Hierarchical Models for Independence Structures of Networks

Kayvan Sadeghi
Department of Statistics
Carnegie Mellon University

May 2014

Joint work with Alessandro Rinaldo
Objective

- The independence structure of networks has not been well-studied.

- Graphical Markov models have been used to model conditional independence structures among random variables.

- The goal is to combine network and graphical models to deal with different possible independence structures.
Network models

• A network is a complete graph where
  – nodes correspond to labeled individuals;
  – edges correspond to binary random variables (also called dyads).

• In a realization of a network, \( i \sim j \) if the random variable corresponding to the edge \( ij \) takes 1, and \( i \not\sim j \) otherwise.
Graphical models

- In graphical models different types of dependence graphs have been used.

- A dependence graph is a graph where
  - nodes correspond to random variables \( X = (X_1, \ldots, X_{|N|}) \);
  - edges indicate some specific types of conditional dependencies among the random variables corresponding to the nodes.

- Here we focus on undirected graphs (conditional independencies), but a similar theory has been developed for bidirected graph (marginal independencies).
Probabilistic independence

- $X \perp Y \mid Z \iff f_{XYZ}(x, y, z) = \frac{f_{XZ}(x, z)f_{YZ}(y, z)}{f_Z(z)}$

- $(X_\alpha)_{\alpha \in V}$: a collection of random variables with a probability distribution $\mathbb{P}$

- Short notation $A \perp B \mid C$ for $X_A \perp X_B \mid X_C$ for $A, B, C$ subsets of $V$
The relationship between networks and dependence graphs

• We are interested in the network $G_n$ with $n = |V|$ nodes and $|E| = \binom{n}{2} := m$ dyads.

• Consider a dependence graph $D$ with $m$ nodes that are binary random variables.

• The basic assumption is that $G_n$ entails the same conditional independence structure among its edges as that of $D$ among its nodes.

• We use a corresponding dependence graph $D$ to model $H$. 
Modeling dependence graphs for modeling networks

- Is it enough to model a dependence graph with \( m \) binary random variables? No!
- Nodes of the network induce a specific labeling to the edges, which is lost by choosing an arbitrarily labelled dependence graph:
- Nodes of \( D \) should be labelled by pairs \( ij \) such that \( 1 \leq i < j \leq n \).
Structural and non-structural approaches to select dependence graph

- **Structural approach:** A given dependence graph is used to model the independence structure for the network.

- **Non-structural approach:** Model selection is performed to obtain a dependence graph.

- Here we deal with the structural approach.
Discrete graphical models

• Suppose that we have a dataset with $N$ observations of $d$ discrete random variables.

• We write the discrete variables as $X = (X_v)_{v \in V}$, and we call the possible values a discrete variable may take its levels.

• We can write a generic observation (or cell) as $i = (i_1, \ldots, i_d)$, and the set of possible cells as $I$.

• We are interested in modelling the probabilities $p(i) = Pr(X = i)$ for $i \in I$. 
Log-linear models

- For a three-dimensional table:
  We write variables as \((A, B, C)\) and a generic cell as \(i = (j, k, l)\);
  the saturated model as:
  \[
  \log p(i) = u + u_A^j + u_B^k + u_C^l + u_{AB}^{jk} + u_{AC}^{jl} + u_{BC}^{kl} + u_{ABC}^{jkl}.
  \]

- By letting \(\tilde{u} = \exp u\),
  \[
p(i) = \tilde{u} \cdot \tilde{u}_A^j \cdot \tilde{u}_B^k \cdot \tilde{u}_C^l \cdot \tilde{u}_{AB}^{jk} \cdot \tilde{u}_{AC}^{jl} \cdot \tilde{u}_{BC}^{kl} \cdot \tilde{u}_{ABC}^{jkl}.
  \]
Hierarchical log-linear models

- The term hierarchical means that if a term is set to zero, all its higher-order relatives are also set to zero.
- Hierarchical models are specified using the maximal interaction terms not set to zero: these are called the generators of the model.
- For example

\[ \log p(i) = u + u^a_j + u^b_k + u^c_l + u^{ab}_{jk} + u^{ac}_{jl} \]

has generators \( \{ab, ac\} \).
Hierarchical log-linear models for undirected graphs

- The generators of a hierarchical log-linear model correspond to the maximal cliques in the undirected graph.

