Goodness of Fit of Social Network Models

David R. HUNTER, Steven M. GOODREAU, and Mark S. HANDCOCK

We present a systematic examination of a real network data set using maximum likelihood estimation for exponential random graph models as well as new procedures to evaluate how well the models fit the observed networks. These procedures compare structural statistics of the observed network with the corresponding statistics on networks simulated from the fitted model. We apply this approach to the study of friendship relations among high school students from the National Longitudinal Study of Adolescent Health (AddHealth). We focus primarily on one particular network of 205 nodes, although we also demonstrate that this method may be applied to the largest network in the AddHealth study, with 2,209 nodes. We argue that several well-studied models in the networks literature do not fit these data well and demonstrate that the fit improves dramatically when the models include the recently developed geometrically weighted edgewise shared partner, geometrically weighted dyadic shared partner, and geometrically weighted degree network statistics. We conclude that these models capture aspects of the social structure of adolescent friendship relations not represented by previous models.

KEY WORDS: Degeneracy; Exponential random graph model; Markov chain Monte Carlo; Maximum likelihood estimation; *p*-star model.

1. INTRODUCTION

Among the many statistical methods developed in recent decades for analyzing dependent data, network models are especially useful for dealing with the kinds of dependence induced by social relations. Applications of social network models are becoming important in a number of fields, such as epidemiology, with the emergence of infectious diseases like AIDS and SARS; business, with the study of "viral marketing"; and political science, with the study of coalition formation dynamics. Much recent effort has been focused on inference for social network models (e.g., Holland and Leinhardt 1981; Strauss and Ikeda 1990; Snijders 2002; Hunter and Handcock 2006), but comparatively little work has tested the goodness of fit of the models.

Data on social relationships often can be represented as a network, or mathematical graph, consisting of a set of nodes and a set of edges, where an edge is an ordered or unordered pair of nodes. This article focuses specifically on network data collected at a nationally representative sample of high schools in the United States. The nodes represent students, and the edges signify friendships between pairs of students. Figure 1 depicts one such network graphically, where the shapes and labels of the nodes represent covariates measured on the students.

We consider exponential family models, in the traditional statistical sense, for network structure. These models have a long history in the networks literature, and we refer to them here as exponential random graph models (ERGMs). The primary contribution of this article is to propose a systematic approach to the assessment of network ERGMs. The models that we examine here achieve a good fit to key structural properties of the network with a small number of covariates. The approach and the findings address a central question in the network literature: Can the global structural features observed in a network be generated by a modest number of local rules?

Another contribution of this article is to demonstrate the use of maximum likelihood to fit reasonable models to network data with hundreds of nodes and obtain results that

are scientifically meaningful and interesting. We have developed an R package (called statnet) to implement the procedures developed in this article; this package is available at *http://csde.washington.edu/statnet*.

It is possible to simulate random networks from a given ERGM—at least in principle—using well-established Markov chain Monte Carlo (MCMC) techniques. More recently, various researchers have been developing techniques to solve a harder problem: calculating approximate maximum likelihood estimates (MLEs) of the ERGM parameters, given an observed network. Although these techniques are conceptually simple (Geyer and Thompson 1992), their practical implementation for relatively large social networks has proven elusive. We are now able to apply these techniques to networks encompassing thousands of nodes, much larger problems than could be tackled until very recently.

In problems for which maximum likelihood estimation previously has been possible in ERGMs, a troubling empirical fact has emerged: When ERGM parameters are estimated and a large number of networks are simulated from the resulting model, these networks frequently bear little resemblance to the observed network (Handcock 2003). This seemingly paradoxical fact arises because even though the MLE makes the probability of the observed network as large as possible, this probability still might be extremely small relative to other networks. In such a case, the ERGM does not fit the data well.

The remainder of this article provides a case study illustrating the application of recently developed models, software, and goodness-of-fit procedures to network data sets from the National Longitudinal Study of Adolescent Health (AddHealth), which is described in Section 2. Section 3 explains the statistical models that we fit to these data. Section 4 illustrates our goodness-of-fit technique on two simple models that do not fit well. Section 5 explains a set of network statistics that are used to build good-fitting models in Section 6.

2. INTRODUCTION TO THE ADDHEALTH SURVEY

The network data on friendships that we study in this article were collected during the first wave (1994–1995) of AddHealth. The AddHealth data come from a stratified sample of schools in the United States containing students in grades 7–12.

David R. Hunter is Associate Professor, Department of Statistics, Pennsylvania University, University Park, PA 16802 (E-mail: *dhunter@stat.psu.edu*). Steven M. Goodreau is Assistant Professor, Department of Anthropology (E-mail: *goodreau@uwashington.edu*), and Mark S. Handcock is Professor, Department of Statistics and Sociology (E-mail: *handcock@uwashington.edu*), University of Washington, Seattle, WA 98195. The authors thank Martina Morris for numerous helpful suggestions. This research was supported by National Institute on Drug Abuse grant DA012831 and National Institute of Child Health and Human Development grant HD041877.
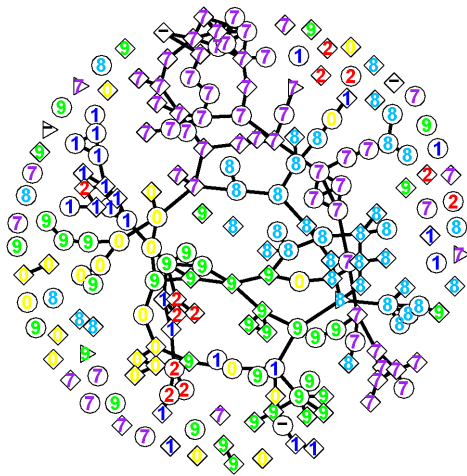
Figure 1. Mutual friendships represented as a network. Shapes of nodes denote sex: circles for female, squares for male, and triangles for unknown. Labels denote the units digit of grade (7–12), or "−" for unknown.

To collect friendship network data, AddHealth staff constructed a roster of all students in a school from school administrators. Students were then provided with the roster and asked to select up to five close male friends and five close female friends. Students were allowed to nominate friends who were outside the school or not on the roster, or to stop before nominating five friends of either sex. Complete details of this and subsequent waves of the study have been given by Resnick et al. (1997) and Udry and Bearman (1998) and can be found at *http://www.cpc.unc.edu/projects/addhealth*. In most cases, the individual school did not contain all grades 7–12; instead, data were collected from multiple schools within a single system (e.g., a junior high school and a high school) to obtain the full set of six grades. In these cases, we use the term "school" to refer to a set of schools from one community.

The full data set contains 86 schools, 90,118 student questionnaires, and 578,594 friendship nominations. Schools with large amounts of missing data, such as special education schools and for school districts that required explicit parental consent for student participation, were excluded from our analysis. Thus our analysis included 59 of the schools, ranging in size from 71 to 2,209 surveyed students. However, in this article we focus primarily on a single illustrative school, school 10, which has 205 students. Our results for school 10 may not necessarily be inferred to the whole population of schools; in particular, as we point out in Section 7, the parameter estimates for school 10 may be numerically quite different for those of other schools, because the parameters may depend on the number of nodes in a complicated way. Yet when we consider all 59 schools, we find remarkably similar qualitative results.

The edges in these raw network data are directed, because it is possible that A could name B as a friend without B naming A. However, in this article we consider the undirected network of *mutual* friendships, those in which both A nominates B and B nominates A. This feature of reciprocating nomination is common to many conceptualizations of friendship.

Each network may be represented by a symmetric $n \times n$ matrix $\mathbf{Y}$ and an $n \times q$ matrix $\mathbf{X}$ of nodal covariates, where n is the number of nodes. The entries of the $\mathbf{Y}$ matrix, termed the *adjacency matrix*, are all 0's and 1's, with $Y_{ij} = 1$ indicating the presence of an edge between i and j. Because self-nomination was disallowed, $Y_{ii} = 0$ for all i. The limit on the number of allowed nominations means that the data are not complete, but for convenience we assume that a lack of nomination in either direction between two individuals means that there is no mutual friendship.

