ELSEVIER

Advances in exponential random graph (p^*) models applied to a large social network

Steven M. Goodreau

University of Washington, Department of Anthropology and Center for Studies in Demography and Ecology, Washington, United States

Abstract

Recent advances in statistical network analysis based on the family of exponential random graph (ERG) models have greatly improved our ability to conduct inference on dependence in large social networks [Snijders, T.A.B., 2002. Markov Chain Monte Carlo Estimation of Exponential Random Graph Models. Journal of Social Structure 3, 2; Pattison, P.E., Robins, G.L., 2002. Neighbourhood-based models for social networks. Sociological Methodology 32, 301–337; Handcock, M.S., 2002. Statistical models for social networks: degeneracy and inference. In: Breiger, R., Carley, K., Pattison, P. (Eds.), Dynamic Social Network Modeling and Analysis. National Academies Press, Washington, DC, pp. 229–240; Handcock, M.S., 2003. Assessing Degeneracy in Statistical Models of Social Networks. Center for Statistics and the Social Sciences, University of Washington, Working Paper No. 39; Snijders, T.A.B., Pattison, P., Robins, G.L., Handcock, M., in press. New specifications for exponential random graph models. Sociological Methodology; Hunter, D.R., Goodreau, S.M. Handcock, M.S., 2005. Goodness of Fit of Social Network Models. Center for Statistics and the Social Sciences, University of Washington, Working Paper No. 47; Goodreau, S.M., Hunter, D.R., Morris, M., 2005. Statistical Modeling of Social Networks: Practical Advances and Results. Center for Studies in Demography and Ecology, University of Washington, Working Paper No. 05-01, previous papers this issue]. This paper applies advances in both model parameterizations and computational algorithms to an examination of the structure observed in an adolescent friendship network of 1681 actors from the National Longitudinal Study of Adolescent Health (AddHealth). ERG models of social network structure are fit using the R package statnet, and their adequacy assessed through comparison of model predictions with the observed data for higher-order network statistics.

For this friendship network, the commonly used model of Markov dependence leads to the problems of degeneracy discussed by Handcock [Handcock, M.S., 2002. Statistical models for social networks: degeneracy and inference. In: Breiger, R., Carley, K., Pattison, P. (Eds.), Dynamic Social Network Modeling and Analysis. National Academic Press, Washington, DC, pp. 229–240; Handcock, M.S., 2003. Assessing Degeneracy in Statistical Models of Social Networks. Center for Statistics and the Social Sciences, University of Washington, Working Paper No. 39]. On the other hand, model parameterizations introduced by Snijders et al. [Snijders, T.A.B., Pattison, P., Robins, G.L., Handcock, M., in press. New specifications for exponential random graph models. Sociological Methodology] and Hunter and Handcock [Hunter, D.R., Handcock,

E-mail address: goodreau@u.washington.edu.

M.S., in press. Inference in curved exponential family models for networks. Journal of Computational and Graphical Statistics] avoid degeneracy and provide reasonable fit to the data. Degree-only models did a poor job of capturing observed network structure; those that did best included terms both for heterogeneous mixing on exogenous attributes (grade and self-reported race) as well as endogenous clustering. Networks simulated from this model were largely consistent with the observed network on multiple higher-order network statistics, including the number of triangles, the size of the largest component, the overall reachability, the distribution of geodesic distances, the degree distribution, and the shared partner distribution. The ability to fit such models to large datasets and to make inference about the underling processes generating the network represents a major advance in the field of statistical network analysis.

The exponential random graph (ERG) class was first posited as an approach to model social network structure two decades ago (Frank and Strauss, 1986), based upon work in spatial statistics (Besag, 1974). The modeling class is extremely general (Wasserman and Pattison, 1996), and as such should in theory be capable of capturing the structure of a wide array of empirical networks, allowing for statistical inference about that structure to be conducted. Nevertheless, most work in this field has focused on a small set of model specifications, most commonly the Markov graphs of Frank and Strauss (1986). Recent work has shown that these commonly used model specifications are in fact not well suited to capturing the processes underlying many empirical networks, due to the problem of model degeneracy (Handcock, 2002, 2003). Degeneracy can be described in brief as the phenomenon in which a seemingly reasonable model can actually be such a bad mis-specification for an observed dataset as to render the observed data virtually impossible under the model. Instead, the social processes encapsulated by the model yield networks such as the full or empty graph, which are qualitatively dissimilar to the observed data and are in fact substantively uninteresting. A more detailed discussion of degeneracy can be found in Robins et al. (2007b).

If model degeneracy were only a sign that a specific model is not well suited to the data to which it is being fit, it may not have proven such a barrier to progress in the field. In other realms, poorly fitting models can often help refine knowledge of data; an analysis of where the model succeeds in capturing the observed data and where it fails can help point the way to subsequent model refinements. In the case of ERG models, however, the maximum likelihood estimates for model parameters are fit using Markov chain Monte Carlo (MCMC) procedures, and degeneracy in fact often prevents model estimation from converging on finite parameter estimates. The use of alternative approximation techniques such as MPLE estimation (Strauss and Ikeda, 1990) does not solve this problem, but simply hides it. The fact that lack of convergence can also result from various features of the MCMC fitting algorithm does not help matters. Collectively, these problems explain the slow application of ERG models to empirical network analysis thus far, despite their potential.

Previous papers in this issue have provided an introduction to the ERG modeling approach (Robins et al., 2007a), examined more deeply the issues behind degeneracy that have impeded the application of these models (Robins et al., 2007b), and provided potential solutions forward (Robins et al., 2007b; Hunter, 2006) In this paper our aim is to apply these developments to a single large (>1600 actors) set of network data to demonstrate their practical use in conducting inference on complex, dependent relational structure.

1. Data set: add health school group 42

The dataset we analyze here is one of the sets of schools from the National Longitudinal Study of Adolescent Health (AddHealth). AddHealth was a stratified school-based sample of US students in grades 7–12. It featured extensive questionnaires on individual characteristics, as well as a module on friendship networks. Students were provided a roster listing all students in the school by name and by unique ID number, and asked to list the ID for up to five best male and five best female friends in rank order. Students were free to nominate fewer than five of either. They were also free to nominate friends who were not in the school (through a special code) or who were in the school but not included on the roster. The dataset here contains two separate schools, both large multi-racial public schools in the urban South; one containing grades 10–12 (which we call the high school) and another school containing grades 7–9 (the junior high school), which served as a feeder for the high school. Students in each school were provided with the rosters for both and allowed to choose from either. Further information on the design of the study can be found in Resnick et al. (1997) and Udry and Bearman (1998) and at http://www.cpc.unc.edu/projects/addhealth.

For this analysis, we consider only those actors who both completed the survey and were on the roster ($n = 1681$). An additional 489 students were on the roster but did not take the survey, while 158 students took the survey but either were not on the roster or improperly filled in their own ID numbers in such a way as to not be matchable to the roster. The total number of uniquely identifiable students in the school is thus 2328 (more could be present but undetectable if they both did not take the survey and were off the roster). Of this the 1681 students in the dataset comprise 72%, a level similar to that for the AddHealth schools generally (Schruth, personal communication). Although we cannot know whether the structure of the missing students' relations are similar to those in the study, we can identify the mean in-degree of students who did not take the survey (3.3) and those who did (3.9) for the raw nondirected data.

We consider the mutualized data, in which an edge is included only if both actors nominate the other. Mutualization is conducted both for a substantive reason (a dually nominated tie may be taken as a form of validated relationship) and for a practical one (some of the network statistics that we consider below have thus far only been defined and implemented for nondirected relations). The number of edges in the resulting nondirected network is 1236, implying a density of 0.00087 (1236/1,412,040 actor pairs) and a mean degree of 1.47. These 1236 edges represented 2472 arcs in the original data, or 35% of the 6985 nominations these students made in total—the other 65% were not reciprocated. Interpretation of the mutualized data may be complicated by the fact that students are only allowed to nominate a fixed number of friends; more relationships might have appeared mutual had students been allowed to nominate an arbitrary number. This effect is dampened by the fact that 53% (895/1681) of the respondents nominated fewer than the maximum number of female friends and 54% (916/1681) nominated fewer than the maximum males.

