(Hunter, Goodreau, and Handcock 2008) & (Hunter et al. 2008)

Hunter, David R., Steven M. Goodreau, and Mark S. Handcock. 2008. “Goodness of Fit of Social Network Models.” *Journal of the American Statistical Association* 103 (481): 248–58. https://doi.org/10.1198/016214507000000446.

Hunter, David R, Mark S Handcock, Carter T Butts, Steven M Goodreau, and Martina Morris. 2008. “Ergm: A Package to Fit, Simulate and Diagnose Exponential-Family Models for Networks.” *Journal of Statistical Software* 24: 1–29. http://www.jstatsoft.org/.

Lusher, Dean, Johan Koskinen, and Garry Robins. 2013. *Exponential Random Graph Models for Social Networks: Theory, Methods, and Applications*. Edited by Dean Lusher, Johan Koskinen, and Garry Robins. *Exponential Random Graph Models for Social Networks*. Cambridge: Cambridge University Press.

**Hunter, Goodreau, and Handcock 2008**:

Assess ERGMs and maximum likelihood to fit models using a friendship network.

Markov chain Monte Carlo (MCMC) is a technique that stimulates a large number of random networks. MCMC maximum likelihood estimation is a stochastic approximation to the likelihood function. Model networks are different from the observed network.

Calculating approximate maximum likelihood estimates (MLEs) using ERGM parameters. MLEs for dyadic independence models are obtained using logistic regression. Inference relies on maximum pseudolikelihood estimation, which uses a standard logistic regression algorithm. Standard error interpretations are not reasonable estimates of the standard deviations of the pseudolikelihood estimators.   
Hunter, Goodreau, and Handcock (2008) begin with a null model and then ass nodal covariates to create a dyadic independence model. The nodal factor effect counts the total number of endpoints with the particular level of a particular factor for each edge in the network. “The homophily statistic for a particular factor gives each edge in the network a score of 0 or 1, depending on whether the two endpoints have matching values of the factor.” Uniform homophily gives a single statistic, whereas differential homophily gives a set of statistics. [Include only a differential homophily or a nodal factor effect for one factor but not both? Or is this only when it is a two-level factor that there is a restriction? What about for a nine factor nodal covariate? Is it appropriate to include differential homophily and nodal factor for one type of nodal covariate?] Ordinal categorical variable is considered . . .  
Goodness-of-fit compares observed network statistics with simulated to determine which structural aspects are important in assessing fit.

Degree distribution, edgewise shared partner distribution, . . . geodesic distance between two nodes equals the length of the shortest path joining those two nodes.

The simplest dyadic dependence model consists only of a subset of the degrees statistics. “The k-star statistics are highly collinear with one another.”

There are two distinct sets of shared partner statistics, the edgewise shared partner statistics and the dyadwise shared partner statistics. For the edgewise shared partner statistics: “We count all of the shared partners for all edges, then we have counted each triangle three times, once for each of its edges.” “K-triangle is defined to be a set of k distinct triangles that share a common edge.”

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.” Frank and Strauss (1986) proposed the Markov random graphs. “Markov random graphs treat all nodes as equivalent, ignoring any covariate information.” “Markov random graph models fail to empirically describe social network data.” The simulated networks need to resemble the observed network for MLE.

Snijder et al. (2006) developed alternating k-triangle, k-towpath, and k-star statistical methods that build dependence models that fit network data sets well enough to make reliable MLEs. If the dependence terms are statistically significant, then we cannot ignore the network structure.

Geometrically weighted degree, geometrically weighted edgewise shared partner, and geometrically weighted dyadwise shared partner “capture high-order dependency structure in networks in a parsimonious fashion while avoiding the problems of degeneracy” making models that include these statistics more robust. MCMC fitting procedure aids in estimating parameters.   
If the observed network is not typical of the simulated network of a particular statistic, then the model is either degenerate or poorly fitted.