The nodal covariate matrix $\mathbf{X}$ includes many measurements on each of the individuals in these networks. Some such measurements, like sex, are not influenced by network structure in any way and are termed *exogenous*. Other covariates may exhibit nonexogeneity: for example, tobacco use may be influenced through friendships. Exogeneity is important, for instance, to guarantee the dyadic independence property that we explain in (3). We focus our analysis on only three covariates: sex, grade, and race. Although the latter two may exhibit some endogeneity (e.g., the influence of friends may affect whether a student fails and must repeat a grade, or which race a student of mixed-race heritage chooses to identify with), we assume that such effects are minimal and consider the attributes to be fixed and exogenous. What we term "race" is constructed from two questions on race and Hispanic origin, with Hispanic origin taking precedence. Thus our categories "Hispanic," "Black," "White," "Asian," "Native American," and "Other" are shorthand names for "Hispanic (all races)," "Black (non-Hispanic)," "White (non-Hispanic)," and so on. This coding follows standard practice in the social science literature.

3. EXPONENTIAL RANDOM GRAPH MODELS

Our overall goal in using ERGMs, also known as p-star models (Wasserman and Pattison 1996), is to model the random behavior of the adjacency matrix $\mathbf{Y}$ conditional on the covariate matrix $\mathbf{X}$. Given a user-defined p-vector $\mathbf{g}(\mathbf{Y}, \mathbf{X})$ of statistics and letting $\boldsymbol{\eta} \in \mathbb{R}^p$ denote the statistical parameter, these models form a canonical exponential family (Lehmann 1983),

$$P_{\boldsymbol{\eta}}(\mathbf{Y} = \mathbf{y}|\mathbf{X}) = \kappa^{-1} \exp\{\boldsymbol{\eta}^t \mathbf{g}(\mathbf{y}, \mathbf{X})\}, \qquad (1)$$

where the normalizing constant $\kappa \equiv \kappa(\boldsymbol{\eta})$ is defined by

$$\kappa = \sum_{\mathbf{w}} \exp\{\boldsymbol{\eta}^t \mathbf{g}(\mathbf{w}, \mathbf{X})\}, \qquad (2)$$

and the sum (2) is taken over the whole sample space of allowable networks $\mathbf{w}$. The objective in defining $\mathbf{g}(\mathbf{Y}, \mathbf{X})$ is to choose statistics that summarize the social structure of the network. The range of substantially motivated network statistics that might be included in the $\mathbf{g}(\mathbf{Y}, \mathbf{X})$ vector is vast (see Wasserman and Faust 1994 for the most comprehensive treatment of these statistics). We consider only a few key statistics here, chosen to represent friendship selection rules operating at a local level. The goal is to test whether these local rules can reproduce the global network patterns of clustering and geodesic distances (Morris 2003).

The development of estimation methods for ERGMs has not kept pace with the development of ERGMs themselves. To understand why this is so, consider the sum of (2). A sample space consisting of all possible undirected networks on n nodes contains $2^{n(n-1)/2}$ elements, an astronomically large number even for moderate n. Therefore, direct evaluation of the normalizing

constant κ in (2) is computationally infeasible for all but the smallest networks—except in certain special cases, such as the dyadic independence model of (3)—and inference using maximum likelihood estimation is extremely difficult. To circumvent this difficulty, we use a technique called MCMC maximum likelihood estimation, in which a stochastic approximation to the likelihood function is built and then maximized (Geyer and Thompson 1992). This and other methods have been considered by Dahmström and Dahmström (1993), Corander, Dahmström, and Dahmström (1998), Crouch, Wasserman, and Trachtenberg (1998), Snijders (2002), and Handcock (2002). Details of the specific technique that we use have been given by Hunter and Handcock (2006), whereas the background of ERGMs in the networks literature has been discussed by Snijders (2002) or Hunter and Handcock (2006).

An important special case of model (1) is the *dyadic independence* model, in which

$$\mathbf{g}(\mathbf{y}, \mathbf{X}) = \sum_{i<j} y_{ij} \mathbf{h}(\mathbf{X}_i, \mathbf{X}_j) \qquad (3)$$

for some function $\mathbf{h}$ mapping $\mathbb{R}^q \times \mathbb{R}^q$ into $\mathbb{R}^p$, where the q-dimensional row vectors $\mathbf{X}_i$ and $\mathbf{X}_j$ are the nodal covariate vectors for the ith and jth individuals. In the context of an undirected network, the word *dyad* refers to a single Y_{ij} for some pair (i, j) of nodes (not to be confused with an *edge*, which requires that $Y_{ij} = 1$). In the ERGM resulting from (3), (1) becomes

$$P_{\boldsymbol{\eta}}(\mathbf{Y} = \mathbf{y}|\mathbf{X}) = \kappa^{-1} \prod_{i<j} \prod \exp\{y_{ij} \boldsymbol{\eta}^t \Delta(\mathbf{g}(\mathbf{y}, \mathbf{X}))_{ij}\}, \quad (4)$$

where

$$\Delta(\mathbf{g}(\mathbf{y}, \mathbf{X}))_{ij} = \mathbf{g}(\mathbf{y}, \mathbf{X})|_{y_{ij}=1} - \mathbf{g}(\mathbf{y}, \mathbf{X})|_{y_{ij}=0} \qquad (5)$$

denotes the change in the vector of statistics when y_{ij} is changed from 0 to 1 and the rest of $\mathbf{y}$ remains unchanged. In (4), the joint distribution of the Y_{ij} is simply the product of the marginal distributions—hence the name "dyadic independence model." The MLE in such a model may be obtained using logistic regression. As the simplest example of a dyadic independence model, we take $p = 1$ and $h(\mathbf{X}_i, \mathbf{X}_j) = 1$, which yields the well-known Bernoulli network, also known as the Erdős–Rényi network, in which each dyad is an edge with probability $\exp\{\eta\}/(1 + \exp\{\eta\})$.

For dyadic dependence models, (4) is not generally true, but nonetheless the right side of this equation is called the *pseudolikelihood*. Until recently, inference for social network models has relied on *maximum pseudolikelihood estimation*, which may be implemented using a standard logistic regression algorithm (Besag 1974; Frank and Strauss 1986; Strauss and Ikeda 1990; Geyer and Thompson 1992). However, it has been argued that maximum pseudolikelihood estimation can perform very badly in practice (Geyer and Thompson 1992) and that its theoretical properties are poorly understood (Handcock 2003). Particularly dangerous is the practice of interpreting standard errors from logistic regression output as though they are reasonable estimates of the standard deviations of the pseudolikelihood estimators. The only estimation technique that we discuss throughout the rest of this article is maximum likelihood estimation.

4. GOODNESS OF FIT FOR DYADIC INDEPENDENCE MODELS

The first dyadic independence model that we consider is perhaps the simplest possible network model, in which $\mathbf{g}(\mathbf{y}, \mathbf{X})$ consists only of $E(\mathbf{y})$, the number of edges in $\mathbf{y}$. This is the Bernoulli, or Erdős–Rényi, network described in Section 3. For AddHealth school 10, the parameter estimate for the Bernoulli network is seen in Table 1 to be -4.625. This may be derived exactly. Because school 10 has 205 nodes and 203 edges, the MLE for the probability that any dyad has an edge is $203/\binom{205}{2}$, or .00971, and the log-odds of this value is -4.625.

The second model that we consider includes edges and also several statistics based on nodal covariates. All of these statistics may be expressed as dyadic independence statistics as in (3); that is, they all are of the form

$$\sum_{i<j} \sum y_{ij} h(\mathbf{X}_i, \mathbf{X}_j) \qquad (6)$$

for a suitably chosen function $h(\mathbf{X}_i, \mathbf{X}_j)$.

First, we include the so-called *nodal factor effects* for each of the factors grade, race, and sex. Given a particular level of a particular factor (i.e., categorical variable), the nodal factor effect counts the total number of endpoints with that level for each edge in the network. In other words,

$$h(\mathbf{X}_i, \mathbf{X}_j) = \begin{cases} 2 & \text{if both nodes } i \text{ and } j \text{ have} \\ & \quad \text{the specified factor level} \\ 1 & \text{if exactly one of } i, j \text{ has the} \\ & \quad \text{specified factor level} \\ 0 & \text{if neither } i \text{ nor } j \text{ has the} \\ & \quad \text{specified factor level.} \end{cases} \qquad (7)$$

This means that the corresponding parameter is the change in conditional log-odds when we add an edge with one endpoint having this factor level—and this change is doubled when both endpoints of the edge share this level. As an example, consider the grade factor, which has levels 7–12 along with one missing-value level, *NA*. These seven levels of grade factor require six separate statistics for the nodal factor effect; one level must be excluded because the sum of all seven equals twice the number of edges in the network, thus creating a linear dependency among the statistics.