The exogenous actor attributes we consider are grade, race and sex. Students who took the survey were free to leave any of these blank that they wished; for all attributes except grade we retained the blank responses, considering all such responses for a given attribute to constitute a single factor level for that attribute in any relevant analyses. Grade was applied to a more extensive set of network statistics than other attributes; as these statistics do not allow for missing values, missing grades were imputed. There were ten students out of the 1681 who left their grade blank. Four of the students revealed at least one mutual friend, and in each case their entire set of friends was in one grade. These four students were assigned the grade of all their friends. Of the remaining six, four answered the question on age, and were assigned the modal grade for

their age (age 12 = grade 7, age 13 = grade 8, etc.). The remaining two, for whom no information was available, were randomly assigned grades with probability proportional to the size of the grade. The variable we term race is taken from a pair of questions on self-identified race and Hispanic origin. Hispanic origin is considered primary; that is, anyone who identifies as Hispanic is categorized as Hispanic regardless of race; all other racial categories represent non-Hispanics. We collapsed all answers other than white, black or Hispanic into a single category other given their small numbers in these data and the absence of any within-group ties for some categories, which prevents convergence to finite parameters for some models.

2. Methods

The ERG modeling class defines the probability of a network with a given set of actors n as

$$\Pr(Y = y) = \left(\frac{1}{\kappa}\right) \exp\{\Sigma_A \eta_A g_A(y)\}$$

The notation $g_A(y)$ represents any possible network statistic, where A indexes the multiple statistics included in a model vector $g(y)$; we will see numerous examples in the next section. η_A represents the coefficients for these terms; their value reflects the change in the conditional log-odds of a tie for each unit increase in g_A that the tie would create. κ represents the normalizing constant, the sum of $\exp(\Sigma_A \eta_A g_A(y))$ over all possible networks with n actors. For all but the smallest networks or simplest models, this normalizing constant prohibits the direct evaluation of the probability for a specific network given the vectors $g(y)$ and η. It also inhibits calculation of the maximum likelihood values for the parameter vector η from an observed network. Strauss and Ikeda (1990) discuss the use of logistic regression to calculate a maximum pseudolikelihood estimate (MPLE), although the quality of this as an estimate of the true maximum likelihood may be poor for models with strong global dependence (Besag, 1986; Handcock, 2003). Geyer and Thompson (1992) on the other hand, provide methods for adapting Markov chain Monte Carlo (MCMC) as a general estimation tool for problems of this type, and Snijders (2002) discusses its use for social networks. To obtain the maximum likelihood estimate (MLE) for the η parameters, we use logistic regression for dyad independence models (since the resulting MPLE estimate is identical to the true MLE) and MCMC for dyadic dependence models. For nondirected networks, dyadic independence models are defined as those in which $P(Y_{ij} = y_{ij})$ is independent of $P(Y_{kl} = y_{kl}) \,\forall\, (i, j) \neq (k, l)$ conditional on the actor attributes; such models typically consist of an edge term and set of terms counting the number of instances of edges among actors with different attribute combinations. Examples of dyadic dependence models include those involving triangle, star or degree terms.

All model fitting and evaluation occurred in the statnet package, a set of statistical network analysis routines in the R environment (Handcock et al., 2004). This package incorporates the new model specifications discussed by earlier papers in this issue as well as a number of algorithmic developments that optimize the process of fitting models to data rapidly and accurately. Some of these features include the use of *R* for user interface and data manipulation but *C* for complex memory-intensive calculations; more efficient storage approaches for sparse networks; the use of hybrid algorithms that allow for rapid movement towards the MLE when far away and more precise refinement once close; and the use of proposal algorithms that speed up mixture of the chain by orders of magnitude over the commonly used simple dyad toggle. Access to the package and more detailed information on its features and use can be found at http://csde.washington.edu/statnet.

For the MCMC-based estimation procedures in this paper, we selected a chain burn-in of 100,000 toggles, an MCMC sample size of 10,000 and an interval between successive samples of 1000 toggles. Chains were started at the MPLE values for the η vector obtained through logistic regression. The chain was run for the above length (100,000 + 10,000 × 1000 = 10.1 million steps), and new estimates of η were obtained from the chain using the Geyer and Thompson (1992) algorithm. The chain was then re-initiated from this updated starting point; this cycle was repeated up to five times to obtain the final estimates of η for each model.

3. Model terms

The ERG model class is general; it includes an infinite number of potential network statistics. Here we focus on statistics that are common in the literature (including many of those discussed in earlier papers in this edition), which are theoretically relevant to these nondirected friendship data, and which are feasible to calculate for networks of size ~1600 actors. Emphasis is on relatively "local" statistics (those in which the probability of a given edge is directly dependent on only a small number of other dyads in the graph)[1] with the hope that these can capture the global structure in the network.

Models for social networks typically contain multiple such terms to capture the various processes at work simultaneously in the creation of social relationships. Two general approaches exist for the selection of specific combinations of statistics to consider for social networks, both of which we employ in the next section. The first involves the derivation of sets of terms from assumptions about the nature of dependence, often via the Hammersley–Clifford Theorem (Besag, 1974). The advantage of such an approach is that the exact nature of the dependence that underlies the model is explicit. One common example of this is the Markov dependence model of Frank and Strauss (1986), which in its homogeneous form leads to terms for edges, triangles and the set of stars (or, equivalently, degree). Another is the partial conditional independence models first explained in Pattison and Robins (2002) and expanded upon by Snijders et al. (in press) and discussed in this issue (Robins et al., 2007a). This work considers a form of "realized" dependence one step larger than Markov; includes the Markov terms as well as sets of terms for *k*-triangles and *k*-twopaths. Including a full set of *k*-star, *k*-triangle, and *k*-twopath terms yields a large number of parameters and can lead to problems of degeneracy and difficulty in interpretation. Instead, Snijders et al. (in press) propose parametric forms for these distributions (the "alternating *k*-triangle", "alternating *k*-star", and "alternating *k*-twopath" statistics) as a way of reducing the parameter space. The first two terms are equivalent to the geometrically weighted edgewise shared partner (GWESP) and dyad-wise shared partner (GWDSP) terms explored by Hunter (2006). The third shares a similar relationship to the geometrically weighted degree (GWD) term of Hunter (2006), although here the relationship involves a minor reparameterization rather than exact equivalence.

An alternative approach is to consider many possible combinations of statistics and observe which combinations empirically yield the best fit for a given data set. This is akin to the various forms of iterative model selection in linear regression or other generalized linear models. This approach allows for great flexibility in comparing models for a given dataset based on observed structure. However, given the complex interactions of some of the higher-order terms, and the non-linearity of their effects, the general ability to approach a well-fitting model through the iterative addition or removal of terms is not well understood. For instance, two different terms

---

[1] For a more in-depth discussion of this concept see Pattison and Robins (2002).

with countervailing effects may be necessary to avoid degeneracy; adding either one alone may provide no insight into the suitability of a model containing both. Nevertheless, this approach can, when combined with a mixture of theoretical grounding and trial and error, achieve success in model fit for some large, complex networks (Hunter et al., 2005; Goodreau et al., 2005).

Below we define each of the constituent terms that appear in any of the models. Model expressions are framed under the assumption of a symmetric matrix given our nondirected data. We first define a set of variables:

- x_{vi}: the attribute value for actor i for attribute v;
- X_{va}: the set of all actors who have value a for attribute v;
- $d_i = \sum_j y_{ij}$: the degree of actor i;
- $sp_{ij} = \sum_k y_{ik} y_{jk}$: the number of shared partners for actors i and j, that is, the number of neighbors they have in common;
- $EP_k(\boldsymbol{y}) = \sum_{i<j} y_{ij} I \{sp_{ij} = k\}$: the k-edgewise shared partner statistic;
- $DP_k(\boldsymbol{y}) = \sum_{i<j} I \{sp_{ij} = k\}$: the k-dyadwise shared partner statistic.

where $I\{\ \}$ is the indicator function, equal to 1 if the enclosed statement is true and 0 if false. Given these, we then have the following network statistics which we consider for entry into our $z(\boldsymbol{y})$ vector for different models:

