indicates error left unexplained in both the fitted and null models. The observed spectrum (shown inset) is subtracted from the plotted values, emphasizing the spectral errors. The preferential attachment model explains more error than the ERGM (represented by more green area in its visualization), but also introduces structure not present in the observed network, producing more new errors (red area) and resulting in a lower net SGOF. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

5.3. Nodal attributes and fit to labeled aspects of graph structure

The desirable properties that the SGOF exhibits are a result of its reliance on the graph Laplacian spectrum. However, this also introduces an important limitation. SGOF can only assess goodness of fit to unlabeled aspects of network structure. To illustrate, consider an observed network characterized by homophily such that there are two classes of nodes and a clear block structure comprising many within-class ties and few between-class ties. A well-fitting network model with a homophily parameter would produce simulated networks with similar block structure to that of the observed network, such that the blocks were determined by node class. In contrast, one could imagine a model that reproduced the block sizes equally well, but produced blocks that mixed both types of nodes. Since node attributes are labelled aspects of graph structure, the

SGOF would not be able to distinguish between these two models. The SGOF only considers the permutation invariant pattern of ties.

Structural statistic- or likelihood-based tools, are therefore recommended for examining goodness of fit to labeled graph properties. However, if such tools are not available or insufficient for some reason, the SGOF could be adapted in certain cases to consider nodal attributes. To use SGOF in the case of the block-model just described, in addition to the standard SGOF one could calculate an SGOF on three additional matrices: the matrix consisting of only those edges connecting class A nodes, the matrix consisting of only those edges connecting class B nodes, and the matrix consisting of only those edges connecting a class A node to a class B node. The SGOF from all three of these modified matrices will only be high if the fitted model accurately reproduces the block structure as it correlates with the nodal class attribute. The utility of this modified
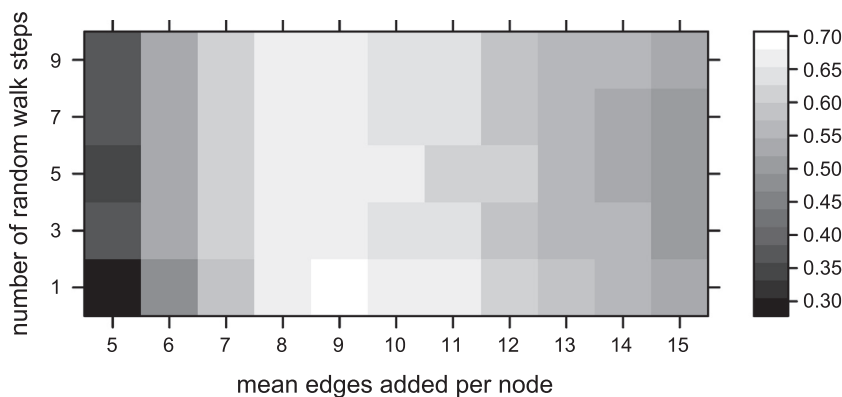
Fig. 2. SGOF for different combinations of parameter values for an algorithm based on Saramäki and Kaski (2004) fitted to the network of Jazz collaborations described in Gleiser and Danon (2003).

application of SGOF is likely to be limited to relatively simple cases with few attributes of interest, but future research may find a more flexible approach to incorporating nodal attributes.

5.4. Statistical properties of Laplacian eigenvalues

Under certain density conditions, the distribution of eigenvalues of the null model follows the 'semi-circle law' (Wigner, 1955; Chung et al., 2003), but these conditions are restrictive enough that we have chosen to calculate the null errors in the SGOF by simulation rather than by reference to the semi-circle law. The statistical properties (e.g. consistency and efficiency) of the eigenvalues of ensembles of networks other than the null model depend on the details of the model from which they are generated. As with the null model, the distribution of eigenvalues from certain narrowly defined models have been studied (Farkas et al., 2001; Bolla, 2004; Zhang et al., 2014). However, it is not clear from the present body of research, what can be said about the statistical properties of these spectra and thus of SGOF in the general case.

Since we cannot derive the statistical properties of the SGOF analytically, in order to provide one practical point of reference, we have conducted a simulation-based exploration of the properties of 100-node density-only models, under a range of densities. These simulations support the following tentative conclusions. The means of individual eigenvalues are stable across sample sizes (where sample size refers to the number of simulated networks from which the mean spectrum is calculated). However, as is typically the case, the standard error of the mean appears to be biased downwards for small samples. Likewise, the standard deviations and the 5th and 95th quantiles of individual eigenvalues from Erdős-Rényi random graphs are asymptotically consistent, but biased downwards for small numbers of simulated networks. As we mentioned above, at least 1000 simulations should be obtained for published results.

We strongly recommend examining the distribution of spectra simulated from fitted models to establish that sufficient sample sizes have been obtained when calculating the SGOF. Future work should seek to derive more general conclusions about the statistical properties of spectral distances for network models.