The second type of nodal statistics that we use is the *homophily statistics*. A homophily statistic for a particular factor gives each edge in the network a score or 0 or 1, depending on whether the two endpoints have matching values of the factor. We distinguish between two kinds of homophily, depending on whether the distinct levels of the factor should exhibit different homophily effects. Thus, for *uniform* homophily, we have a single statistic, defined by

$$h(\mathbf{X}_i, \mathbf{X}_j) = \begin{cases} 1 & \text{if } i \text{ and } j \text{ have the same level of the factor} \\ 0 & \text{otherwise.} \end{cases}$$

On the other hand, for *differential* homophily, we have a set of statistics, one for each level of the factor, where each is defined by

$$h(\mathbf{X}_i, \mathbf{X}_j) = \begin{cases} 1 & \text{if } i \text{ and } j \text{ both have the} \\ & \quad \text{specified factor level} \\ 0 & \text{otherwise.} \end{cases} \qquad (8)$$

Table 1. Estimated coefficients and standard errors for the parameters of three simple models that consider only network structure but no nodal covariate information

	Model			
Coefficient	Edges only	Edges plus GWESP	Edges plus GWDSP	Edges plus GWD
Edges	$-4.625_{(.07)}{}^{*}$	$-5.280_{(.10)}{}^{*}$	$-4.812_{(.197)}{}^{*}$	$-1.423_{(.50)}{}^{**}$
GWESP		$1.544_{(.10)}{}^{*}$		
GWDSP			$.047_{(.046)}$	
GWD				$-1.305_{(.20)}{}^{*}$

NOTE: The GWESP statistic $v(\mathbf{y}; \tau)$, the GWDSP statistic $w(\mathbf{y}; \tau)$, and the GWD statistic $u(\mathbf{y}; \tau)$ all use $\tau = .25$.

*Significant at the .001 level.

**Significant at the .01 level.

Note that for sex (a two-level factor), we may include a differential homophily effect or a nodal factor effect, but not both. This is because in an undirected network, there are only three types of edges—male–male, female–female, and male–female—so only two statistics are required to completely characterize the sexes of both endpoints of an edge, provided that the overall edge effect is also in the model. A differential homophily effect (two statistics) plus a nodal factor effect (one statistic) would entail redundant information.

In addition to the nodal factor and homophily effects, one final set of terms in our second dyadic independence model (summarized as model I in Table 2) involves the grade factor. This is an ordinal categorical variable, and we may expect the propensity to form friendships to depend on the difference between two individuals' grade values (e.g., seventh graders may be more likely to form friendships with eighth graders than with twelfth graders). Although we could add a new model term for

each possible pairing of two grade levels, a far more parsimonious model considers only the absolute difference of grade values,

$$h(\mathbf{X}_i, \mathbf{X}_j) = \begin{cases} 1 & \text{if } |\text{grade}_i - \text{grade}_j| = C \\ & \text{for some constant } C \\ 0 & \text{otherwise.} \end{cases} \quad (9)$$

In our models we added terms according to (9) for $C = 1$, $C = 2$, and $C = 3$. (We could not let $C = 0$, because this would introduce a linear dependence with the homophily statistics.) This has the effect of combining $C = 4$ and $C = 5$, along with any pairs for which grade is missing on one individual, into a single reference category.

Note that all schools have two sexes and six grades, but only some have additional NA categories for these factors. Furthermore, the number of races present varies considerably from school to school. Parameters are excluded from the model when it can be determined in advance that the MLE will be undefined.

Table 2. Estimated coefficients (and standard errors) for two models applied to AddHealth school 10

Coefficient	Model I	Model II	Coefficient	Model I	Model II
Edges	$-10.45_{(1.44)}{}^{***}$	$-3.49_{(1.92)}$	AD (grade) = 1	$3.71_{(1.26)}{}^{**}$	$3.41_{(1.42)}{}^{*}$
GWESP		$.83_{(.13)}{}^{***}$	AD (grade) = 2	$2.43_{(1.28)}$	$2.42_{(1.48)}$
GWD		$-2.01_{(.35)}{}^{***}$	AD (grade) = 3	$1.52_{(1.42)}$	$1.43_{(1.62)}$
GWDSP		$.50_{(.09)}{}^{***}$			
DH (grade 7)	$6.83_{(1.37)}{}^{***}$	$6.00_{(1.56)}{}^{***}$			
NF (grade 8)	$-.54_{(.73)}$	$-.34_{(.78)}$	DH (grade 8)	$7.81_{(1.56)}{}^{***}$	$6.48_{(1.64)}{}^{***}$
NF (grade 9)	$.72_{(.53)}$	$.64_{(.59)}$	DH (grade 9)	$5.04_{(1.39)}{}^{***}$	$4.52_{(1.58)}{}^{**}$
NF (grade 10)	$.59_{(.57)}$	$.55_{(.59)}$	DH (grade 10)	$5.76_{(1.44)}{}^{***}$	$4.96_{(1.59)}{}^{**}$
NF (grade 11)	$1.12_{(.53)}{}^{*}$	$.97_{(.60)}$	DH (grade 11)	$4.98_{(1.39)}{}^{***}$	$4.32_{(1.54)}{}^{**}$
NF (grade 12)	$1.51_{(.60)}{}^{*}$	$1.23_{(.60)}{}^{*}$	DH (grade 12)	$4.64_{(1.48)}{}^{**}$	$4.11_{(1.58)}{}^{**}$
NF (grade NA)	$4.05_{(1.14)}{}^{***}$	$3.86_{(1.30)}{}^{**}$			
			DH (White)	$1.58_{(.62)}{}^{*}$	$1.55_{(.68)}{}^{*}$
NF (Black)	$.45_{(.39)}$	$.51_{(.42)}$	DH (Black)	$1.18_{(1.27)}$	$.92_{(1.55)}$
NF (Hisp)	$-.419_{(.34)}$	$-.23_{(.33)}$	DH (Hisp)	$1.17_{(.41)}{}^{**}$	$.87_{(.43)}{}^{*}$
NF (Nat Am)	$-.460_{(.30)}$	$-.21_{(.32)}$	DH (Nat Am)	$1.70_{(.42)}{}^{***}$	$1.31_{(.43)}{}^{**}$
NF (Other)	$-1.14_{(.75)}$	$-.61_{(.69)}$			
NF (Race NA)	$1.23_{(.61)}{}^{*}$	$1.53_{(.89}$			
NF (Female)	$.09_{(.09)}$	$.09_{(.10)}$	UH (Sex)	$.78_{(.15)}{}^{***}$	$.67_{(.16)}{}^{***}$
NF (Sex NA)	$-.43_{(.47)}$	$-.18_{(.47)}$			

NOTE: NF, node factor; AD, absolute difference; DH, differential homophily; UH, uniform homophily. Model I contains terms for edges and the 25 nodal covariate terms described in Section 4. Model II contains all of the terms in model I plus three additional terms—GWESP, GWDSP, and GWD—each with $\tau = .25$. Differential homophily terms for Grade NA, Race Other, Race NA, and Sex NA are omitted because no edges are observed between two actors sharing these attribute values.

*Significant at the .05 level.

**Significant at the .01 level.

***Significant at the .001 level.

Such cases occur for node factor effects when only a small number of students have the factor level, and they all have 0 friendships, or for homophily terms when there are no ties between two students with a given factor level. For example, in AddHealth school 10, grade is a 7-level factor, sex is a 3-level factor, and race is a 4-level factor, and our dyadic independence model contains 25 parameters: 1 for edges, 6 for the grade factor effect, 6 for differential homophily on grade (excluding the NA category), 5 for the race factor effect, 4 for differential homophily on race (excluding the NA and Other categories), 2 for the sex factor effect, and 1 for uniform homophily on sex. The fitted values of these 25 parameters are presented as model I in Table 2.

Our graphical tests of goodness-of-fit require a comparison of certain observed network statistics with the values of these statistics for a large number of networks simulated according to the fitted ERGM. The choice of these statistics determines which structural aspects of the networks are important in assessing fit. We propose to consider three sets of statistics: the degree distribution, the edgewise shared partner distribution, and the geodesic distance distribution.

