Graphical Models for Network Data

Stephen E. Fienberg

Department of Statistics, Living Analytics Research Centre, Heinz College, Machine Learning Department, and Cylab
Carnegie Mellon University

(Based on joint work with Sonja Petrović, Alessandro Rinaldo, and Xiaolin Yang)

Fields Institute Workshop on Graphical Models
April 16, 2012

Graphs as Metaphors

- Representation of statistical structures in terms of graphs $G = \{V, E\}$, is a useful metaphor that allows us to exploit the mathematical language of graph theory and some relatively simple results.
- Graphs often provide powerful representations for the interpretation of models.
- Vertices and edges have different meaning in different statistical settings.

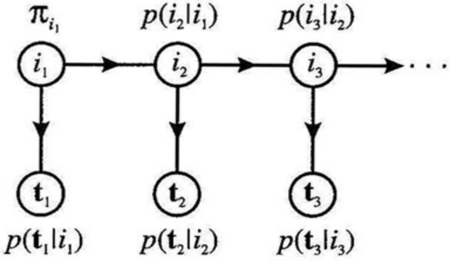
Graphical Representations of Statistical Models

	Variables	Individuals
Directed	a	b
Undirected	c	d

- a—HMMs, state-space models, Bayes nets, causal models (DAGs), recursive partitioning models
- b—social networks, trees, citation and email networks
- c—covariance selection models, log-linear models, multivariate time-series models
- d—relational networks, co-authorship networks

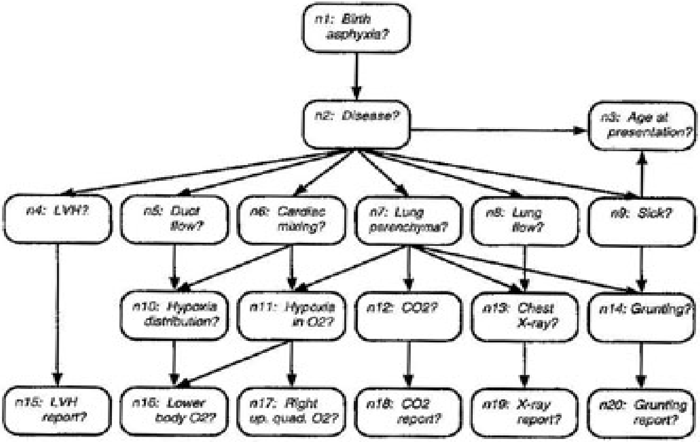
Note that a and c refer to probability models, while b and d are used to describe observed data.

a—HMMs, State-Space Models



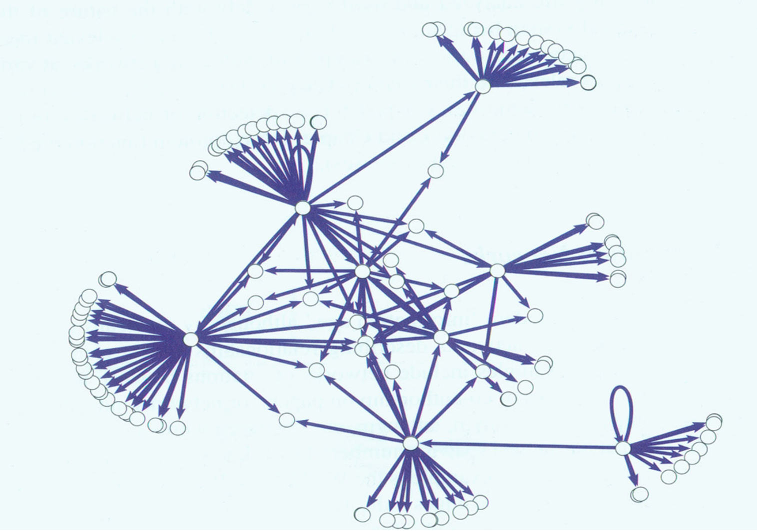
a—Causal Models, DAGs

- CHILD network (blue babies) (Cowell et al.,1999)