- $L(\boldsymbol{y})$: $\sum_{i<j} y_{ij}$: the edge statistic;
- $T(\boldsymbol{y})$: $\sum_{i<j<m} y_{ij} y_{jm} y_{im}$: the triangle statistic;
- $D_k(\boldsymbol{y})$: $\sum_i I \{d_i = k\}$: the k-degree statistic;
- $S_k(\boldsymbol{y})$: $\sum_{m=1}^{n-1} \binom{m}{k} D_m(y)$: the k-star statistic;
- $M_{v,a}(\boldsymbol{y})$: $\sum_{i \in X_{va}} d_i$: the edge statistic for attribute v level a;
- $H_{v,a}(\boldsymbol{y})$: $\sum_{i \in X_{va}, j \in X_{va}, i<j} y_{ij}$: the differential homophily statistic for attribute v level a;
- $U_v(\boldsymbol{y})$: $\sum_{i<j} y_{ij} I \{x_{vi} = x_{vj}\}$: the uniform homophily statistic for attribute v;
- $A_v(\boldsymbol{y})$: $\sum_{i<j} y_{ij} |x_{vi} - x_{vj}|$: the absolute difference statistic for attribute v;
- $u(\boldsymbol{y}, \theta_1)$: $e^{\theta_1} \sum_{k=1}^{n-1} \{1 - (1 - e^{-\theta_1})^k\} D_k(y)$: the geometrically weighted degree statistic with parameter θ_1;
- $v(\boldsymbol{y}, \theta_2) = e^{\theta_2} \sum_{k=1}^{n-2} \{1 - (1 - e^{-\theta_2})^k\} EP_k(y)$ the geometrically weighted edgewise shared partner statistic with parameter θ_2;
- $w(\boldsymbol{y}, \theta_3)$: $e^{\theta_3} \sum_{k=1}^{n-2} \{1 - (1 - e^{-\theta_3})^k\} DP_k(y)$ the geometrically weighted dyad-wise shared partner statistic with parameter θ_3;

$L(\boldsymbol{y})$, $T(\boldsymbol{y})$, $D_k(\boldsymbol{y})$ and $S_k(\boldsymbol{y})$ have long histories in the network literature and are of general familiarity. $M_{v,a}(\boldsymbol{y})$, $H_{v,a}(\boldsymbol{y})$, $U_v(\boldsymbol{y})$ and $A_v(\boldsymbol{y})$ are attribute-specific terms that capture the way in which actor attributes structure relations. $M_{v,a}(\boldsymbol{y})$ models main effects, allowing each level of an attribute to form edges with different propensities. $H_{v,a}(\boldsymbol{y})$ models a separate tendency for members of each attribute level to form ties within attribute class, while $U_v(\boldsymbol{y})$ models a single such within-group preference for all groups. $A_v(\boldsymbol{y})$ models the tendency for tie probabilities to change monotonically as the absolute value of the difference in attribute value for two actors increases; this term is only defined for valued or ordinal attributes.

The statistics $u(\mathbf{y}, \theta_1)$, $v(\mathbf{y}, \theta_2)$ and $w(\mathbf{y}, \theta_3)$ represent the parametric higher-order terms of Snijders et al. (in press), re-parameterized following Hunter and Handcock (2006) and Hunter (2006; Eqs. (14), (25), and (26), respectively). The geometrically weighted degree term $u(\mathbf{y}, \theta_1)$ represents a parametric form of the degree distribution; $v(\mathbf{y}, \theta_2)$ represents a parametric form of clustering, and is equivalent to alternating k-triangles. Finally, $w(\mathbf{y}, \theta_3)$ can be thought of as a parametric form on degrees of structural equivalence (an interpretation not previously discussed in the literature), because it considers actors who may or may not be tied to one another, but are tied to various numbers of the same other actors. This term is equivalent to the alternating k-twopaths of Snijders et al. (in press).

Note that all three of the θ terms can take any positive value. To see what this range can represent, consider $v(\mathbf{y}, \theta_2)$. As θ_2 takes values approaching infinity, the $v(\mathbf{y}, \theta_2)$ statistic approaches three times the number of triangles in the graph.[2] For θ_2 close to 0, $v(\mathbf{y}, \theta_2)$ approaches the number of edges in the graph that are part of at least one triangle. This latter statistic still incorporates dyadic dependence, but in a form that is strongly capped; once a pair of actors is in one triangle, they have no special propensity to form any more. θ_2 can be set at any value in between to capture the scale and magnitude of clustering; the value of θ_2 that maximizes the model likelihood can also be estimated using the methods of Hunter and Handcock (2006) and Hunter (2006), although this can dramatically increase necessary computation time. θ_1 and θ_3 can similarly be estimated or considered fixed. In the analysis below we adopt a fixed value of 0.5 for all three θ terms. This is based on work exploring the likelihood for some reduced models (in this and other smaller schools) at multiple levels for the θ terms, identifying that the likelihood surface is quite flat in the range 0.25–1, and the maximum is generally close to 0.5. Free-varying θ values were not used, as this approach did not converge for all models presented here.

Theoretically derived combinations of the above terms that we consider include the homogeneous Markov model of Frank and Strauss (1986), in which the statistics included are $L(\mathbf{y})$, $T(\mathbf{y})$, and $S_k(\mathbf{y})$ for $k \in \{2, \ldots, n-1\}$.[3] We also examined a reduced Markov model commonly examined in the networks literature (e.g. Robins et al., 2007b) which contains only $L(\mathbf{y})$, $T(\mathbf{y})$, $S_2(\mathbf{y})$ and $S_3(\mathbf{y})$. Given the current interest in degree-only models in some branches of network literature, we also fit a model containing the terms $D_n(\mathbf{y})$, in which the degree distribution is completely captured in a fully saturated non-parametric form, to see how well this model fits the overall structure of the network. This formulation represents the nondirected analogue of that discussed by Snijders and van Duijn (2002). The saturated degree model necessarily fits the degree distribution as well or better than any of the various parameterized degree distributions currently discussed in the literature.

We also employ the method of adding each of a set of proposed terms to the current model, seeing which yields the most improvement in model fit, and selecting that to add to the model before beginning the next iteration of model-building. Some of these terms are functions of others (e.g. $U_v(\mathbf{y})$ is the sum of $H_{v,a}(\mathbf{y})$ across all α) so the inclusion of one eliminates the consideration of the other in subsequent rounds. Trying every model for our list of proposed statistics is simply too extensive, and an examination of the data indicates that some terms are clearly fundamental;

---

[2] To see this, consider that as θ_2 gets large, $e^{-\theta_2}$ approaches 0 from the right, and $(1 - e^{-\theta_2})^k$ approaches $1 - k\,e^{-\theta_2}$ since all other terms in the expansion are of second-order or more in $e^{-\theta_2}$. Thus $1 - (1 - e^{-\theta_2})^k$ approaches $k\,e^{-\theta_2}$, and the entire quantity $v(\mathbf{y}, \theta_2)$ approaches $\sum_{k=1}^{n-2} k\mathrm{EP}_k(y)$, or the grand total of all the shared partners for each edge. This formula counts each triangle exactly three times, once with each edge as the base for counting shared partners.

[3] The inclusion of $S_k(y)$ terms for k larger than the largest degree in the graph plus one is unnecessary, as the parameter value for $S_k(y)$ when k equals the largest degree plus one will be negative infinity and prevent all larger stars.

we begin with only these terms of obvious central importance as a way of reducing the number of models to fit, as seen below.

4. Model selection and goodness of fit

In order to examine the goodness of fit models, we use three general approaches:

1. Check for degeneracy and model convergence: a minimum requirement for a model to fit well is for estimation of parameters to converge on finite parameter values. It must also be non-degenerate, that is, not place all of its probability mass on a few networks entirely unlike the observed network, such as a full or empty network.
2. Compare the Akaike information criterion (AIC) between models. Models that exhibit dyadic independence can be fit with standard logistic regressions, which yields a likelihood measure for the model. Those models that are dyad dependent must be fit with MCMC, which also yields an estimate of the likelihood. These likelihood values are only approximate since not all graphs are possible under the AddHealth data set; the limitation to an out-degree of ten in the data collection, for instance, is one such form of restriction. We ignore such effects here, and use the given likelihood to calculate AIC in order to compare models, with lower AIC implying a significant increase in model fit. Note that as the models become increasingly complex and include more dyadic dependence terms, the approximations for the likelihood become increasingly less precise (under current methods), and additional methods for model selection must be found. For more information on the use of AIC for ERGM model fit, see Hunter et al. (2005).
3. Goodness of fit plotting for higher order statistics. This approach is described in detail in Hunter et al. (2005). The logic entails generating new networks according to the probability distribution implied by the fit model. Because the normalizing constant is still present in the fit models, this must be done using the same MCMC approach employed during the estimation procedure. A statistic of interest is then calculated on the original network and on the set of networks generated from the model, and these are plotted for comparison. If the original network is inconsistent with the networks generated from the model, this suggests that the structure of the network differs from those predicted by the model, and the model is not well fit. Multiple statistics can be compared visually to provide detailed information about the systematic ways in which the data and the model predictions differ.