The degree distribution for a network consists of the values $D_0/n, \ldots, D_{n-1}/n$, where D_k/n equals the proportion of nodes that share edges with exactly k other nodes. The edgewise shared partner distribution consists of the values $EP_0/E, \ldots, EP_{n-2}/E$, where E denotes the total number of edges and EP_k equals the number of edges whose endpoints both share edges with exactly k other nodes. (The D_k and EP_k statistics are explained in much greater detail in Sec. 5.) Finally, the geodesic distance distribution consists of the relative frequencies of the possible values of geodesic distance between two nodes, where the geodesic distance between two nodes equals the length of the shortest path joining those two nodes (or infinity if there is no such path). For instance, because two nodes are at geodesic distance 1 if and only if they are connected by an edge, and because there are $\binom{n}{2}$ possible pairs of nodes, the first value of the geodesic distance distribution equals $E/\binom{n}{2}$. The last value, the fraction of dyads with infinite geodesics, is also called the fraction "unreachable."

We chose to include the degree statistics because of the tremendous amount of attention paid to them in the networks literature. We included the shared partner statistics based on the work of Snijders, Pattison, Robins, and Handcock (2006) and Hunter and Handcock (2006), and because, as we show in Section 6, the addition of a parametric formula involving $EP_0, \ldots, EP_{n-2}$ improves model fit dramatically. Therefore, these statistics appear to contain a great deal of relevant network information. Furthermore, (13) demonstrates that the triangle count, ubiquitous in the networks literature, is a function of the shared partner statistics. Finally, the geodesic distance statistics are the basis for two of the most common measures of centrality, a fundamental concept in social network theory (Wasserman and Faust 1994, p. 111) and clearly are relevant to the speed and robustness of diffusion across networks. They also represent higher-order network statistics not directly related to any of the statistics included in our models, and thus provide a strong independent criterion for goodness of fit.

Figure 2 depicts the results of 100 simulations for school 10 from the fitted dyadic independence models given in Tables 1

and 2. The vertical axis in each plot is the logit (log-odds) of the relative frequency, and the solid line represents the statistics for the observed network. We can immediately see that the models do an extremely poor job capturing the shared partner distribution. They perform relatively well for the degree distribution and the geodesics distribution, considering their simplicity. Adding the attribute-based statistics improves the fit of the geodesic distribution considerably. The lack of fit in the shared partner plot reflects the fact that the model strongly underestimates the amount of local clustering present in the data. The models predict friends to have no friends in common most of the time and one friend in common occasionally, whereas in the original data they have up to five friends in common. Although we present plots for only one school here, the qualitative results for other schools follow a small number of similar patterns.

In Sections 5 and 6 we present some modifications to the models seen here that fit much better as measured both by the graphical criterion that we have used here and by more traditional statistical measures, such as the Akaike information criterion (AIC). The fact that the simple dyadic independence models do not appear to fit the data well is not surprising; after all, such models are merely logistic regression models in which the responses are the dyads. That we must move beyond dyadic independence to construct models that fit social network data well is a result of the fact that the formation of edges in a network depends on the existing network structure itself.

5. DEGREE, SHARED PARTNER, AND OTHER NETWORK STATISTICS

A simplistic ERGM that is not a dyadic independence model is one in which $\mathbf{g}(\mathbf{y}, \mathbf{X})$ consists only of a subset of the degree statistics $D_k(\mathbf{y})$, $0 \le k \le n - 1$. The degree of a node in a network is the number of neighbors that it has, where a neighbor is a node with which it shares an edge. We define $D_k(\mathbf{y})$ to be the number of nodes in the network $\mathbf{y}$ that have degree k. Note that the $D_k(\mathbf{y})$ statistics satisfy the constraint $\sum_{i=0}^{n-1} D_i(\mathbf{y}) = n$, so we may not include all n degree statistics among the components of the vector $\mathbf{g}(\mathbf{y}, \mathbf{X})$; if we did, then the coefficients in model (1) would not be identifiable. A common reformulation of the degree statistics is given by the k-star statistics $S_1(\mathbf{y}), \ldots, S_{n-1}(\mathbf{y})$, where $S_k(\mathbf{y})$ is the number of k-stars in the network $\mathbf{y}$. A k-star (Frank and Strauss 1986) is an unordered set of k edges that all share a common node. For instance, "1-star" is synonymous with "edge." Because a node with i neighbors is the center of $\binom{i}{k}$ k-stars (but the "common node" of a 1-star may be considered arbitrarily to be either of two nodes), we see that

$$S_k(\mathbf{y}) = \sum_{i=k}^{n-1} \binom{i}{k} D_i(\mathbf{y}), \qquad 2 \le k \le n - 1, \qquad \text{and}$$

$$S_1(\mathbf{y}) = \frac{1}{2} \sum_{i=1}^{n-1} i D_i(\mathbf{y}). \tag{10}$$

Note that an edge is the same as a 1-star, so $E(y) = S_1(y)$. The k-star statistics are highly collinear with one another. For example, any 4-star automatically comprises four 3-stars, six 2-stars, and four 1-stars (or edges).
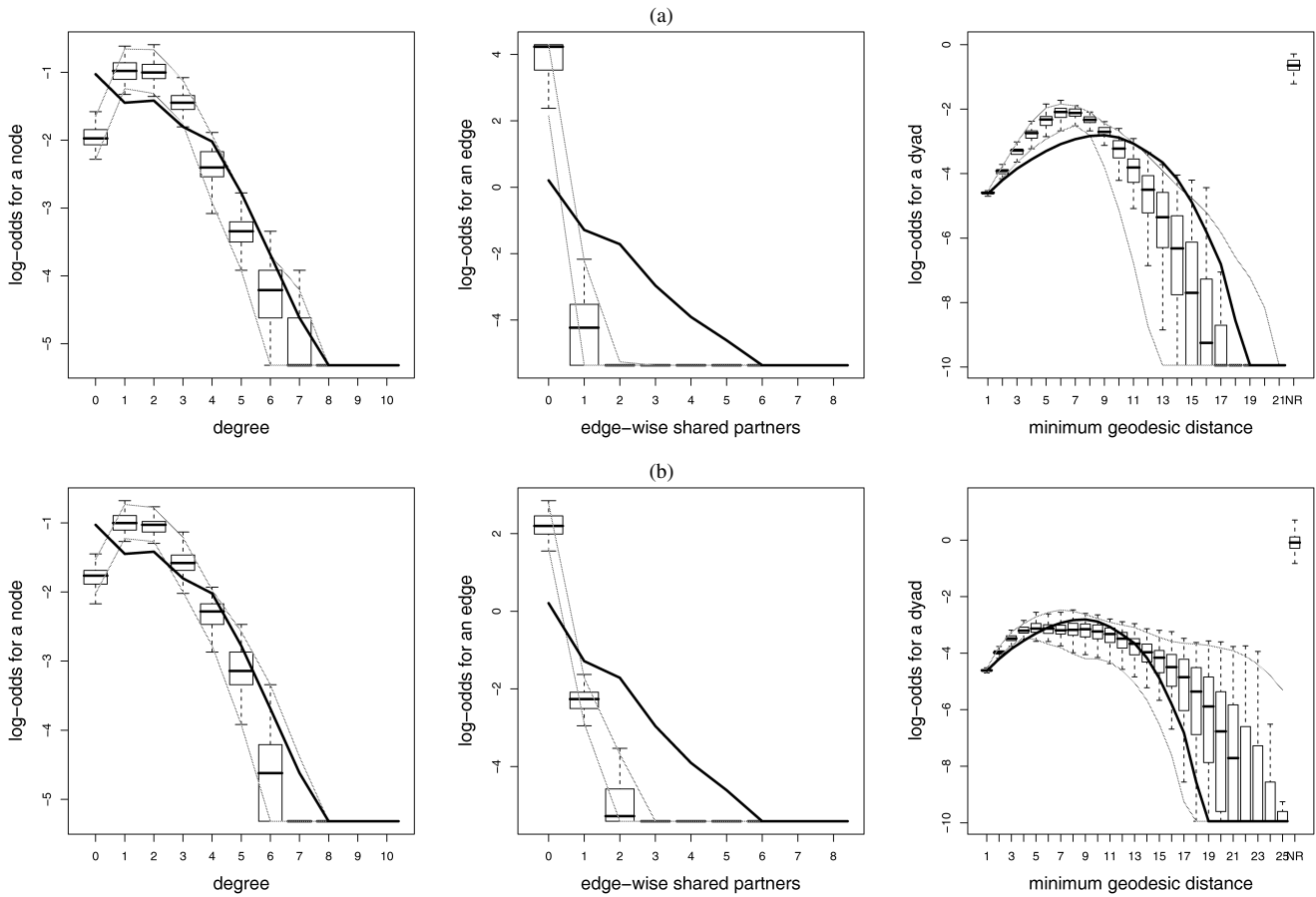
Figure 2. Simulation results for dyadic independence models. (a) School 10, edges only (Bernoulli or Erdős–Rényi model). (b) School 10, edges and covariates. In all plots, the vertical axis is the logit of relative frequency, the school 10 statistics are indicated by the solid lines, the boxplots include the median and interquartile range, and the light-gray lines represent the range in which 95% of simulated observations fall.