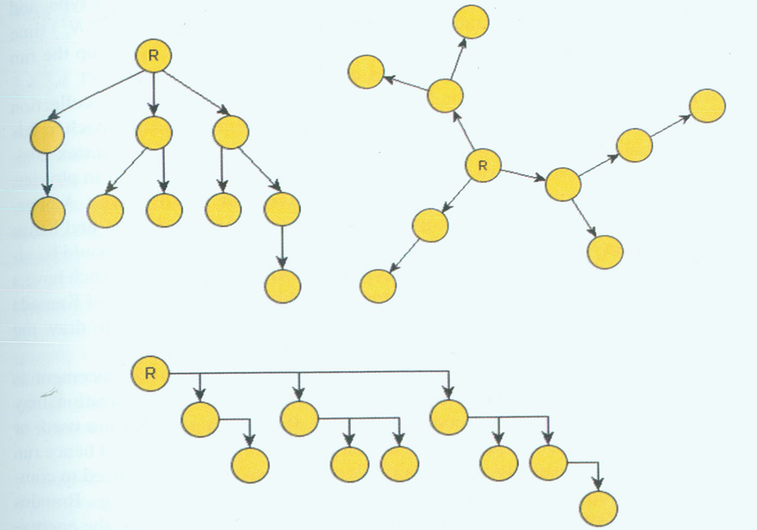
b—Social Networks

- AIDS blog network (Kolaczyk, 2009)



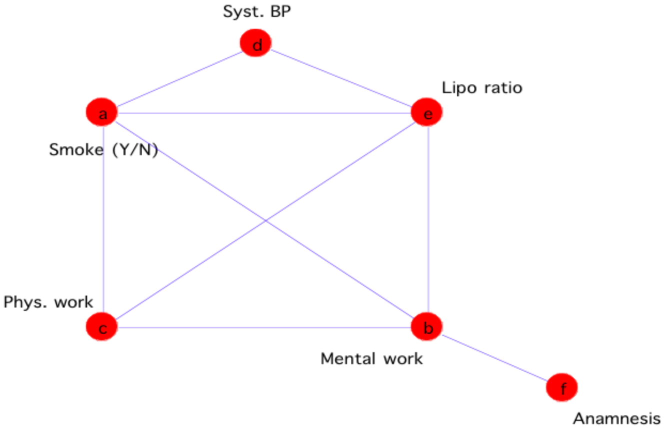
b—Trees

- Ancestral Trees (Kolaczyk, 2009)

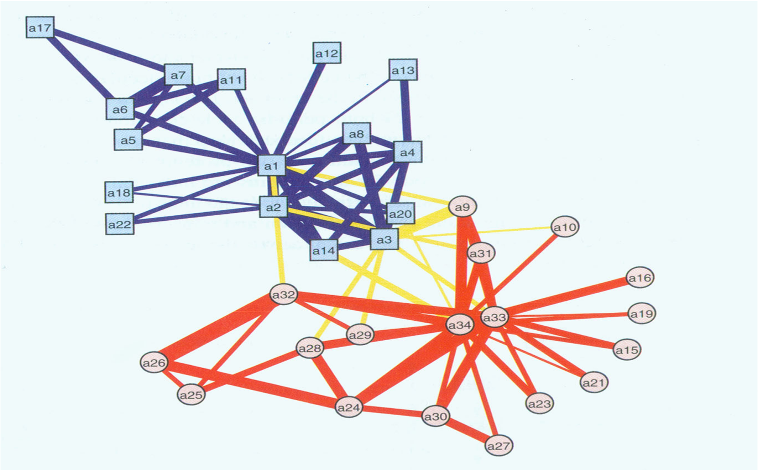


c—Log-linear Models

- Prognostic factors for coronary heart disease for Czech autoworkers—2^6 table (Edwards and Hrvanek, 1985)

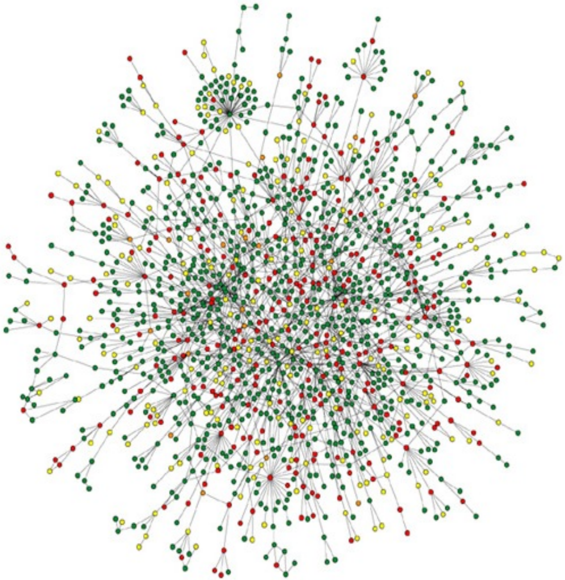


- Zachary's "karate club" network (Zachary, 1977; Kolaczyk, 2009)



d—Yeast Protein-Protein Interaction

- Airoldi et al. (2008)



Graphical Models for Variables—Rest of Workshop!!