For this approach, the network statistics we compare include the degree distribution (across all actors), the shared partner distribution (across all edges), and the geodesic distance distribution (across all dyads). Degree and geodesic distance are well known in the networks literature; shared partner, on the other hand, is a relatively new concept capturing the patterns of clustering in a network. The shared partner value for an edge (the sp_{ij} statistic defined earlier) represents the number of actors to whom both actors in that edge are tied. This distribution across all ties provides a picture not only of the magnitude of clustering but also the scales at which it occurs. Each of these terms is plotted on the log-odds scale for the sake of greater visibility across the covered range of values. Although we have selected distributions of three statistics to compare, the approach is far more general; any statistic of interest may be considered across graphs, and selection of such statistics may be guided by a combination of network theory and an investigator's individual objectives.

Table 1
Fit for basic dyadic independence models

Model #	Terms	# Terms	AIC	Interpretation
1	$\{L(\mathbf{y})\}$	1	19878.0	All edges equiprobable (Bernoulli/Erdös-Rényi model)
6	$\{L(\mathbf{y}), M_{\text{race,a}}(\mathbf{y})\}$	7	19826.0	Differential levels of edge formation by race
7	$\{L(\mathbf{y}), M_{\text{grade,a}}(\mathbf{y})\}$	6	19843.1	Differential levels of edge formation by grade
8	$\{L(\mathbf{y}), H_{\text{race,a}}(\mathbf{y})\}$	6	19367.2	Differential assortative mixing by race
9	$\{L(\mathbf{y}), H_{\text{grade,a}}(\mathbf{y})\}$	7	17584.2	Differential assortative mixing by grade
10	$\{L(\mathbf{y}), U_{\text{race}}(\mathbf{y})\}$	2	19450.6	Uniform assortative mixing by race
11	$\{L(\mathbf{y}), U_{\text{grade}}(\mathbf{y})\}$	2	17637.5	Uniform assortative mixing by grade
12	$\{L(\mathbf{y}), H_{\text{grade,a}}(\mathbf{y}), H_{\text{race,a}}(\mathbf{y}), M_{\text{race,a}}(\mathbf{y}), M_{\text{grade,a}}(\mathbf{y})\}$	23	16851.7	Differential levels and assortative mixing by race and by grade

5. Results

We begin with the Bernoulli model (model 1), that with only a single term to capture the density of the network. The AIC for this model is listed in Table 1; the parameter estimates in Table 2 (along with their standard errors estimated using the method of Geyer, 1994), and the

Table 2
Parameter values $\eta(\mathbf{y})$ for select models (standard error in parentheses)

	Model 01	Model 04	Model 12	Model 24	Model 27
$L(\mathbf{y})$	$-7.04\,(0.03)^*$	$-5.32\,(0.27)^*$	$-11.03\,(0.39)^*$	$-10.89\,(0.39)^*$	$-11.14\,(0.38)^*$
$u(\mathbf{y}, 0.5)$		$-1.60\,(0.18)^*$			
$v(\mathbf{y}, 0.5)$		$1.85\,(0.05)^*$		$1.64\,(0.04)^*$	$1.64\,(0.04)^*$
$w(\mathbf{y}, 0.5)$		$-0.22\,(0.04)^*$			
$H_{\text{grade,07}}(\mathbf{y})$			$5.10\,(0.38)^*$	$4.57\,(0.37)^*$	$4.68\,(0.37)^*$
$H_{\text{grade,08}}(\mathbf{y})$			$4.17\,(0.32)^*$	$3.78\,(0.32)^*$	$3.70\,(0.32)^*$
$H_{\text{grade,09}}(\mathbf{y})$			$4.37\,(0.32)^*$	$3.89\,(0.31)^*$	$3.71\,(0.31)^*$
$H_{\text{grade,10}}(\mathbf{y})$			$2.68\,(0.22)^*$	$2.43\,(0.21)^*$	$2.40\,(0.22)^*$
$H_{\text{grade,11}}(\mathbf{y})$			$1.36\,(0.20)^*$	$1.09\,(0.20)^*$	$1.08\,(0.19)^*$
$H_{\text{grade,12}}(\mathbf{y})$			$2.46\,(0.22)^*$	$2.10\,(0.21)^*$	$2.19\,(0.21)^*$
$H_{\text{race,white}}(\mathbf{y})$			$1.55\,(0.15)^*$	$1.32\,(0.14)^*$	$1.30\,(0.14)^*$
$H_{\text{race,black}}(\mathbf{y})$			$2.07\,(0.15)^*$	$1.79\,(0.14)^*$	$1.97\,(0.15)^*$
$H_{\text{race,Hisp.}}(\mathbf{y})$			$-0.02\,(0.23)$	$0.07\,(0.25)$	$-0.01\,(0.25)$
$H_{\text{race,other}}(\mathbf{y})$			$0.68\,(0.22)^*$	$0.50\,(0.23)^*$	$0.27\,(0.24)$
$M_{\text{grade,08}}(\mathbf{y})$			$0.40\,(0.24)$	$0.38\,(0.24)$	$0.32\,(0.24)$
$M_{\text{grade,09}}(\mathbf{y})$			$0.38\,(0.24)$	$0.36\,(0.24)$	$0.41\,(0.24)$
$M_{\text{grade,10}}(\mathbf{y})$			$0.86\,(0.21)^*$	$0.83\,(0.21)^*$	$0.79\,(0.21)$
$M_{\text{grade,11}}(\mathbf{y})$			$1.67\,(0.22)^*$	$1.57\,(0.21)^*$	$1.54\,(0.21)$
$M_{\text{grade,12}}(\mathbf{y})$			$1.22\,(0.21)^*$	$1.14\,(0.21)^*$	$1.09\,(0.21)$
$M_{\text{race,black}}(\mathbf{y})$			$-0.47\,(0.11)^*$	$-0.39\,(0.11)^*$	$-0.44\,(0.10)$
$M_{\text{race,Hisp.}}(\mathbf{y})$			$0.72\,(0.11)^*$	$0.65\,(0.11)^*$	$0.67\,(0.11)$
$M_{\text{race,other}}(\mathbf{y})$			$0.51\,(0.12)^*$	$0.44\,(0.10)^*$	$0.55\,(0.12)$
$M_{\text{sex,female}}(\mathbf{y})$					$0.23\,(0.04)$
$M_{\text{sex,NA}}(\mathbf{y})$					$0.14\,(0.22)$

* $p < 0.05$.

goodness-of-fit plots in Fig. 1. Not surprisingly, this simplistic model does not capture the larger statistics of the original network compared in the goodness of fit plots.

We next consider the standard Markov model (model 2). This model does not converge to finite parameter estimates; in multiple runs under varying conditions the value of the triangle parameter runs to positive infinity and other terms to negative infinity; the true maximum likelihood either occurs at infinite parameters, or the observed network is so unlikely under the finite maximum likelihood parameters for this model that the fitting procedure cannot converge. Model 3, the reduced Markov model, yields the same outcome. Model 4, a homogeneous form of the realization-dependent model of Snijders et al. (in press), provides much better results; its parameter estimates are included in Table 1 and a plot of its goodness of fit statistics are shown in Fig. 1b. The model converges, and is largely capable of capturing the distribution of degree and shared partners, network features that it was modeling using a simple parameterized form. However, the model's ability to capture the higher-order structure of path lengths is limited. A model that completely parameterizes the degree distribution (model 5) but contains no other terms captures the degree distribution perfectly (we would be in trouble if it did not); yet it is also completely off for both the clustering and the geodesics (Fig. 1c); for geodesic length, it is even further from the source data than the Bernoulli model is. Even capturing the degrees perfectly tells one very little about other relatively local (shared partner distribution) and global (geodesic distribution) network structure.

Since none of the homogeneous models captured all of the network structure under consideration, we turn to the process of iteratively adding terms depending on their improvement to the model fit. We return to the Bernoulli model as our starting point. Mixing matrices of partnerships by race and by grade (not shown) display strong tendencies for within-race and within-grade friendships. Mixing on these attributes is undoubtedly a strong element in network formation, and will thus presumably be included in any model that accurately captures the network generation process. Thus, at first we consider only six models relating to these two statistics:

- Model 6: $g(y) = \{L(y), M_{race,a}(y)\}$ for $a \in \{$black, Hispanic, other$\}$ with white as a reference category;
- Model 7: $g(y) = \{L(y), M_{grade,a}(y)\}$ for $a \in \{8, 9, 10, 11, 12\}$ with grade 7 as a reference category;
- Model 8: $g(y) = \{L(y), H_{race,a}(y)\}$ for $a \in \{$white, black, Hispanic, other$\}$;
- Model 9: $g(y) = \{L(y), H_{grade,a}(y)\}$ for $a \in \{7, 8, 9, 10, 11, 12\}$;
- Model 10: $g(y) = \{L(y), U_{race}(y)\}$;
- Model 11: $g(y) = \{L(y), U_{grade}(y)\}$.