The shared partner statistics are another useful class of statistics. We define two distinct sets of shared partner statistics, the *edgewise* shared partner statistics and the *dyadwise* shared partner statistics. The edgewise shared partner statistics are denoted as $EP_0(\mathbf{y}), \ldots, EP_{n-2}(\mathbf{y})$, where $EP_k(\mathbf{y})$ is defined as the number of unordered pairs $\{i, j\}$ such that $y_{ij} = 1$ and i and j have exactly k common neighbors (Hunter and Handcock 2006). The requirement that $y_{ij} = 1$ distinguishes the edgewise shared partner statistics from the dyadwise shared partner statistics $DP_0(\mathbf{y}), \ldots, DP_{n-2}(\mathbf{y})$: We define $DP_k(\mathbf{y})$ to be the number of pairs $\{i, j\}$ such that i and j have exactly k common neighbors. In particular, it is always true that $DP_k(\mathbf{y}) \geq EP_k(\mathbf{y})$, and in fact $DP_k(\mathbf{y}) - EP_k(\mathbf{y})$ equals the number of unordered pairs $\{i, j\}$ for which $y_{ij} = 0$ and i and j share exactly k common neighbors.

Because there are $E(\mathbf{y})$ edges and $\binom{n}{2}$ dyads in the entire network, we obtain the identities

$$E(\mathbf{y}) = \sum_{i=0}^{n-2} EP_i(\mathbf{y}) \tag{11}$$

and

$$\binom{n}{2} = \sum_{i=0}^{n-2} DP_i(\mathbf{y}). \tag{12}$$

Furthermore, we can obtain the number of triangles in $\mathbf{y}$ by considering the edgewise shared partner statistics. Whenever $y_{ij} = 1$, the number of triangles that include this edge is exactly the number of common neighbors shared by i and j. Therefore, if we count all of the shared partners for all edges, then we have counted each triangle three times, once for each of its edges; in other words,

$$T(\mathbf{y}) = \frac{1}{3} \sum_{i=0}^{n-2} i EP_i(\mathbf{y}). \tag{13}$$

A related formula involving the dyadwise shared partner statistics is obtained by noting that each triangle automatically comprises three 2-stars; therefore, $S_2(\mathbf{y}) - 3T(\mathbf{y})$ is the number of 2-stars for which the third side of the triangle is missing. We conclude that

$$S_2(\mathbf{y}) - 3T(\mathbf{y}) = \sum_{i=0}^{n-2} i [DP_i(\mathbf{y}) - EP_i(\mathbf{y})]. \tag{14}$$

Combining (14) and (13) produces

$$S_2(\mathbf{y}) = \sum_{i=0}^{n-2} i DP_i(\mathbf{y}).$$

Because a 2-star is also a path of length 2, $S_2(\mathbf{y})$ is sometimes referred to as the twopath statistic.
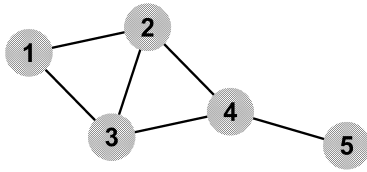
Figure 3. For this simple five-node network, the edgewise and dyadwise shared partner distributions are $(EP_0, \ldots, EP_3) = (1, 4, 1, 0)$ and $(DP_0, \ldots, DP_3) = (2, 6, 2, 0)$; the k-triangle and k-two-path distributions are $(T_1, T_2, T_3) = (2, 1, 0)$ and $(P_1, P_2, P_3) = (10, 1, 0)$.

Finally, we summarize two additional sets of statistics, due to Snijders et al. (2006), that we use in Section 6. First, the triangle statistic generalizes to the set of k-triangle statistics, where a k-triangle is defined to be a set of k distinct triangles that share a common edge. In particular, a 1-triangle is the same as a triangle. Second, the 2-star statistic (also known as the twopath statistic) generalizes to the set of k-twopath statistics, where a k-twopath is a set of k distinct 2-paths joining the same pair of nodes. In particular, a 1-twopath is the same as a 2-star or a 2-path. Snijders et al. (2006) actually coined the term "k-independent 2-path," but we simplify this to k-twopath in this article.

As a concrete example, we note that the simple network of Figure 3 contains 2 1-triangles, 1 2-triangle, 10 1-twopaths, and 1 2-twopath. (Note that the 2-twopath joining nodes 1 and 4 is the same as the 2-twopath joining nodes 2 and 3, although it is counted only once.) We denote the number of k-triangles and k-twopaths in the network $\mathbf{y}$ by $T_k(\mathbf{y})$ and $P_k(\mathbf{y})$. Just as the degree statistics $D_i(\mathbf{y})$ are related to the k-star statistics $S_k(\mathbf{y})$ by (10), the edgewise and dyadwise shared partner statistics are related to the k-triangle and k-twopath statistics by the equations

$$T_k(\mathbf{y}) = \sum_{i=k}^{n-2} \binom{i}{k} EP_i(\mathbf{y}), \qquad 2 \leq k \leq n-2,$$

and

$$P_k(\mathbf{y}) = \sum_{i=k}^{n-2} \binom{i}{k} DP_i(\mathbf{y}), \qquad 1 \leq k \leq n-2, k \neq 2.$$

The cases not covered here are those of $T_1(\mathbf{y})$, given in (13), and $P_2(\mathbf{y})$, the number of 4-cycles, which includes an extra factor of $1/2$, because any 4-cycle can be considered a 2-path between two distinct pairs of nodes,

$$P_2(\mathbf{y}) = \frac{1}{2} \sum_{i=2}^{n-2} \binom{i}{2} DP_i(\mathbf{y}).$$

6. GOODNESS OF FIT FOR DYADIC DEPENDENCE MODELS

A fundamental principle of social network analysis is that dependence among edges is a guiding force in the formation of networks; that is, the Y_{ij} that make up an adjacency matrix $\mathbf{Y}$ are not independent of one another. For instance, it seems reasonable that two potential edges should be correlated if they involve the same individual (so that Y_{ij} is not independent of Y_{ik} for distinct i, j, and k). Starting from this simple assumption, Frank and Strauss (1986) proposed the so-called Markov random graphs. They showed that an ERGM with a few simple network statistics, among them the number of triangles, suffices under this assumption. As originally proposed, these homogeneous Markov random graphs treat all nodes as equivalent, ignoring any covariate information.

But what if we have covariate information about the nodes? Because nodal attributes are clearly important in most, if not all, social networks, it is not surprising that homogeneous Markov random graph models fail to empirically describe social network data. But is it true that the Y_{ij} are dependent even when we condition on the observations X? We might take the approach suggested by Markov random graphs, adding covariate information of the sort discussed in Section 4 to Markov random graph models and then performing statistical tests to determine whether such terms as the number of triangles are statistically significant after the effects of covariates are accounted for. Unfortunately, this approach fails. The reasons for this failure are deep (see Handcock 2002, 2003), but in a nutshell, the models obtained by adding covariate information to Markov-inspired ERGMs simply do not fit network data well. Without a model that yields simulated networks resembling the observed network, maximum likelihood estimation itself is doomed. (For a discussion of maximum likelihood estimation algorithms, see Snijders 2002; Hunter and Handcock 2006.) This failure motivated the work of Snijders et al. (2006) in developing the alternating k-triangle, k-twopath, and k-star statistics that we explain in this section.

These new statistics finally allow us to build dependence models that fit network data sets well enough to enable reliable maximum likelihood estimation routines. Thus we can augment these models by adding covariate-only terms, then try to determine whether the dependence terms are statistically significant. If not, then independence models will suffice, so we may simply perform "network" analysis by ignoring the network structure altogether and performing logistic regression on the independent responses Y_{ij}. We can see (somewhat reassuringly from the standpoint of social networks research) that the dependence among dyads appears to persist even after nodal covariate information is taken into account.