- The following Markov conditions are equivalent:
  - Pairwise Markov Property: For all nonadjacent pairs of vertices, i and j, $i \perp j \mid K \setminus \{i, j\}$.
  - Global Markov Property: For all triples of disjoint subsets of K, whenever a and b are separated by c in the graph, $a \perp b \mid c$.
  - Local Markov Property: For every vertex i, if c is the boundary set of i, i.e., $c = bd(i)$, and $b = K \setminus \{i \cup c\}$, then $i \perp b \mid c$.
- All discrete graphical models are log-linear.
- Gaussian graphical model selection problem involves estimating the zero-pattern of the inverse covariance or concentration matrix.
- For DAGs, we continue to use Markov properties but also exploit partial ordering of variables.
- Always assume individuals or units are independent r.v.'s.

Models for Individuals/Units in Networks

- Graph describes *observed* adjacency matrix.
  - Usually use 1 for presence of an edge, and 0 for absence.
  - Can also have weights in place of 1's.
- Except for Erdös-Rényi-Gilbert model, where occurrence of edges corresponds to iid Bernoulli r.v.'s, units are *dependent.*
- Simplest generalization of E-R-G model assumes that dyads are independent—e.g., the p_1 model of Holland and Leinhardt, which has additional parameters for reciprocation in directed networks.
- Exponential Random Graph Models (ERGMs) that include "star" and "triangle" motifs no longer have dyadic independence.
- Can have multiple relationships measure on same individuals/units.

Erdös-Rényi-Gilbert Model

- In $G(n, M)$ model, we chose graph uniformly at random from the collection of all graphs which have n nodes and M edges—hypergeometric distribution associated with the degree of a node.
- In $G(n, p)$ model, we connect nodes in graph independently, with constant probability p, Now M is random and has a binomial distribution with probability $\binom{\binom{n}{2}}{M} p^M (1-p)^{\binom{n}{2}-M}$.
- interesting probability structure, especially as as n and M get large, but not much of interest statistically for a fixed n or M since basically we are in a simple distributional setting.
- This changes when we let p vary depending on the nodes it connects, and especially when we allow edges to be directed and dependent.

Holland and Leinhardt p_1 model

- n nodes, random occurrence of directed edges.
- Describe the probability of an edge occurring between nodes i and j:
  - $\log P_{ij}(0,0) = \lambda_{ij}$
  - $\log P_{ij}(1,0) = \lambda_{ij} + \alpha_i + \beta_j + \theta$
  - $\log P_{ij}(0,1) = \lambda_{ij} + \alpha_j + \beta_i + \theta$
  - $\log P_{ij}(1,1) = \lambda_{ij} + \alpha_i + \beta_j + \alpha_j + \beta_i + 2\theta + \rho_{ij}$
- 3 common forms:
  - $\rho_{ij} = 0$ (no reciprocal effect)
  - $\rho_{ij} = \rho$ (constant reciprocation factor)
  - $\rho_{ij} = \rho + \rho_i + \rho_j$ (edge-dependent reciprocation)
- When edges are undirected, p_1 reduces to the beta model.

Estimation for p_1

- The likelihood function for the p_1 model is clearly of exponential family form.
- For the constant reciprocation version, we have

$$\log p_1(x) \propto x_{++}\theta + \sum_i x_{i+}\alpha_i + \sum_j x_{+j}\beta_j + \sum_{ij} x_{ij}x_{ji}\rho \qquad (1)$$

- Get MLEs using iterative proportional fitting—method scales.
- Holland-Leinhardt explored goodness of fit of model empirically by comparing $\rho_{ij} = 0$ vs. $\rho_{ij} = \rho$.
  - Standard asymptotics (normality and χ^2 tests) aren't applicable; no. parameters increases with no. of nodes.
- Fienberg and Wasserman (1981) use edge-dependent reciprocation model to test $\rho_{ij} = \rho$.
- *Algebraic Statistics*: Petrović et al. (2010) provide Markov bases; Rinaldo et al. (2011) characterize MLE existence.
- Goldenberg et al. (2010) review these and related models.