All six of these models are independence models, that is, the MLE can be found using logistic regression. The effect of each model on the AIC is shown in Table 1. We note that differential homophily by grade (the set of six $H_{grade,a}(y)$ statistics) has the largest effect, and include these terms in all subsequent models; we then consider the four models that include $L(y), H_{grade,a}(y)$ and one additional term set. Note that there is no need to consider a model for $g(y) = \{L(y), H_{grade,a}(y), U_{grade}(y)\}$; once the $H_{grade,a}(y)$ terms are in the model, the $U_{grade}(y)$ term is redundant. Of these four models, all yield a large reduction in AIC, with $L(y), H_{grade,a}(y), H_{race,a}(y)$ yielding the largest. We continue this process (not shown) and discover that in fact the maximal model:

- Model 12: $g(y) = \{L(y), H_{grade,a}(y), H_{race,a}(y), M_{grade,a}(y), M_{race,a}(y)\}$

is significantly better fitting than any of the sub-models as measured by AIC, with terms entering in the order above. We call this the "race/grade base model", with all subsequent models
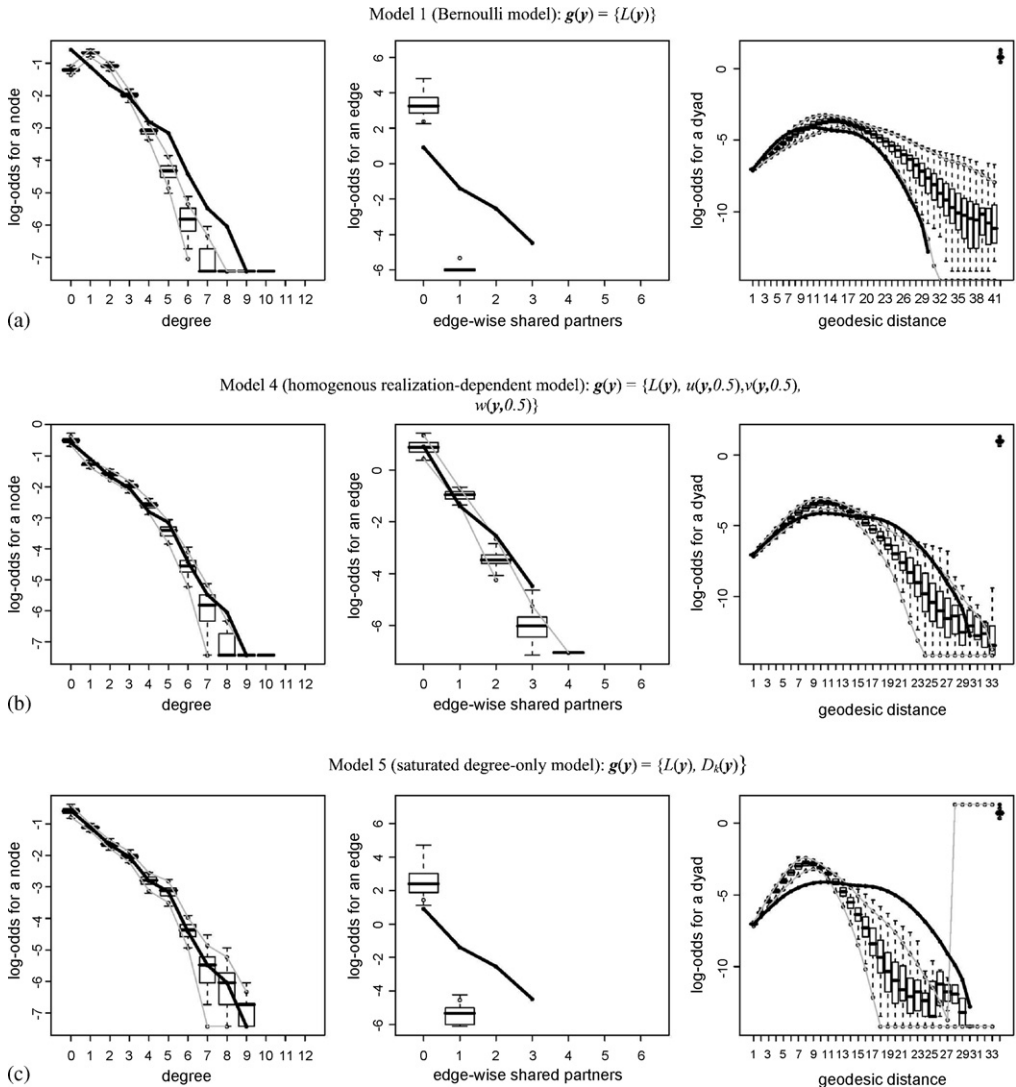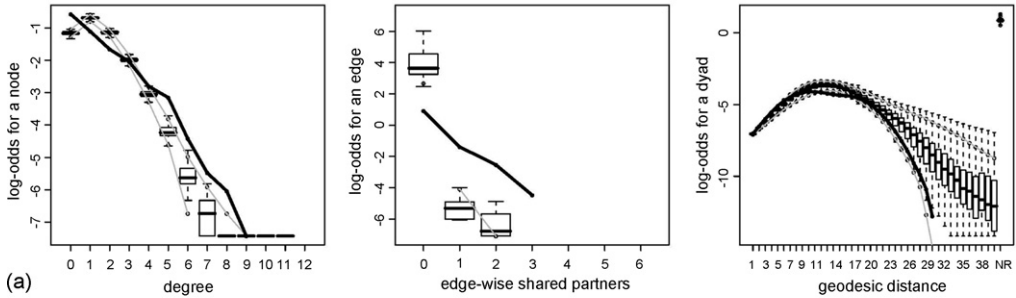
Fig. 1. Goodness-of-fit boxplots for select models. Plots compare degree distribution, shared partner distribution, and geodesic distribution between observed data and 100 networks simulated from the given model. *X*-axis represents the value of the respective distribution. *Y*-axis represents the fraction of nodes, edges, or dyads, respectively, exhibiting that value. *Y*-axis is on the log-odds scale for greater readability. Solid line represents original network; boxplots display the distribution of the same statistics across the 100 simulated networks. For example, 5.7% of the nodes in the original data (96/1681) have degree 4. This can be seen in the first panel of Fig. 1a, where the original data shows the log-odds of degree 4 equaling $\ln(0.057/(1 - 0.057)) = -2.8$. Similarly, the second figure in each panel plots the log-odds that an edge will posses a given number of shared partners, and the third panel plots the log-odds that are dyad will possess a given geodesic length. (a) Model 1 (Bernoulli model): $g(y) = \{L(y)\}$. (b) Model 4 (homogenous realization-dependent model): $g(y) = \{L(y), u(y, 0.5), v(y, 0.5), w(y, 0.5)\}$. (c) Model 5 (saturated degree-only model): $g(y) = \{L(y), D_k(y)\}$.

Model 12 (race/grade base model): $g(y) = \{L(y), H_{grade,a}(y), H_{race,a}(y), M_{grade,a}(y),$
$M_{race,a}(y)\}$

(a)

Model 24 (race/grade/GWESP model): $g(y) = \{$Race/grade base model, $v(y, 0.5)\}$

(b)
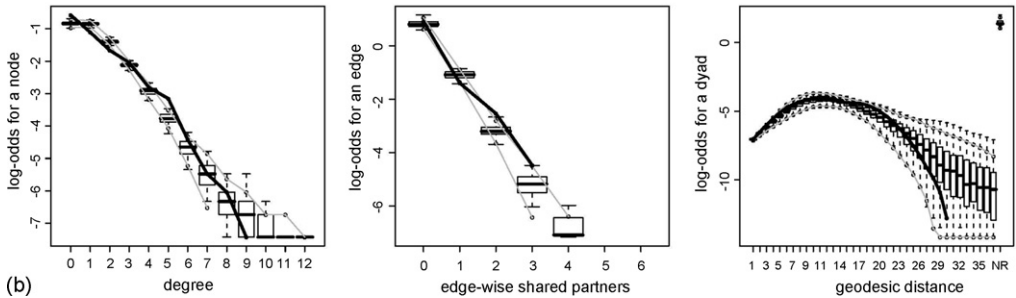
Fig. 2. Goodness-of-fit boxplots for select models (see Fig. 1 for explanation). (a) Model 12 (race/grade base model): $g(y) = \{L(y), H_{grade,a}(y), H_{race,a}(y), M_{grade,a}(y), M_{race,a}(y)\}$. (b) Model 24 (race/grade/GWESP model): $g(y) = \{$race/grade base model, $v(y, 0.5)\}$.

including this set of terms, and include its features in Table 2 and Fig. 2a. From Table 2 we see that each grade exhibits significant levels of within-group homophily, as do whites, blacks, and those of "other" race (the H parameters). In addition, we see increasing levels of baseline relationship formation with increasing grade relative to grade 7; Hispanics and students of "other" race have significantly higher, and blacks significantly lower, baseline relation formation effects than do whites.