“Social relations generally exhibit local clustering” “Neither homophily nor shared partners alone is sufficient to explain the clustering observed in the study’s friendship network.”   
The degree distribution is modeled by the GWD statistic. “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.”

The alternating k-star statistic

Hunter, Goodreau, and Handcock (2008) explain that by augmenting dependence models with covariate-only terms, one can assess whether the dependence terms are statistically significant. If they are not, then independence models might suffice, indicating that the network analysis could proceed by ignoring the network structure and using logistic regression on the independent responses. Their analysis reveals that dyadic dependence persists even after accounting for nodal covariate information.

Hunter, Goodreau, and Handcock (2008) also discuss the challenge of fitting many dyadic dependence models due to severe numerical difficulties in estimation. To determine how well a model fits, Hunter, Goodreau, and Handcock (2008) recommend comparing observed network statistics with those obtained from simulating many networks using the fitted ERGM. If the observed network significantly deviates from the simulated networks for a particular statistic, it indicates either model degeneracy or poor fit. For instance, while individual-level attributes can recreate some global properties of the network, such as geodesic distribution, they may underestimate local clustering as captured by shared partner distribution. This finding underscores the necessity of testing various network statistics to develop a comprehensive understanding of model fit.

Goodreau et al. (2009) provide a detailed exploration of goodness of fit (GoF) of the adolescent social networks models. Their methodology involves comparing observed network data with simulated networks generated from ERGMs using estimated coefficients. This comparison is visualized in GoF plots for various network statistics.

**Hunter et al. 2008**:

ERGMs describes the local forces that shape the whole network.

The propensity for individuals to form triangles of grant proposals. When looking at a network visualization, nodes appear to cluster in groups, and ERGM quantifies the strength of this intra-group effect. Group membership is not a measured attribute of the node but a structure of relations dependent on more general neighborhood effects. [ERGMs allow me to investigate long paths and cycles.]

Exogenous covariates and dyadic independence are node attributes incorporated into an ERGM.

`Nodematch` is called differential homophily for [attribute name] `nodefactor` is a main effect of [attribute name]. \*Interpretation in section 3.2\*

“A dyad in a network is a random variable representing the state of the relationship(s) between two given nodes.” “A Dyadic independence term is a term in an ERGM for which the corresponding network change statistic(s) in the change statistic vector” may always be calculated without knowing anything about the particular network.

ERGM models should have linear independence among the terms in an ERGM.

“A dyadic independence ERGM is an ERGM whose terms are dyadic independence terms.”

Dyadic dependence models:

“Dyads that do not share a node are conditionally independent”

[Local configurations are nested where a single tie between two nodes forms a dyad, a node with two ties is a 2-star, and three nodes that are tied together form a triad (Lusher, Koskinen, and Robins 2013, 22).]

A degree distribution parameterization counts k-stars, making the degree statistics directly interpretable regarding the concurrency of partnerships.

Curved exponential-family models include the ERGM terms geometrically weighted degree (GWD) and geometrically weighted edgewise shared partner (GWESP). These terms are linear combinations of an entire distribution of degree or shared partner statistics.

GWESP

[What is edgewise shared partners (ESP)?]

GWESP

What does geometrically weighted mean?

[A positive transitivity parameter indicates a tendency of triadic closer (Lusher, Koskinen, and Robins 2013, 74) or a higher number of triangles in the network. Suppose the observed network’s triangles are near the mean of the simulated networks’ triangles. In that case, it implies that the observed network’s transitivity is not significantly different from what the null model would predict.]

@software{your-key-here,

title = {Statnet: Tools for the Statistical Modeling of Network Data},

author = {Pavel N. Krivitsky, Mark S. Handcock and Hunter, David R. and Butts,

Carter T. and Klumb, Chad and Goodreau, Steven M. and Morris, Martina},

date = {2003/2023},

url = {https://statnet.org},

organization = {Statnet Development Team}

}