Exponential Random Graph Models (ERGMs)

- Let X be a $n \times n$ adjacency matrix or a 0-1 vector of length $\binom{n}{2}$ or a point in $\{0,1\}^n$).
- Identify a set of *network statistics*

$$t = (t_1(X), \ldots, t_k(X)) \in \mathbb{R}^k$$

  and construct a distribution such that t is a vector of *sufficient statistics*.
- This leads to an *exponential family* model of the form:

$$P_\theta(X = x) = h(x) \exp\{\theta \cdot t - \psi(\theta)\},$$

  where
  - $\theta \in \Theta \subseteq \mathbb{R}^k$ is the natural parameter space;
  - $\psi(\theta)$ is a normalizing constant (often intractable);
  - $h(\cdot)$ depends on x only.

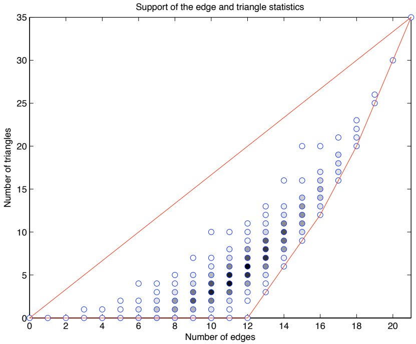
Likelihood Methods for ERGMs

- Likelihood methods are more complex than exponential family structure might suggest.
- Pseudo-estimation using independent logistic regressions, one per node.
- Can get MLEs via MCMC.
- Problem of degeneracy or near degeneracy:
  - MLEs don't exist—maximize on the boundary.
  - Likelihood function is not well-behaved and most observable configurations are near the boundary.

ERGMs: 7-node Example–I

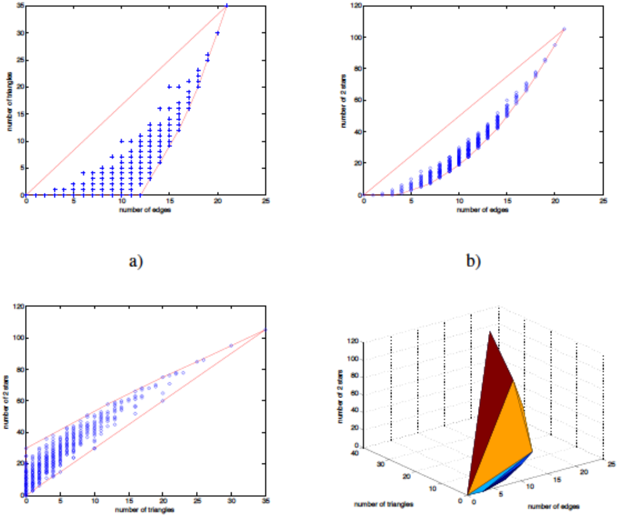
Set of all graphs on 7 nodes when the sufficient statistics are the number of edges and number of triangles.

- There are $2^{21} = 2,097,152$ possible graphs!
- There are only 110 different configurations for the 2-dimensional sufficient statistics.
- Note: many points on the boundary (including the empty and complete graph)



ERGMs: 7-node Example–II

Consider the set of all graphs on 7 nodes when the sufficient statistics are the no. of edges, no. 2-stars, and no. of triangles.



Inference for Models for Individuals/Units in Networks

- Relevant asymptotics has number of nodes, $n \to \infty$.
- When there are node-specific parameters, asymptotics are far more complex.
- Maximum likelihood approaches available for ERGMs.
- For blockmodels, with constant structure within blocks, there is asymptotic theory.
  - Related literature on "community formation" and "modularity." Bickel and Chen (2009)

Pseudo-likelihood for ERGMs

- Frank and Strauss (1986) and Strauss and Ikeda (1990), following ideas of Besag and work on Markov random field models, considered conditional probability $P(X_{ij} = 1|X_{ij}^c)$ where X_{ij}^c is the graph after removing edge (i, j).

$$
\begin{aligned}
P(X_{ij} = 1|X_{ij}^c) &= \frac{\exp\left[\theta \cdot (T(X_{ij}^+) - T(X_{ij}^-))\right]}{1 + \exp\left[\theta \cdot (T(X_{ij}^+) - T(X_{ij}^-))\right]} \\
&= \frac{\exp\left[\theta \cdot \delta(X_{ij}^c)\right]}{1 + \exp\left[\theta \cdot \delta(X_{ij}^c)\right]}
\end{aligned}
$$

  where X_{ij}^+ and X_{ij}^- represent graphs setting $X_{ij} = 1$ or 0, X_{ij}^c denotes remainder of network , and $\delta(x_{ij}^c)$ is change of SSs when x_{ij} changes from 0 to 1.
- This has form of logistic regression model.