We may at this point wish to revisit the standard Markov model (model 2) and reduced Markov model (model 3), combining each with the race/grade terms (models 13 and 14, respectively) to determine whether the presence of these heterogeneous mixing terms helps to eliminate the non-convergence of the model. It does not; the triangle parameter continues to run to infinity.

With our base model in place, we now open up the process to explore a wider variety of terms. These include:

- Model 15: $g(y) = \{$race/grade base model, $T(y)\}$;
- Model 16: $g(y) = \{$race/grade base model, $D_k(y)\}$ for $k \in \{1:8\}$;
- Model 17: $g(y) = \{$race/grade base model, $S_k(y)\}$ for $k \in \{2:8\}$;
- Model 18: $g(y) = \{$race/grade base model, $M_{sex,a}(y)\}$ for $a \in \{$female, N/A$\}$, with male as a reference category;
- Model 19: $g(y) = \{$race/grade base model, $U_{sex}(y)\}$;

Table 3
Fit for dyadic dependence models[a]

Model #	Additional model terms	Total terms	AIC	Interpretation
15	$T(\mathbf{y})$	20	[b]	Triangle effect
16	$D_k(\mathbf{y})$, $k \in \{1{:}8\}$	27	12104.5	Non-parametric model for degrees
17	$S_k(\mathbf{y})$, $k \in \{2{:}8\}$	26	[b]	Non-parametric model for stars
18	$M_{\text{sex}}(\mathbf{y})$	21	16852.8	Differential levels of edge formation by sex
19	$U_{\text{sex}}(\mathbf{y})$	20	16775.1	Uniform assortative mixing by sex
20	$U_{\text{building}}(\mathbf{y})$	20	16691.9	Uniform assortative mixing by building (junior high vs. high school)
21	$A_{\text{grade}}(\mathbf{y})$	20	16670.9	Absolute difference effect for grade (linear drop-off in log-odds of tie formation with increasing difference in grade
22	$EP_k(\mathbf{y})$, $k \in \{0{:}3\}$	23	[c]	Non-parametric model for edgewise shared partners—a form of clustering
23	$u(\mathbf{y},0.5)$	20	16462.0	GWD (a parameterized form of the degree distribution)
24	$v(\mathbf{y}, 0.5)$	20	11242.0	GWESP (a parameterized form of edgewise shared partners
25	$w(\mathbf{y}, 0.5)$	20	[c]	GWDSP (a parameterized form of dyadwise shared partners or multiple twopaths)

[a] All models contain $L(\mathbf{y})$, $H_{\text{grade},a}(\mathbf{y})$, $H_{\text{race},a}(\mathbf{y})$, $M_{\text{grade},a}(\mathbf{y})$, $M_{\text{race},a}(\mathbf{y})$ in addition to the model terms explained in the table body.
[b] Model did not converge in any of multiple runs.
[c] Model converged on maximum likelihood parameters but did not yield an estimate for the likelihood value.

- Model 20: $\mathbf{g}(\mathbf{y}) = \{$race/grade base model, $U_{\text{building}}(\mathbf{y})\}$, i.e. junior high versus high school;
- Model 21: $\mathbf{g}(\mathbf{y}) = \{$race/grade base model, $A_v(\mathbf{y})\}$ for grade;
- Model 22: $\mathbf{g}(\mathbf{y}) = \{$race/grade base model, $EP_k(\mathbf{y})\}$ for $k \in \{0{:}3\}$;
- Model 23: $\mathbf{g}(\mathbf{y}) = \{$race/grade base model, $u(\mathbf{y}, 0.5)\}$;
- Model 24: $\mathbf{g}(\mathbf{y}) = \{$race/grade base model, $v(\mathbf{y}, 0.5)\}$;
- Model 25: $\mathbf{g}(\mathbf{y}) = \{$race/grade base model, $w(\mathbf{y}, 0.5)\}$.

The first round, adding each of these terms or sets of terms, yields the model fits in Table 3. The standard Markov terms of triangle and individual star parameters again led to a non-converging model. Of those that did converge, all represent an improvement in AIC. The addition of the clustering term $v(y)$ is by far the strongest of these (AIC = 11242.0), with a saturated non-parametric version of the degree distribution $D_k(\mathbf{y})$ in second place (AIC = 12104.5). We add the $v(y)$ term to our model, and include this new model's parameters in Table 2 and goodness of fit plots in Fig. 2b.

Note that when the above models are run repeatedly, the estimates for the log-likelihood for both the $v(y)$ and $D_k(\mathbf{y})$ models varied widely (on the order of 10^1), even though the estimates for the parameter values are highly consistent. The value for $v(y)$ is always lower, justifying its selection as the best-fitting additional term; however, the methods used for approximating the value of the likelihood begin to lose precision at this point of model complexity; subsequent model comparison is done using the goodness of fit plots alone.

Our current best-fitting model, model 24 captures most features of the higher-order network structure. This include the edgewise shared partner distribution, which is perhaps not surprising since the model contains a term for it, although again this single parameter model was not guaranteed to fit the entire distribution. The degree distribution is largely fit, with the exception of the lowest terms (0–2), which are still slightly off. Most of the geodesic distribution is fit, although the tail end of the observed distribution diverges from the model prediction. Note, however, that these graphs are on the log scale, so the visual effect of these low-probability events is greatly
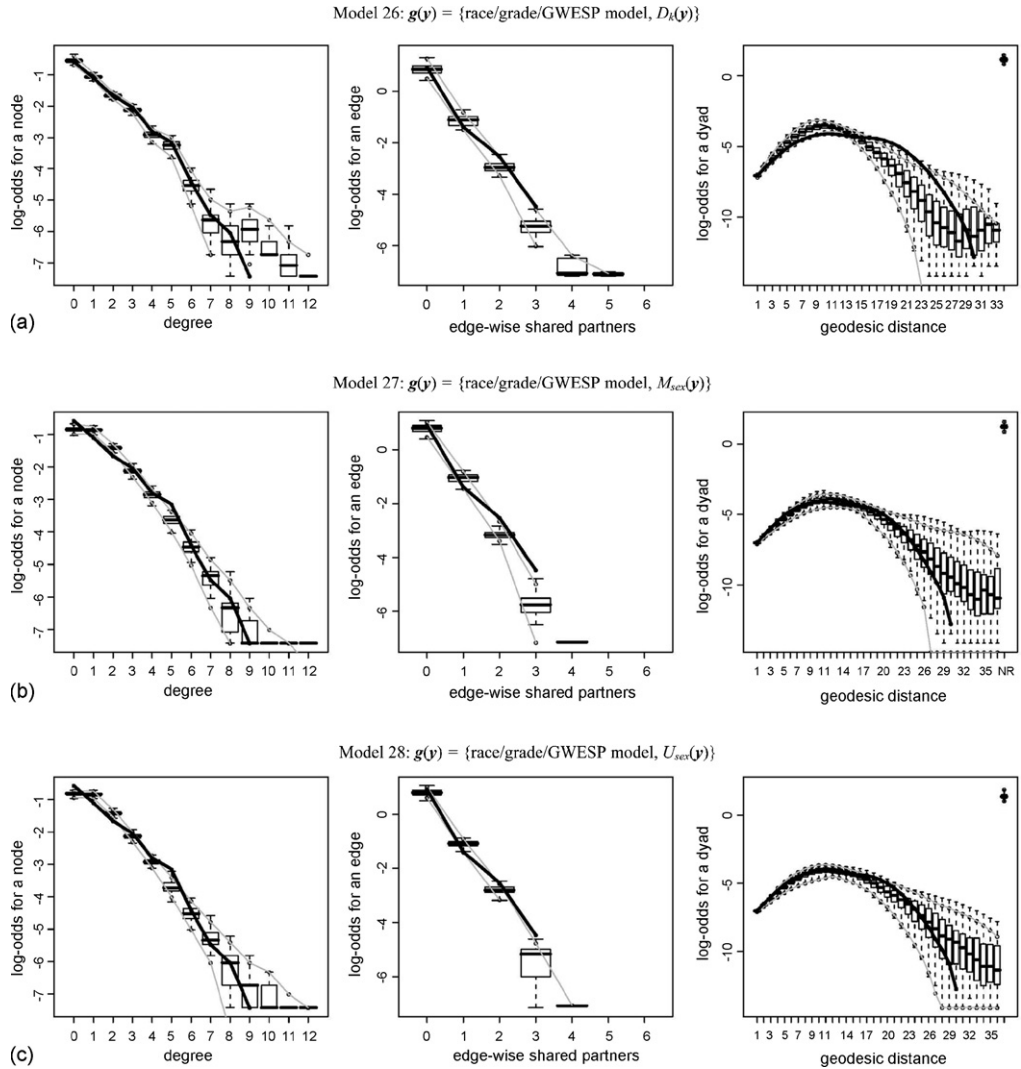