Consider using the shared partner statistics and the degree statistics defined in Section 5 to build an ERGM. For instance, it is possible to add one new term to the model for each of the edgewise shared partner statistics $EP_1, \ldots, EP_{n-2}$—we omit EP_0 to avoid the linear dependence of (11)—but this typically leads to a model with excessive flexibility. As Hunter and Handcock (2006) pointed out, it is often better to restrict the parameter space to avoid problems of degeneracy. Toward this end, we define the statistics

$$u(\mathbf{y}; \tau) = e^\tau \sum_{i=1}^{n-2} \{1 - (1 - e^{-\tau})^i\} D_i(\mathbf{y}), \qquad (15)$$

$$v(\mathbf{y}; \tau) = e^\tau \sum_{i=1}^{n-2} \{1 - (1 - e^{-\tau})^i\} EP_i(\mathbf{y}), \qquad (16)$$

and

$$w(\mathbf{y}; \tau) = e^\tau \sum_{i=1}^{n-2} \{1 - (1 - e^{-\tau})^i\} DP_i(\mathbf{y}), \qquad (17)$$

where in each case τ is an additional parameter. We refer to these three statistics as geometrically weighted degree, edgewise shared partner, and dyadwise shared partner statistics.

Although the definitions of u, v, and w may appear somewhat unusual, we chose them to coincide with the alternating k-star, alternating k-triangle, and alternating k-twopath statistics of Snijders et al. (2006),

$$u(\mathbf{y}; \tau) = 2S_1(\mathbf{y}) - \frac{S_2(\mathbf{y})}{(e^\tau)^1} + \cdots + (-1)^n \frac{S_{n-1}(\mathbf{y})}{(e^\tau)^{n-2}}, \quad (18)$$

$$v(\mathbf{y}; \tau) = 3T_1(\mathbf{y}) - \frac{T_2(\mathbf{y})}{(e^\tau)^1} + \cdots + (-1)^{n-3} \frac{T_{n-2}(\mathbf{y})}{(e^\tau)^{n-3}}, \quad (19)$$

and

$$w(\mathbf{y}; \tau) = P_1(\mathbf{y}) - \frac{2P_2(\mathbf{y})}{(e^\tau)^1} + \cdots + (-1)^{n-3} \frac{P_{n-2}(\mathbf{y})}{(e^\tau)^{n-3}}. \quad (20)$$

Actually, the alternating k-star statistic in (18) is not identical to the alternating k-star statistic of Snijders et al. (2006), although it is in a certain sense equivalent from a modeling perspective. Hunter (2007) discussed this issue at length. As Snijders et al. (2006) explained, these three statistics appear to capture high-order dependency structure in networks in a parsimonious fashion while avoiding the problems of degeneracy described by Handcock (2002, 2003).

The τ parameters in (16), (17), and (15) are not canonical exponential family parameters like η in (1); rather, if $\tau = (\tau_1, \tau_2, \tau_3)$ is considered unknown and (η, τ) is the full parameter vector, then the ERGM forms a curved exponential family, which complicates the estimation procedure. Hunter and Handcock (2006) addressed this more complicated situation; however, for the purposes of this article, we make the simplifying assumption that each τ is fixed and known. In our model-fitting procedure, we tried a range of different values of τ on several schools and found that for each statistic, the goodness-of-fit plots (as in Fig. 4) were nearly indistinguishable for different values of τ in the range that we tested (.1–1.5). Values far outside this range resulted in models that could not be fit. Based on these results, we use a fixed value of $\tau = .25$ for all the models that we discuss.

As an example, we take $\mathbf{g}(\mathbf{y}, \mathbf{X})$ to consist of only two terms, the edge statistic and the geometrically weighted edgewise shared partner (GWESP) statistic. In this case the ERGM of (1) becomes

$$P_\eta(\mathbf{Y} = \mathbf{y} | \mathbf{X}) = \kappa^{-1} \exp\{\eta_1 E(\mathbf{y}) + \eta_2 v(\mathbf{y}; \tau)\}. \quad (21)$$

We fit model (21), as well as similar models using the geometrically weighted dyadwise shared partner (GWDSP) and geometrically weighted degree (GWD) statistics, to AddHealth school 10. The results are given in Table 1.

Many dyadic dependence models create such severe numerical difficulties in estimation (Handcock 2002, 2003) that we are unable to fit them for a large number of different networks of different sizes. Models with the GWESP, GWDSP, and GWD statistics appear to be more robust. Using our MCMC fitting procedure, we could estimate their parameters on many of the AddHealth schools, the first such application of maximum likelihood estimation to a dyadic dependence model for a range of

different-sized networks with hundreds of nodes. As a case in point, consider Figure 5, in which we successfully fit a dyadic dependence model to the largest school in the sample, with 2,209 nodes, and obtained reasonable parameter estimates. (We discuss this school further in Sec. 7.)

As described in Section 4, one way to develop an idea of how well a model fits is by comparing a set of observed network statistics with the range of the same statistics obtained by simulating many networks from the fitted ERGM. If the observed network is not typical of the simulated network for a particular statistic, then the model is either degenerate (if the statistic is among those included in the ERGM vector $\mathbf{g}[\mathbf{y}, \mathbf{X}]$) or poorly fitted (if the statistic is not included). Figure 4 depicts simulation results for school 10 for the three dyadic-dependent ERGMs in Table 1; Figure 6 depicts model II from Table 2.

For both school 10 and many of the other smaller AddHealth schools, a simple model containing only individual-level attributes [Fig. 2(b)] does a respectable job of recreating the geodesic distribution of the observed data, a global property of the network. At the same time, it strongly underestimates the amount of local clustering as captured by the shared partner distribution. The former observation is encouraging, because information on attribute matching is far easier to collect than other types of network data in most real-world settings where only a sample of nodes is available. Gathering such information requires questions only about the attributes of respondents' partners, not their actual identities. The latter observation tells us that not all features of the network can be ascribed to purely dyadic-level phenomena; this fact is not surprising, because it is the very basis for the field of network analysis. Finally, the fact that a simple model is strongly predictive of one higher-order network property (geodesics) and strongly divergent from another (shared partner) suggests that various network statistics should be tested to develop a robust sense of goodness of fit.

Comparing Figure 2(b) with Figures 4(a) and 4(b) shows that incorporating the heterogeneity of actors through nodal covariates was more important for model fit than either modeling degree or edgewise shared partners alone. This should not be too surprising; we expect that nodal covariates to be very important in predicting most types of social relationships, and certainly high school friendships are no exception.