Maximum Pseudo-likelihood Estimation for ERGMs

- Pseudo-likelihood treats logistic regression components as if they were independent and sums over all edges:

$$l_P(\theta; x) = \theta \cdot \sum_{ij} \delta(x_{ij}^c) x_{ij} - \sum_{ij} \log(1 + \exp(\theta^T \delta(x_{ij}^c))) \qquad (2)$$

- Simple Markov basis structure for pseudo-likelihood. Fienberg, Petrović, and Rinaldo (unpub.)

Theorem (Yang, Fienberg, and Rinaldo (unpub.))

The existence of the MPLE implies the existence of the MLE. The converse is false.

- *Implications*:
  - If we use MPLEs we may act as if the likelihood is well-behaved when it is not.
  - Even if MLEs exist, actual values of MPLEs and MLEs can differ substantially.

MLE vs. MPLE—7 node Example

	MPLE	NMPLE	MPLE-EN
MLE	129	22	101
NMLE	0	138	0

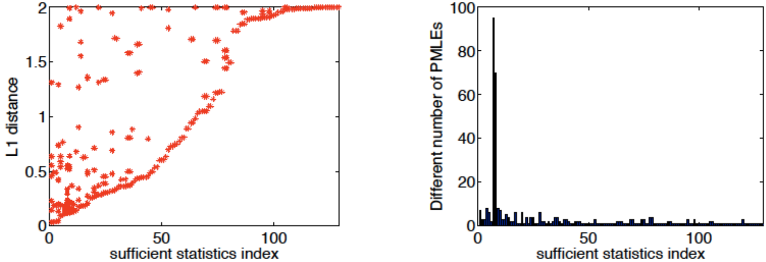


Figure 2: 7-node graph: a). The L_1 difference between the distribution under the MLE and MPLE estimated from graphs with the same sufficient statistics. b). The number of different estimates of MPLE estimated from graphs with the same sufficient statistics corresponding to (a)

Connecting the Two Graphical Approaches

- There is link between graphical models for for variables and graphical models for networks, not just a common metaphor.
- Frank and Strauss (1986) introduce a pairwise Markov property for individual-level undirected network models.
- Key is the construction of the dual graph.

The Network Dual Graph

- *Dual Graph*: Set up conditional independence graph, $G^* = \{V^*, E^*\}$, whose nodes are edges from original graph, $G = \{V, E\}$.
- X_{ij} and $X_{i'j'}$ are conditionally independent given the other r.v.'s X_{kl} iff they do not share a vertex in $G^* = \{V^*, E^*\}$.
- G is *Markov* if G^* contains no edge between disjoint sets (s, t) and (u, v) in V.
- Cliques in a Markov random graph are stars (edges are 1-stars) of various orders and triangles.
- If there are no edges in the dual graph G^*, then we essentially have the beta model.

Undirected Markov Random Graph Models

Theorem

For homogeneous (exchangeable) graphs, distribution of X now satisfies the pairwise Markov property iff

$$Pr\{X = x\} \propto exp\{\sum_{k=1}^{N_V-1} \theta_k S_k(x) + \theta_\tau T(x)\}$$

where $S_k(x)$ is no. of k-stars and $T(x)$ is no. triangles.

- Many other ERGMs don't have this property, e.g., those with alternating *k*-stars and alternating triangles.

Directed Markov Random Graph Models

- Analogous approach to construction of dual graph for situation with directed edges.
- Now the vertices of G^* are paired corresponding to dyads from G.
- Cliques are the original dyads, and various stars and triangular structures.
- If model contains no stars or triangles, it reduces to p_1.

Open Problems

- Some network models have “nice”, non-degenerate behavior.
  - Dyadic independence models such as p_1.
  - Simple blockmodels.
  - Blockmodels that build on dyadic independence structures.
- *Question*: Are there other ERGMs, and in particular Markov Random Graph Models, that are “nice”?
- *Question*: Where does decomposability in dual graph fit in?