5.5. Co-spectral graphs

One possible limitation to the usefulness of SGOF is introduced by the phenomenon of co-spectral graphs, or non-isomorphic graphs with the same spectrum (Godsil and McKay, 1982; Harary et al., 1971). Enumerations of co-spectral graphs (Brouwer and Spence, 2009; Cvetković, 2012; Haemers and Spence, 2004) indicate that co-spectral graphs are most common among networks

of nine nodes and decrease in frequency as the number of nodes and edges increase. Given that most interesting networks are much larger than nine nodes, and given the overall rarity of graphs with co-spectral partners, we do not believe there is any a priori reason to believe that the utility of SGOF is substantially limited by this phenomenon, but future work should confirm this.

5.6. Adjacency spectra

We have relied on the spectrum of the graph Laplacian to implement SGOF, as it has a long history of being used as a discriminating signature for complex phenomena. It is also relatively easy to interpret in terms of network structure as we discussed in Section 2.2. That said, future work may reveal circumstances in which the adjacency spectrum is a more discriminating tool for goodness of fit. If this occurs, SGOF could be combined with an adjacency spectrum-based goodness of fit measure.

6. Conclusion

We have proposed a new measure of goodness of fit for network models based on the spectrum of the graph Laplacian: "spectral goodness of fit" (SGOF), and provided code with which SGOF can be easily implemented. The properties of SGOF fill gaps left by the current set of goodness of fit indicators, making it complementary to existing methods.

Unlike AIC, SGOF is always straightforward to calculate, does not depend on model form, and allows comparison of non-nested models as well as comparison between statistical and algorithmic models of network structure.

Hypothesis tests based on structural statistics also have these good properties, but it is hard to decide between models that are all rejected or all accepted by the test, and they require the researcher to know which statistics are important for a given network.

SGOF does not have these drawbacks. The scale of SGOF is also easier to interpret than an AIC for network models. Analogous to the standard R^2, the SGOF statistic measures the percent improvement in network structure explained over a null model. For fitted models that are superior to the null model (usually an Erdős-Rényi random graph), SGOF lies between 0 and 1.

In sum, SGOF is complementary to prior methods of assessing goodness of fit. It fills major gaps left by other techniques, but it does not dominate them in that it does have certain limitations of its own. Most of all, we hope that SGOF's strength of easily comparing goodness of fit among dissimilar models will facilitate the use and refinement of network models and results from outside of any given researcher's own methodological tradition.

References

Anderson, C.J., Wasserman, S., Faust, K., 1992. Building stochastic blockmodels. Soc. Netw. 14 (1), 137–161.

Barabási, A.-L., Albert, R., 1999. Emergence of scaling in random networks. Science 286 (5439), 509–512.

Belkin, M., Niyogi, P., 2003. Laplacian eigenmaps for dimensionality reduction and data representation. Neural Comput. 15 (6), 1373–1396.

Bolla, M., 2004. Distribution of the eigenvalues of random block-matrices. Linear Algebra Appl. 377, 219–240.

Bollobás, B., Nikiforov, V., 2004. Graphs and Hermitian matrices: eigenvalue interlacing. Discret. Math. 289 (1), 119–127.

Brouwer, A.E., Haemers, W.H., 2011. Spectra of Graphs. Springer.

Brouwer, A.E., Spence, E., 2009. Cospectral graphs on 12 vertices. Electron. J. Comb. 16 (20), 1.

Chung, F., 2005. Laplacians and the cheeger inequality for directed graphs. Ann. Comb. 9 (1), 1–19.

Chung, F., Lu, L., Vu, V., 2003. Spectra of random graphs with given expected degrees. Proc. Natl. Acad. Sci. U. S. A. 100 (11), 6313–6318.

Chung, F.R.K., 1997. Spectral Graph Theory, vol. 92. American Mathematical Soc.

Csardi, G., Nepusz, T., 2006. The igraph software package for complex network research. InterJournal, Complex Systems: 1695 200, 6, URL: http://igraph.sf.net

Cvetković, D., 2012. Spectral recognition of graphs. Yugoslav J. Oper. Res. 22 (2).

Cvetković, D.M., Rowlinson, P., Simić, S., 1997. Eigenspaces of Graphs, vol. 66. Cambridge University Press.

Donetti, L., Neri, F., Muñoz, M.A., 2006. Optimal network topologies: Expanders, cages, Ramanujan graphs, entangled networks and all that. J. Stat. Mech.: Theor. Exp. 2006 (08), P08007.

Doreian, P., Batagelj, V., Ferligoj, A., 2005. Structural analysis in the social sciences. Structural Analysis in the Social Sciences, vol. 25. Cambridge University Press.

Farkas, I.J., Derényi, I., Barabási, A.-L., Vicsek, T., 2001. Spectra of "real-world" graphs: Beyond the semicircle law. Phys. Rev. E 64 (2), 026704.

Fiedler, M., 1973. Algebraic connectivity of graphs. Czechoslov. Math. J. 23 (2), 298–305.

Frank, O., Strauss, D., 1986. Markov graphs. J. Am. Stat. Assoc. 81 (395), 832–842.