Fig. 3. Goodness-of-fit boxplots for select models (see Fig. 1 for explanation). (a) Model 26: $g(y) = \{$race/grade/GWESP model, $D_k(y)\}$. (b) Model 27: $g(y) = \{$race/grade/GWESP model, $M_{sex}(y)\}$. (c) Model 28: $g(y) = \{$race/grade/GWESP model, $U_{sex}(y)\}$. (d) Model 29: $g(y) = \{$race/grade/GWESP model, $U_{building}(y)\}$. (e) Model 30: $g(y) = \{$race/grade/GWESP model, $A_{grade}(y)\}$. (f) Model 31: $g(y) = \{$race/grade/GWESP model, $u(y,0.5)\}$. (g) Model 32: $g(y) = \{$race/grade/GWESP model, $w(y, 0.5)\}$.
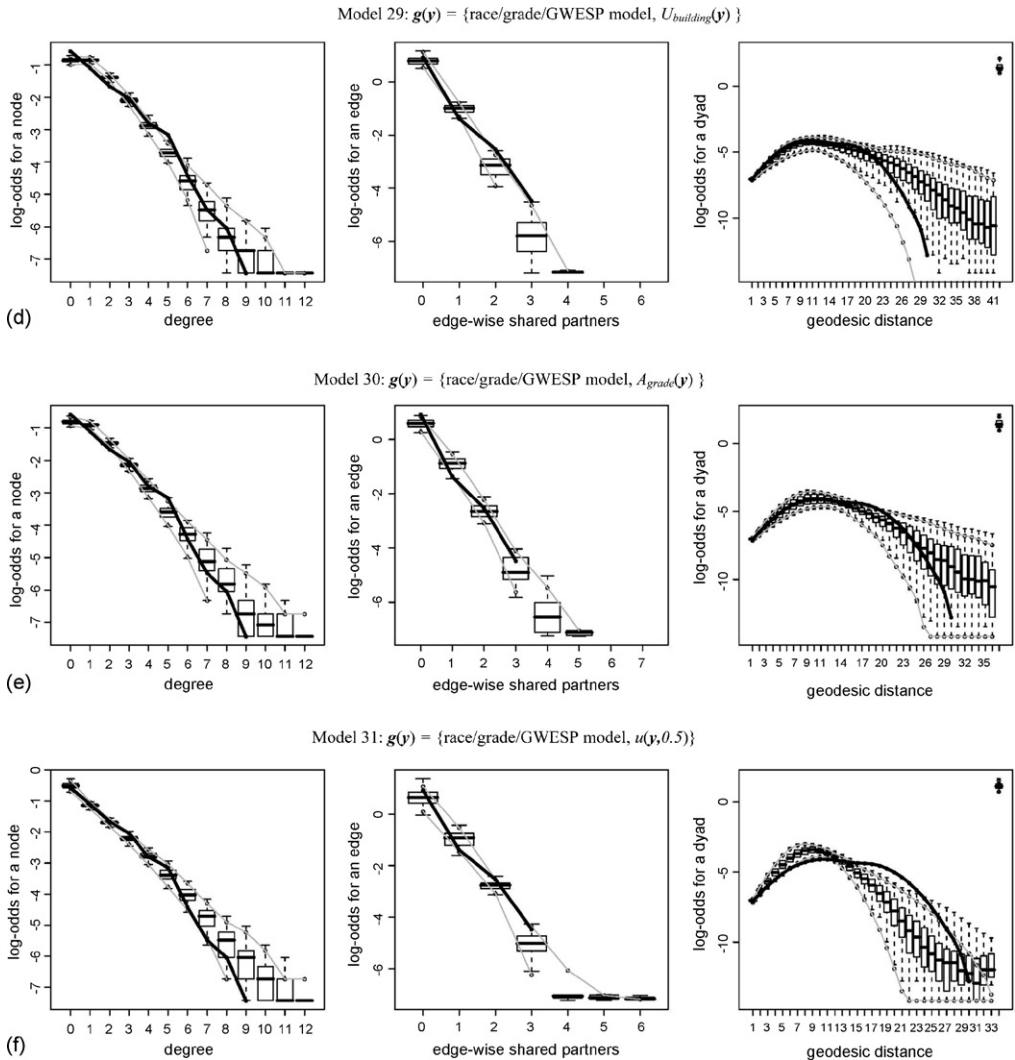
Fig. 3. (*Continued*)

heightened. A log-odds of -12 represents only 8 dyads in this network of more than 1.4 million dyads, so only a very small fraction of the data are included in this tail. Nevertheless, a complete model would be expected to capture this feature of the original data as well as any other feature.

In the next round, we re-consider all of the terms that did not enter in the last round, excluding the triangle statistic $T(y)$ and the individual edgewise shared partner statistics $EP_k(y)$, given the previous inclusion of $v(y, \theta_2)$. The model containing individual star terms $S_k(y)$ is again degenerate; goodness of fit plots for all of the remainder are shown in Fig. 3.

Adding parametric terms for the degree distribution (model 26) does fix the problem that our previous model did not accurately predict the $0°$ and $1°$. However, in the process it wreaks havoc on the geodesic distribution; that is, it improves the fit of the local structure but mis-specifies the global structure in the process. From visual inspection of the goodness-of-fit plots, the model that
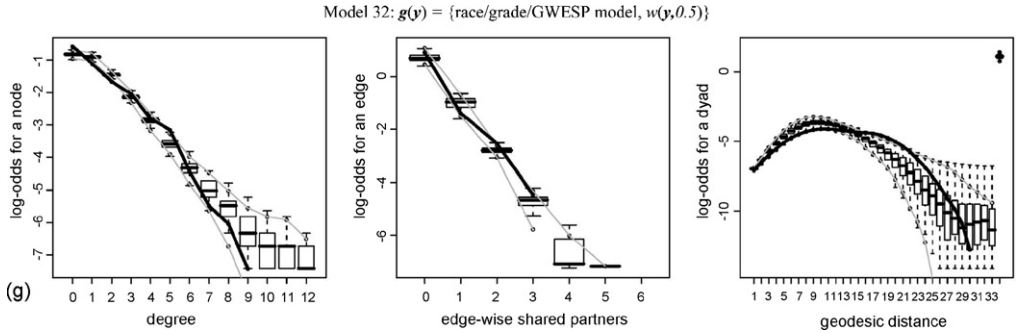
Model 32: $g(y) = \{$race/grade/GWESP model, $w(y,0.5)\}$

Fig. 3. (*Continued*).

most improves fit includes edge statistics for the different values of the sex attribute (model 27). Note that there are actually three values for sex in the data (male, female, plus N/A for those students who let the question blank) so that an edge term by sex adds two statistics to the model (female, N/A, with male as the reference category). These terms captures the fact that female students have significantly more mutual friends in these data than do male (i.e. $M_{sex,female}(y)$ is significant and positive), while students of unknown sex are intermediate. The addition of this simple component extends out the range over which the geodesic distribution is well-captured; in the previously best-fitting model (model 24), the data and model appear to diverge around geodesic length 16, while in this model the divergence begins more around 24.