Social relations generally exhibit local clustering, and in this case we observe that the simple Bernoulli model drastically underpredicts the number of shared partners people should have, even though it captures the degree distribution well. Such clustering can come from at least two different sources: actors matching on exogenous attributes and actors forming partnerships on the basis of existing shared partners. The two are fundamentally different; the former is dyadic-independent, using factors exogenous to the network structure; whereas the latter is dyadic-dependent and reflects the transitivity property that friends of my friends are more likely to be my friends. The modeling here shows that neither homophily nor shared partners alone is sufficient to explain the clustering observed in this friendship network. The same is true of other AddHealth schools (see the plots available at *http://csde.washington.edu/networks*). Indeed, Table 2 shows that the homophily effects are smaller in magnitude in model II (which includes the shared partner statistics) than in model I.
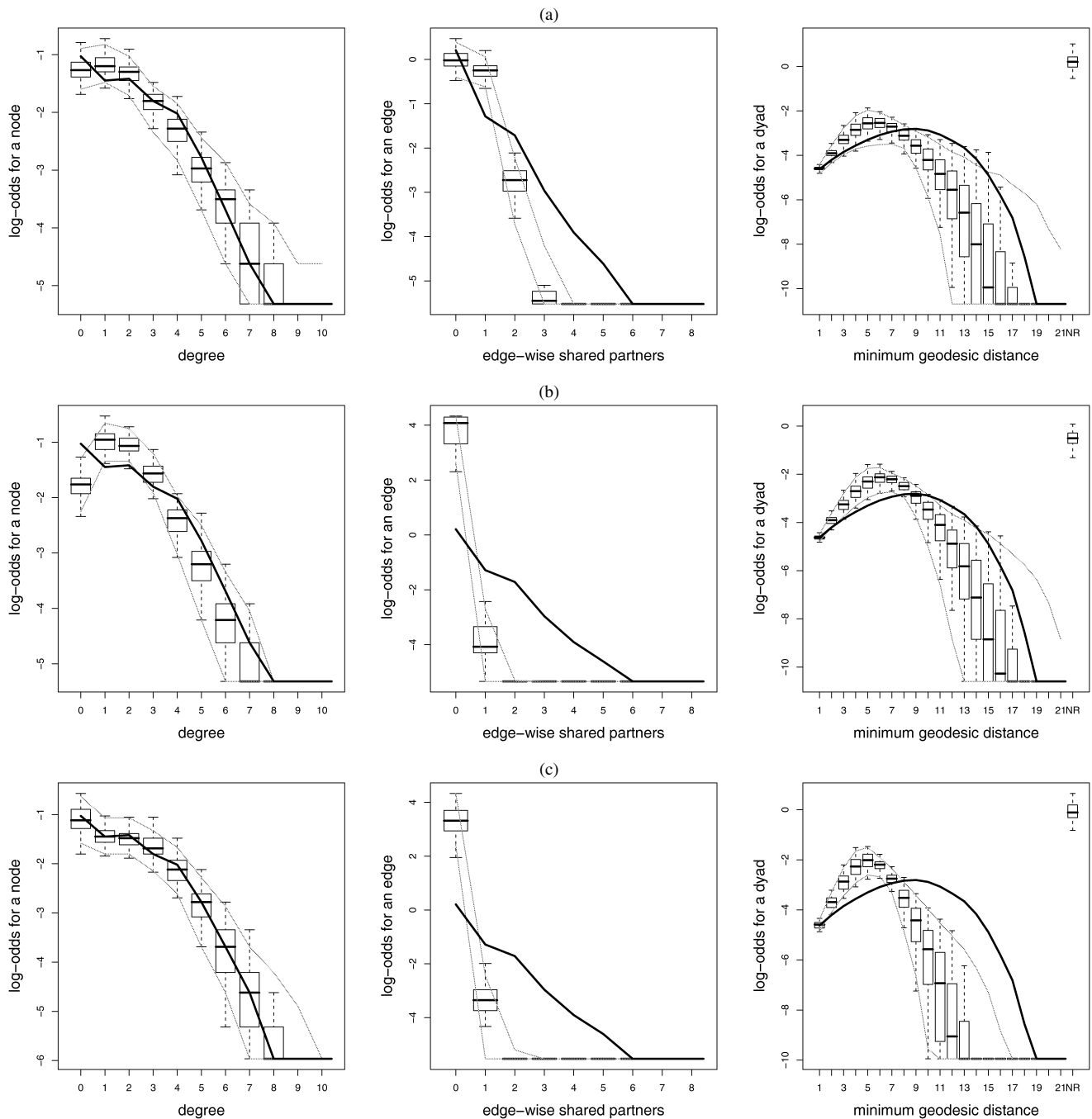
Figure 4. Simulation results for dyadic dependence ERGMs of Table 1. (a) School 10, edges and GWESP ($\tau = .25$). (b) School 10, edges and GWDSP ($\tau = .25$). (c) School 10, edges and GWD ($\tau = .25$).

In this setting a simple one-term Bernoulli model [Fig. 2(a)] turned out to fit the degree distribution fairly well, although it failed slightly in underestimating the number of high-degree nodes. Adding the degree distribution to the model in the form of the GWD statistic [Fig. 4(c)] appeared to remedy this defect. However, neither of these two models reproduced the clustering and geodesic distances observed in this network. In fact, comparing the rightmost plots in Figures 2(a) and 4(c) shows that adding the GWD statistic by itself appears to have worsened the fit to the geodesic distance distribution. We may explain this phenomenon by noting (in Fig. 1) that the students tend to form friendships preferentially within their own

grade, creating longer-than-expected geodesics between pairs of students in different grades. Adding only the GWD term to the model makes this problem worse, because it tends to create highly connected individuals (while still ignoring grade), thus shortening the geodesics even more. But whatever the cause of this phenomenon, the most important message here is that there are hazards inherent in focusing only on the degree distribution when trying to build realistic models for networks.

One might wonder how our graphical methods for assessing goodness of fit compare with more traditional methods, such as the AIC (Akaike 1973) or the Bayes information criterion BIC
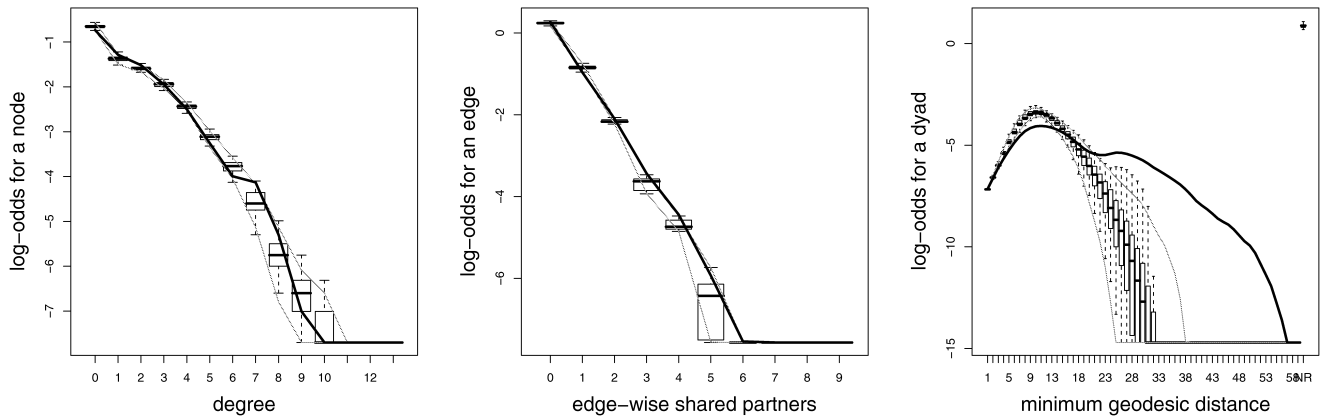
Figure 5. Goodness-of-fit plots for school 44, the largest AddHealth school with 2,209 nodes. The clear lack of fit in the geodesic distribution is typical of this model for the larger AddHealth schools, even though the same model tends to fit well on smaller schools. School 44, edges, covariates, GWESP, GWDSP, and GWD (all $\tau = .25$).

(Schwarz 1978), where

$$\text{AIC}(M) = -2(\text{maximized loglikelihood under } M)$$
$$+ 2(\text{\# of parameters in } M) \quad (22)$$

and

$$\text{BIC}(M) = -2(\text{maximized log-likelihood under } M)$$
$$+ \log N(\text{\# of parameters in } M), \quad (23)$$

where M denotes a particular ERGM and N denotes the sample size. The goal is to minimize $\text{AIC}(M)$ or $\text{BIC}(M)$ as a function of M. Unfortunately, these traditional methods entail several problems. For one thing, the assumptions used to justify the AIC and BIC are not met here, because our observations are not an independent and identically distributed sample. In fact, it is not even clear how to evaluate $\text{BIC}(M)$, because there is no easy way to determine the effective sample size N. For any dyadic independence model, $N = \binom{n}{2}$, the number of dyads. However, when dependence among dyads exists, the effective sample size can be smaller than $\binom{n}{2}$. Finally, as we pointed out in Section 3, it is not possible to evaluate the likelihood function directly for most ERGMs except in the case of dyadic independence models, where the likelihood equals the pseudolikelihood (4).

Thus any attempt to achieve model selection through the AIC or BIC is approximate at best. However, even using the roughly approximated AIC, we find that the conclusions of the graphical goodness-of-fit procedures are borne out in the sense that models that produce large reductions in the (approximate) AIC also seem to yield considerably better fits in the graphical plots, and those with smaller reductions in the AIC have less pronounced effects on the plots. However, the goodness-of-fit plots provide a richer picture than the AIC alone. From these plots, a number of features of the relationships between these models and the network structure become clear. For instance, both the plots and the AIC indicate that incorporating the heterogeneity of actors through nodal covariates is far more important to model fit than modeling either degree or shared partners alone. Yet the plots are more informative than the AIC results in the sense that they tell which structural features are fit well and which are not.