Gleiser, P.M., Danon, L., 2003. Community structure in jazz. Adv. Complex Syst. 6 (04), 565–573.

Godsil, C.D., McKay, B.D., 1982. Constructing cospectral graphs. Aequ. Math. 25 (1), 257–268.

Haemers, W.H., Spence, E., 2004. Enumeration of cospectral graphs. Eur. J. Comb. 25 (2), 199–211.

Handcock, M.S., 2003. Assessing degeneracy in statistical models of social networks. J. Am. Stat. Assoc., Citeseer.

Handcock, M.S., Hunter, D.R., Butts, C.T., Goodreau, S.M., Krivitsky, P.N., Morris, M., 2014. ergm: Fit, Simulate and Diagnose Exponential-Family Models for Networks. The Statnet Project (http://www.statnet.org) URL CRAN.R-project.org/package=ergm. R package version 3.1.2.

Harary, F., King, C., Mowshowitz, A., Read, R.C., 1971. Cospectral graphs and digraphs. Bull. Lond. Math. Soc. 3 (3), 321–328.

Harris, K.M., Udry, J.R., 2008. National Longitudinal Study of Adolescent Health (Add Health), 1994–2008. Carolina Population Center, University of North Carolina-Chapel Hill, Chapel Hill, NC/Inter-university Consortium for Political and Social Research, Ann Arbor, MI.

Holland, P.W., Leinhardt, S., 1981. An exponential family of probability distributions for directed graphs. J. Am. Stat. Assoc. 76 (373), 33–50.

Hunter, D.R., Handcock, M.S., 2006. Inference in curved exponential family models for networks. J. Comput. Graph. Stat. 15 (3).

Hunter, D.R., Goodreau, S.M., Handcock, M.S., 2008. Goodness of fit of social network models. J. Am. Stat. Assoc. 103 (481).

Hunter, D.R., Handcock, M.S., Butts, C.T., Goodreau, S.M., Morris, Martina, 2008. ergm: A package to fit, simulate and diagnose exponential-family models for networks. J. Stat. Softw. 24 (3), 1–29.

Krivitsky, P.N., 2012. Exponential-family random graph models for valued networks. Electron. J. Stat. 6, 1100.

Mohar, B., Alavi, Y., 1991. The Laplacian spectrum of graphs. Graph Theor. Comb. Appl. 2, 871–898.

Newman, M.E.J., 2006. Modularity and community structure in networks. Proc. Natl. Acad. Sci. U. S. A. 103 (23), 8577–8582.

Newman, M.E.J., Girvan, M., 2004. Finding and evaluating community structure in networks. Phys. Rev. E 69 (2), 026113.

Newman, M.E.J., Strogatz, S.H., Watts, D.J., 2001. Random graphs with arbitrary degree distributions and their applications. Phys. Rev. E 64 (2), 026118.

Pothen, A., Simon, H.D., Liou, K.-P., 1990. Partitioning sparse matrices with eigenvectors of graphs. SIAM J. Matrix Anal. Appl. 11 (3), 430–452.

Saramäki, J., Kaski, K., 2004. Scale-free networks generated by random walkers. Physica A 341, 80–86.

Schur, I., 1923. Uber eine klasse von mittelbildungen mit anwendungen auf die determinantentheorie. Sitzungsberichte der Berliner Mathematischen Gesellschaft 22, 9–20.

Schweinberger, M., 2012. Statistical modelling of network panel data: Goodness of fit. Br. J. Math. Stat. Psychol. 65 (2), 263–281.

Schweinberger, M., Snijders, T.A.B., 2003. Settings in social networks: a measurement model. Sociol. Methodol. 33 (1), 307–341.

Shen-Orr, S.S., Milo, R., Mangan, S., Alon, U., 2002. Network motifs in the transcriptional regulation network of Escherichia coli. Nat. Genet. 31 (1), 64–68.

Snijders, T.A.B., 2001. The statistical evaluation of social network dynamics. Sociol. Methodol. 31 (1), 361–395.

Steglich, C., Snijders, T.A.B., Pearson, M., 2010. Dynamic networks and behavior: Separating selection from influence. Sociol. Methodol. 40 (1), 329–393.

Turk, M., Pentland, A., 1991. Eigenfaces for recognition. J. Cogn. Neurosci. 3 (1), 71–86.

Vázquez, A., 2003. Growing network with local rules: Preferential attachment, clustering hierarchy, and degree correlations. Phys. Rev. E 67 (5), 056104.

Watts, D.J., Strogatz, S.H., 1998. Collective dynamics of 'small-world' networks. Nature 393 (6684), 440–442.

Wigner, E.P., 1955. Characteristic vectors of bordered matrices with infinite dimensions. Ann. Math. 54, 8–564.

Zhang, X., Rao Nadakuditi, R., Newman, M.E.J., 2014. Spectra of random graphs with community structure and arbitrary degrees. Phys. Rev. E 89 (4), 042816.