- For example, for the generator set \( \{ab, ac\} \):

\[
p(i) = (\tilde{u}.\tilde{u}^a_j.\tilde{u}^b_k.\tilde{u}^{ab}_{jk})(\tilde{u}^c_l.\tilde{u}^{ac}_{jl}),
\]

which implies \( b \perp \perp c \mid a \), as needed.
Markov dependence property for undirected graphs

- We assume a dependence graph satisfies the **Markov dependence property** (Frank and Strauss 1986) if when $ij \not\sim kl$ in $G$ then $ij \not\sim kl$ in $D$.

- Notice that the statement holds only in one direction; therefore, such dependence graphs are subgraphs of the line graph of $G_n$.

- This implies that cliques in $D$ correspond to stars and triangles in $G$.

- In most practical cases, this is a plausible assumption.
Frank and Strauss: Markov Graphs

Figure 2. Dependence Graph $D$ of $G$ in Example 2.

Theorem 2.4. If $\{v, w\}$ is a consistent pair of nodes with $\operatorname{Pr}(G)$, then $\{v, w\}$ is a sufficient subgraph of $G$. This follows from (3.7).

In the general case of $n$ nodes, $G$ is a $k$-star if $n-k$ nodes have $(n-k-1)$ neighbors and $k$ nodes are incident with $n-k$ edges. This follows from (3.8). Graph $D$ (two illustrations of the correspondence between cliques of $D$ and sufficient subgraphs of $G$).
Network models

- For $x$, an observed network, exponential random graph models are of from $P(x) = \exp\{\sum_i s_i(x)\theta_i - \psi(\theta)\}$, where $s_i(x)$ is sufficient statistics, $\theta_i$ natural parameter, and $\psi(\theta)$ the normalizing constant.

- Models in the literature that assume dyadic independence (Erdős-Rényi, Beta, $P_1$, etc.) can be generalized to capture the independence structure of $D$.

- We show the method for Erdős-Rényi and Beta.

- For Erdős-Rényi the sufficient statistics is the number of edges and for beta is the degree sequence.
Hierarchical Erdős-Rényi models

- The baseline for empty independence graph:
  \[ p(x) = \tilde{u} \prod_{i<j} \exp\{x_{ij}q\}, \]  
  where \( q = \log\left(\frac{p}{1-p}\right) \).

- Hierarchical Erdős-Rényi models:
  \[ p_X(x) = \prod_{C \in \mathcal{C}} \tilde{u}_{i_C} = \tilde{u} \exp\{\sum_{C \in \mathcal{C}} \gamma_C \prod_{c \in C} x_c\}, \]
  where \( C \) is of from \( ij \), and

- \( \gamma_C = \begin{cases} 
  q^{(r)}, & \text{if } C = \{ii_1, ii_2, \ldots, ii_r\}; \\
  t, & \text{if } C = \{ij, ik, jk\}. 
\end{cases} \)
Example

\[ P(x) = \tilde{u} \exp\{q^{(1)} x_{12} + q^{(1)} x_{13} + q^{(1)} x_{23} + q^{(2)} x_{12}x_{13} + q^{(2)} x_{13}x_{23}\}. \]
Hierarchical Erdős-Rényi models

- Number of parameters is $r$: the size of the largest clique in $D$.
- Model hierarchical: $q^{(r)} = 0 \implies q^{(r+1)} = 0$.
- Model in exponential family form:

$$p_X(x) = \exp\left\{\sum_{r=1}^{n} q^{(r)} s^{(r)}_{C(r)}(x) + t.s^{(t)}_{C(t)}(x) - \psi(q, t)\right\},$$

$C^{(r)}$: set of all cliques with $r$ nodes in $D$,
$s^{(r)}_{C(r)}(x)$: the number of $r$-stars in $G$ with edges that are members of $C^{(r)}$;
- The saturated model (where $D$ is the line graph of a complete graph):
  The $k$-stars and triangle model.
- Normalizing constant in closed form, but depending on the cliques of the dependence graph.
Hierarchical Beta models

- The baseline for empty independence graph:
  \[ p_\beta(x) = \prod_{i<j} p_{ij}^{x_{ij}} (1 - p_{ij})^{1-x_{ij}} = \tilde{u} \prod_{i<j} e^{(\beta_i + \beta_j) x_{ij}}. \]