We have necessarily selected only a short list of network statistics to compare between observed data and model predictions. One may choose to compare others as well. As examples, we provide plots for two additional statistics in Fig. 4: the total number of triangles in the network (Fig. 4A) and the size of the largest component (Fig. 4B). Each of these statistics represent a single measure on the graph as opposed to a distribution across elements within the graph; thus these comparisons can be made using simple histograms. For both statistics the observed data (157 triangles, 775
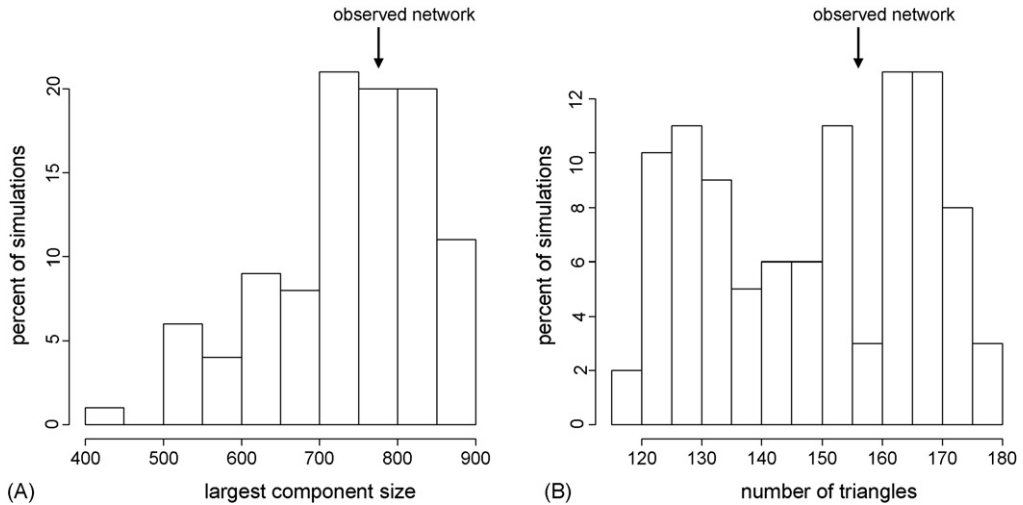

Fig. 4. Additional goodness of fit measures, model 27.

actors in the largest component) fall well into the mass of the predictions made by model 27, providing further evidence for model fit.

Additional iterations of model-building for this set of terms, using model 27 as a base, do not appear to further advance model fit, and are not shown.

6. Discussion

The advances discussed in earlier papers in this issue have placed generalized statistical inference for dependence models in large social networks on a firmer footing than have previously existed. In the process of applying these models to the friendship network in this paper, we have obtained some insight into the underlying social processes that could (or could not) have generated that set of friendships. First, there appears to be both exogenous (attribute-based) and endogenous (shared partner) processes of friend selection at work among these students, and neither alone is capable of capturing all aspects of the network structure. We can use the values of the parameters (which represent log-odds ratios for a tie) to see that the bias towards within-grade ties is generally stronger than for within-race ties, although both types are strong and significant. The magnitude of the parameter for $v(y, \theta_2)$ is not easy to compare directly with the others given the complexity of the shared partner formulation; however, this effect is significant. Terms that did not seem to contribute substantially to model fit include the degree terms (parametric or non-parametric) and the dyadwise shared partner terms. In short, to capture most of the global structure for these friendships, one needed only relatively local phenomena—attribute mixing and patterns of shared partners or k-triangles. Information on attribute-based mixing can be obtained from local network studies using standard techniques for sampling respondents. Information on patterns of shared partners can be obtained either from asking respondents how many partners they have in common with each of their partners (which may or may not be reliable), or through a one-step contact trace, both of which are considerably easier and cheaper than a complete census of a network for most real-world populations. If the results here are found to be general, and much of the global structure of networks is determined at these local or semi-local levels, then this finding bodes well for the future analysis of empirical social network structure, as well as its determinants and effects, using sampled egocentric data.

In addition to learning about those friendships in this school, we also learned much about the various parameterizations of ERG models. The results here are in agreement with recent work showing that the commonly used Markov model is degenerate for empirical networks that are not very small. The alternative forms proposed by Snijders et al. (in press) and Hunter and Handcock (2006), which have similar underlying interpretations but more robust properties, not only avoid degeneracy, but are validated as empirically useful. These parameterizations, combined with advances in computational algorithms, now allow us to conduct general statistical inference on networks between one and two orders of magnitude larger than those that social network analysts have long been used to considering. Although numerous refinements are doubtless still to come, it appears that the long-awaited promise of statistical network analysis may finally have arrived.

References

Besag, J.E., 1974. Spatial interaction and the statistical analysis of lattice systems. Journal of the Royal Statistical Society B 36, 192–236.

Besag, J.E., 1986. On the statistical analysis of dirty pictures. Journal of the Royal Statistical Society Series B 48, 259–302.

Frank, O., Strauss, D., 1986. Markov graphs. Journal of the American Statistical Association 81, 832–842.

S.M. Goodreau / Social Networks 29 (2007) 231–248

Geyer, C.J., 1994. On the convergence of Monte Carlo maximum likelihood calculations. Journal of the Royal Statistical Society Series B 56, 261–274.

Geyer, C.J., Thompson, E.A., 1992. Constrained Monte Carlo maximum likelihood for dependent data. Journal of the Royal Statistical Society Series B 54, 657–699.

Goodreau, S.M., Hunter, D.R., Morris, M., 2005. Statistical Modeling of Social Networks: Practical Advances and Results. Center for Studies in Demography and Ecology, University of Washington, Working Paper No. 05-01.

Handcock, M.S., 2002. Statistical models for social networks: degeneracy and inference. In: Breiger, R., Carley, K., Pattison, P. (Eds.), Dynamic Social Network Modeling and Analysis. National Academies Press, Washington, DC, pp. 229–240.

Handcock, M.S., 2003. Assessing Degeneracy in Statistical Models of Social Networks. Center for Statistics and the Social Sciences, University of Washington, Working Paper no. 39.

Handcock, M.S., Hunter, D.R., Butts, C.T., Goodreau, S.M., Morris, M., 2004. ergm: An R Package for the Statistical Modeling of Social Networks. Center for Studies in Demography and Ecology, University of Washington, unpublished manual (http://csde.washington.edu/statnet/statnetmanual-0.8.pdf).

Hunter, D.R., 2006. Curved Exponential Family Models for Social Networks. Social Networks, doi:10.1016/j.socnet.2006.08.005.

Hunter, D.R., Handcock, M.S., 2006. Inference in curved exponential family models for networks. Journal of Computational and Graphical Statistics 15, 565–583.

Hunter, D.R., Goodreau, S.M., Handcock, M.S., 2005. Goodness of Fit of Social Network Models. Center for Statistics and the Social Sciences, University of Washington, Working Paper No. 47.

Pattison, P.E., Robins, G.L., 2002. Neighbourhood-based models for social networks. Sociological Methodology 32, 301–337.

Resnick, M.D., Bearman, P.S., Blum, R.W., Bauman, K.E., Harris, K.M., Jones, J., Tabor, J., Beuhring, T., Sieving, R.E., Shew, M., Ireland, M., Bearinger, L.H., Udry, J.R., 1997. Protecting adolescents from harm. Findings from the National Longitudinal Study on Adolescent Health. Journal of the American Medical Association 278, 823–832.

Robins, G., Pattison, P., Kalish, Y., Lusher, D., 2007a. An introduction to exponential random graph (p^*) models for social networks. Social Network 29, 173–191.

Robins, G., Snijders, T.A.B., Wang, P., Handcock, M.S., Pattison, P., 2007b. Recent developments in exponential random graph (p^*) models for social networks. Social Networks 29, 192–215.

Snijders, T.A.B., 2002. Markov chain Monte Carlo estimation of exponential random graph models. Journal of Social Structure 3, 2.

Snijders, T.A.B., van Duijn, M.A.J., 2002. Conditional maximum likelihood estimation under various specifications of exponential random graph models. In: Hagberg, J. (Ed.), Contributions to Social Network Analysis, Information Theory, and Other Topics in Statistics: A Festschrift in Honour of Ove Frank. University of Stockholm, Department of Statistics, Stockholm, pp. 117–134.

Snijders, T.A.B., Pattison, P., Robins, G.L., Handcock, M., in press. New specifications for exponential random graph models. Sociological Methodology, 36.

Strauss, D., Ikeda, M., 1990. Pseudo-likelihood estimation for social networks. Journal of the American Statistical Association 85, 204–212.

Udry, J.R., Bearman, P.S., 1998. New methods for new research on adolescent sexual behavior. In: Jessor, R. (Ed.), New Perspectives on Adolescent Risk Behavior. Cambridge University Press, New York, pp. 241–269.

Wasserman, S., Pattison, P., 1996. Logit models and logistic regressions for social networks. I. An introduction to Markov graphs and p^*. Psychometrika 60, 401–425.