Roles for Latent Variables

- For graphical models for variables:
  - Natural for many models, e.g., HMMs.
  - Arise naturally in Hierarchical Bayesian structures. Hyperparameters are latent quantities.
- For models for individuals/units in networks:
  - Random effects versions of node-specific models such as p_1.
  - Arise naturally in hierarchical Bayesian approaches, such as Mixed Membership Stochastic Blockmodels and latent space models.
- Can also use latent structure to infer network links from data on variables for individuals, e.g., as in relational topic models.

Role of Time/Dynamics

- For graphical models for variables:
  - Time gives ordering to variables and assists in causal models.
  - Note distinction between position of underlying “latent” quantity over time and the actual manifest measurement associated with it, which is often measured retrospectively.
- Dynamic models for individual-based networks:
  - Continuous-time stochastic process models for event data, perhaps aggregated into chunks.
  - Discrete-time transition models, perhaps embedded into continuous time process, e.g., see Hanneke et al. (2010)

Drawing Inferences From Subnetworks and Subgraphs

- Inferences from Subgraphs
  - Conditional independence structure allows for local message passing and inference from cliques and regular subgraphs when there are separator sets that isolate components.
  - Interpretation in terms of GLM regression coefficients always depends on the other variables in the model.
- Inferences from Subnetworks
  - Most properties observed in subnetworks don't generalize to full network, and vice versa, e.g., power laws for degree distributions.
  - Problem is dependencies among nodes and boundary effects for subnetworks.
  - Missing edges are generally not missing at random, except for some sampling settings, e.g., see Handcock and Gile (2010).
- The forgoing suggests that we can't use cross-validation for all but simplest network models.

Summary

- Two types of settings:

	Variables	Individuals
Directed	a	b
Undirected	c	d

  - For a and c we use conditional independence ideas to model probabilistic relations among variables.
  - for b and d we use graph to represent observed data.
- Independence comes into play in network settings only of dyadic independence.
- ERGMs have heuristic appeal but often display degenerate behavior.
- Markov Random Graph models invoke the Markov property we inherit from more traditional graphical model settings.
- Whether something nice flows from Markov Random Graph structure remains an open issue.

References

Airoldi, E M., Blei, D. M., Fienberg, S. E., and Xing, E. P. (2008) Mixed Membership Stochastic Blockmodels. *Journal of Machine Learning Research*, **9**, 1981–2014.

Bickel, P. and Chen, A. (2009) A Nonparametric View of Network Models and Newman-Girvan and Other Modularities. *PNAS,* **106** *(50), 21068–21073.*

Bishop, Y. M. M., Fienberg, S. E., and Holland, P. W. (1975) Discrete Multivariate Analysis: Theory and Practice. MIT Press. Reprinted by Springer (2007).

Cowell, R. G., Dawid, A. P., Lauritzen, S. L., and Spiegelhalter, D. J. (1999) Probabilistic Networks and Expert Systems. Springer.

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

Goldenberg, A., Zheng, A. X., Fienberg, S. E., and Airoldi E. M. (2010) A Survey of Statistical Network Models. Foundations and Trends in Machine Learning, **2** *(2), 129–233.*

Handcock, M. S. and Gile, K. J. (2010) Modeling Social Networks from Sampled Data. Annals of Applied Statistics, **4***, 5–25.*

References

Hanneke, S., Fu, W. and Xing, E. P. (2010) Discrete Temporal Models of Social Networks. *Electronic J. Statistics*, **4**, 585–605.

Lauritzen, S. (1996) *Graphical Models*. Oxford Univ. Press.

Kolaczyk, E. D. (2009) *Statistical Analysis of Network Data: Methods and Models.* Springer.

Petrović, S., Rinaldo, A., and Fienberg, S. E. (2010) Algebraic Statistics for a Directed Random Graph Model with Reciprocation. *Proceedings of the Conference on Algebraic Methods in Statistics and Probability,* Contemporary Mathematics Series, AMS.

Rinaldo, A., Fienberg, S. E. and Zhou, Y. (2009) On the Geometry of Discrete Exponential Families with Application to Exponential Random Graph Models. *Electronic J. Statistics*, **3**, 446–484.

Rinaldo, A., Petrović, S., and Fienberg, S. E. (2011) On the Existence of the MLE for a Directed Random Graph Network Model with Reciprocation. `arXiv:1010.0745v1`

Zachary, W. W. (1977) An Information Flow Model for Conflict and Fission in Small Groups. *J. Anthropological Research*, **33**, 452–473.