7. DISCUSSION

Only with the recent development of the R package statnet, which is available on the Internet at *http://csde. washington.edu/statnet*, has it been possible to reliably apply likelihood-based inference for ERGMs to networks of hundreds of nodes. Therefore, it is possible to consider other aspects of these models, such as how well they fit the observed data, as we
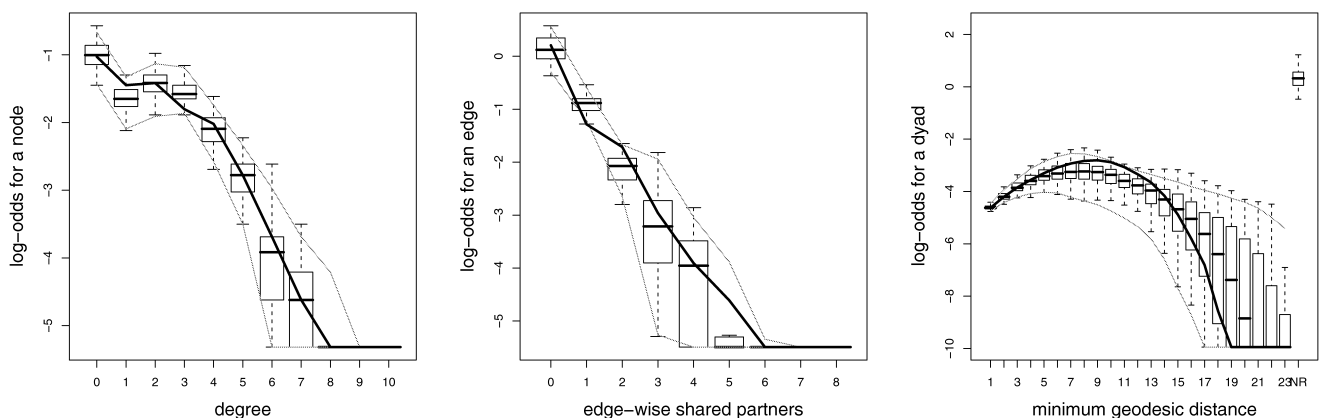

Figure 6. Simulation results for model II of Table 2. School 10, edges, covariates, GWESP, GWDSP, and GWD (all $\tau = .25$).

do in this article for the particular case of high school friendship networks.

Choosing an appropriate set of network statistics on which to compare the observed network with networks simulated from the fitted model is an important task in the graphical goodness-of-fit studies that we advocate in this article. If possible, these statistics should match the purpose of the network estimation and simulation. It may not be immediately clear what kinds of network properties are relevant; in fact, this might be precisely the question in which we are interested in the first place. For many social relations, theory may suggest that people do not look beyond more than one or two layers of network neighbors, so adequately modeling statistics such as the edgewise shared partner distribution might be expected to get higher-order statistics correct as well.

Comparing different AddHealth schools shows that many significant model parameters have remarkably similar qualitative patterns. Even the numerical values of the MLEs are often quite similar across friendship networks. However, when comparing networks with different numbers of nodes, it is important that the values of the parameter estimates are not necessarily comparable. The question of how to modify ERGMs so that their coefficients are directly comparable without regard to n, the number of nodes, is a very important issue in network modeling. Furthermore, as pointed out in Section 6, the related question of the effective sample size of a network on n nodes for a particular ERGM is important if we have any hope of applying model selection methods, such as the BIC, that depend on sample size. But this is a question for the future; for now, the science of likelihood-based methods for fitting ERGMs is still in its early stages.

Although the most complete and best-fitting model presented here appears to come close to capturing the higher-order network statistics examined for school 10 and many of the smaller schools, the same is not true for many of the larger schools. For instance, consider Figure 5, based on the largest school in our sample, with 2,209 nodes. This and other large schools depart from the fitted model in a similar way; the model underpredicts the number of long geodesics and overpredicts the number of short ones. In effect, the real social networks are more "stringy" than our best-fitting model predicts. Goodreau (2007) has provided a more detailed analysis of a large school.

As this empirical application has shown, both exogenous nodal covariates and endogenous network effects can play an important role in the generative processes that give rise to network structure. There is no a priori reason to assume that all networks will have the same structure, and the methods proposed here provide a systematic framework for evaluation of models that can be adapted to test a wide range of hypotheses. In the context of mutual friendships among high-school adolescents, geometrically weighted degree, edgewise shared partner, and dyadwise shared partner statistics—equivalent to the alternating k-star, k-triangle, and k-twopath statistics of Snijders et al.

(2006)—do a credible job of capturing the aggregate network structures of interest.

REFERENCES

Akaike, H. (1973), "Information Theory and an Extension of the Maximum Likelihood Principle," in *Second International Symposium on Information Theory*, eds. B. N. Petrov and F. Csaki, Budapest: Akademiai Kaidó, pp. 267–281.

Besag, J. (1974), "Spatial Interaction and the Statistical Analysis of Lattice Systems," *Journal of the Royal Statistical Society*, Ser. B, 36, 192–225.

Corander, J., Dahmström, K., and Dahmström, P. (1998), "Maximum Likelihood Estimation for Markov Graphs," Research Report 1998:8, University of Stockholm, Dept. of Statistics.

Crouch, B., Wasserman, S., and Trachtenberg, F. (1998), "Markov Chain Monte Carlo Maximum Likelihood Estimation for p^* Social Network Models," presented at the XVIII International Sunbelt Social Network Conference, Sitges, Spain.

Dahmström, K., and Dahmström, P. (1993), "ML Estimation of the Clustering Parameter in a Markov Graph Model," research report, University of Stockholm, Dept. of Statistics.

Frank, O., and Strauss, D. (1986), "Markov Graphs," *Journal of the American Statistical Association*, 81, 832–842.

Geyer, C. J., and Thompson, E. (1992), "Constrained Monte Carlo Maximum Likelihood for Dependent Data," *Journal of the Royal Statistical Society*, Ser. B, 54, 657–699.

Goodreau, S. M. (2007), "Advances in Exponential Random Graph (p^*) Models Applied to a Large Social Network," *Social Networks*, 29, 231–248.

Handcock, M. S. (2002), "Statistical Models for Social Networks: Inference and Degeneracy," in *Dynamic Social Network Modelling and Analysis: Workshop Summary and Papers*, eds. R. Breiger, K. Carley, and P. E. Pattison, Washington, DC: National Academy Press, pp. 229–240.

——— (2003), "Assessing Degeneracy in Statistical Models of Social Networks," Working Paper 39, Center for Statistics and the Social Sciences, University of Washington, available at *http://www.csss.washington.edu/Papers/*.

Holland, P. W., and Leinhardt, S. (1981), "An Exponential Family of Probability Distributions for Directed Graphs," *Journal of the American Statistical Association*, 76, 33–50.

Hunter, D. R. (2007), "Curved Exponential Family Models for Social Networks," *Social Networks*, 29, 216–230.

Hunter, D. R., and Handcock, M. S. (2006), "Inference in Curved Exponential Family Models for Networks," *Journal of Computational and Graphical Statistics*, 15, 565–583.

Lehmann, E. L. (1983), *Theory of Point Estimation*, New York: Wiley.

Morris, M. (2003), "Local Rules and Global Properties: Modeling the Emergence of Network Structure," in *Dynamic Social Network Modelling and Analysis: Workshop Summary and Papers*, eds. R. Breiger, K. Carley, and P. E. Pattison, Washington, DC: National Academy Press, pp. 174–186.

Resnick, M. D., Bearman, P. S., Blum, R. W., et al. (1997), "Protecting Adolescents From Harm: Findings From the National Longitudinal Study on Adolescent Health," *Journal of the American Medical Association*, 278, 823–832.

Schwarz, G. (1978), "Estimating the Dimension of a Model," *The Annals of Statistics*, 6, 461–464.

Snijders, T. A. B. (2002), "Markov Chain Monte Carlo Estimation of Exponential Random Graph Models," *Journal of Social Structure*, 3, available at *www.cmu.edu/joss/content/articles/volume3/Snijders.pdf*.

Snijders, T. A. B., Pattison, P. E., Robins, G. L., and Handcock, M. S. (2006), "New Specifications for Exponential Random Graph Models," *Sociological Methodology*, 36, 99–154.

Strauss, D., and Ikeda, M. (1990), "Pseudolikelihood Estimation for Social Networks," *Journal of the American Statistical Association*, 85, 204–212.

Udry, J. R., and Bearman, P. S. (1998), "New Methods for New Research on Adolescent Sexual Behavior," in *New Perspectives on Adolescent Risk Behavior*, ed. R. Jessor, New York: Cambridge University Press, pp. 241–269.

Wasserman, S., and Faust, K. (1994), *Social Network Analysis: Methods and Applications*, Cambridge, U.K.: Cambridge University Press.

Wasserman, S., and Pattison, P. E. (1996), "Logit Models and Logistic Regression for Social Networks: I. An Introduction to Markov Graphs and p^*," *Psychometrika*, 61, 401–425.