- Hierarchical Beta models:
  \[ p_X(x) = \prod_{C \in C} \tilde{u}_{i_C} = \tilde{u} \exp \left\{ \sum_{C \in C} \gamma_C \prod_{c \in C} x_c \right\}, \]

where \( C \) is of from \( ij \), and

- \( \gamma_C = \begin{cases} 
  \beta_i^{(r)}, & \text{if } C = \{ii_1, ii_2, \ldots, ii_r\}, \ r \geq 1 \\
  \tau_i + \tau_j + \tau_k, & \text{if } C = \{ij, ik, jk\}.
\end{cases} \)
Example

\[ P(x) = \tilde{u} \exp\{(\beta_1^{(1)} + \beta_2^{(1)})x_{12} + (\beta_1^{(1)} + \beta_3^{(1)})x_{13} + (\beta_2^{(1)} + \beta_3^{(1)})x_{23} + \beta_1^{(2)} x_{12} x_{13} + \beta_3^{(2)} x_{13} x_{23}\}. \]
Hierarchical Beta models

- Number of parameters is $nr$.

- Model hierarchical: $\beta_i^{(r)} = 0 \Rightarrow \beta_i^{(r+1)} = 0$.

- Model in exponential family form:

$$p_X(x) = \exp\{\sum_{i=1}^{n} \sum_{r=1}^{n} \beta_i^{(r)} d_{i,C_i^{(r)}}(x) + \beta_i^{(t)} d_{i,C_i^{(t)}}(x) - \psi(\beta)\},$$

$C_i^{(r)}$: the set of all cliques with $r$ nodes in $D$ such that all their nodes share $i$, $d_{i,C_i^{(r)}}(x)$: the number of $r$-stars in $G$ with hub $i$ and endpoints that are pairs of $i$ in members of $C_i^{(r)}$.

- The saturated model for line graph of a complete graph: sufficient statistics are the number of $r$-stars with $i$ as the hub; and the number of triangles that contain $i$. 
Parameter estimation

- Normalizing constant is in closed form $\Rightarrow$ there is an explicit formula for the likelihood function and its gradient; e.g. for HER:

$$\frac{\partial l(q)}{\partial q^{(i)}} = s^{(i)}(x) - \frac{\sum_{r=1}^{m} \sum_{H \subseteq D(r)} c^{(i)}(H) \exp\left\{\sum_{r'=1}^{\min(d,r)} c^{(r')}(H)q^{(r')}\right\}}{1 + \sum_{r=1}^{m} \sum_{H \subseteq D(r)} \exp\left\{\sum_{r'=1}^{\min(d,r)} c^{(r')}(H)q^{(r')}\right\}}.$$

- We apply the gradient decent method to obtain the MLE.

- The method is computationally demanding due to many terms in the sum and large values in $\exp$:

- For sparse dependence graphs, i.e., number of edges of order of a constant, the method works reasonably well.
Erdös-Rényi vs. hierarchical Erdös-Rényi: Simulation results

- We simulated binary data with conditional dependencies of a dependence graph using Gibbs sampling.

- We applied the likelihood ratio test to compare Erdös-Rényi vs. hierarchical Erdös-Rényi.

\[ D = 2l_{HER}(q) - 2l_{ER}(q_1, \ldots, q_r). \]

- Under the null hypothesis that these two models are not different
  \[ D \sim \chi_{r-1}. \]
Example for $n = 12$:

20 simulations form a star dependence graph

20 simulations form a dense dependence graph with max clique size 4
Related algebraic and geometric questions

- Find the model polytopes based on a given dependence graph in order to study the MLE existent.

- The identifiability problem for these models.

- **Markov moves**: How to move on the space of networks so that we keep sufficient statistics, which are based on the dependence graph, unchanged.

- **Model selection** for the non-structural approach: Model selection for binary variables with certain restrictions.
Summary

- Modeling the independence structure of networks is an essential but neglected task.

- In order to model independence structure for networks, we model binary graphical models with nodes labelled as $ij$ from.

- We generalized the models for network by using the hierarchical models in the graphical models literature to come up with models that capture certain independence structures.

- The selection of dependence graphs from data is an important task for further work.

- There is a lot of interesting geometry in these models to discover.

- A similar method has been developed for bidirected graphs, which are those that capture marginal independence.