Modeling Network Data

seminar

Jürgen Lerner

University of Konstanz

Coordination Meeting Konstanz, 16 April 2013
Topic of this seminar.

Statistical models for social network data.
Topic of this seminar.

Statistical models for social network data.

Social networks consist of **actors** and **relations** among them.

- **actors**: persons, organizations, companies, countries, ...  
- **relations**: friendship, asking for advice, communication, collaboration, trade, war, ...
Topic of this seminar.

Statistical models for social network data.

Statistics can formulate precise statements about uncertainty.

What would happen, if we measured the data again?

▶ at a different point in time,
▶ on a different set of actors,
▶ with different environmental factors, . . .

Estimate expected outcome ± variability

⇒ to explain and predict social relations and behavior.
Network models may serve several purposes.

**Explaining** social relations and/or behavior
- search for rules that govern the evolution of social networks.

**Predicting** social relations and/or behavior
- learn from given data and predict the data yet to come.

**Random generation of networks** that look like real data
- algorithm engineering; empirical estimation of average runtime or performance;
- simulation of network processes (e.g., information spreading, spread of disease).
A graph is a pair $G = (V, E)$, where $V$ is a finite set of vertices and $E$ the set of edges.

- **undirected graph**: $E \subseteq \binom{V}{2} = \{\{u, v\}; \ u, v \in V\}$
- **directed graph**: $E \subseteq V \times V = \{(u, v); \ u, v \in V\}$

**Interpretation:**
- vertices correspond to actors
- edges form the relation among them

Graphs can be *attributed* and/or *time-dependent*. 
Random graph models.

A *random graph model* is a probability space \((\mathcal{G}, P)\), where \(\mathcal{G}\) is a set of graphs and \(P: \mathcal{G} \to [0, 1]\) a probability function, satisfying

\[
\sum_{G \in \mathcal{G}} P(G) = 1.
\]
(I) Exponential random graph models (ERGM).

The *ERGM class* consists of random graph models \((\mathcal{G}, P_\theta)\) whose probability function \(P_\theta\) can be written as

\[
P_\theta(G) = \frac{1}{\kappa(\theta)} \exp \left( \sum_{i=1}^{k} \theta_i \cdot g_i(G) \right)
\]

with

- \(g_i: \mathcal{G} \rightarrow \mathbb{R}\) for \(i = 1, \ldots, k\) (statistics);
- \(\theta_i \in \mathbb{R}\) for \(i = 1, \ldots, k\) (parameters); \(\theta = (\theta_1, \ldots, \theta_k)\);
- *normalizing constant* \(\kappa\) defined by

\[
\kappa(\theta) = \sum_{G' \in \mathcal{G}} \exp \left( \sum_{i=1}^{k} \theta_i \cdot g_i(G') \right).
\]
(I) Commonly used statistics.

\[ P_\theta(G) = \frac{1}{\kappa(\theta)} \exp \left( \sum_{i=1}^{k} \theta_i \cdot g_i(G) \right) \]

<table>
<thead>
<tr>
<th>(g_i(G))</th>
<th>effect</th>
</tr>
</thead>
<tbody>
<tr>
<td>number of edges</td>
<td>density</td>
</tr>
<tr>
<td>edges connecting same attribute</td>
<td>homophily</td>
</tr>
<tr>
<td>number of triangles</td>
<td>transitivity</td>
</tr>
<tr>
<td>number of (\ell)-stars</td>
<td>pref. attachment</td>
</tr>
</tbody>
</table>
(II) Stochastic actor-oriented models.

Models for **longitudinal** network data: networks observed at $M \geq 2$ points in time $G_1, \ldots, G_M$.

Specify probabilities $P(G_{t_i} | G_{t_{i-1}})$ for the network at time $t_i$ given its state at the preceding observation time $t_{i-1}$.

Transition from $G_{t_{i-1}}$ to $G_{t_i}$ modeled by a stochastic process:

- at a given moment one probabilistically selected actor $u$ has the opportunity to change;
- actor $u$ might change one of his/her out-going ties $(u, v)$ to maximize a random utility function

$$f_u(\beta, G(u \rightarrow v)) = \sum_{k=1}^{K} \beta_k s_{uk}(G(u \rightarrow v)) + U_u$$
The topic of this seminar are models that extend or modify one of these frameworks making them applicable to more general network data.
Organizational points.
General information.

Project webpage:
http://www.inf.uni-konstanz.de/algo/lehre/ss13/seminar/

Participants independently get a published paper introducing a model extension or model alternative.

Participants give a presentation and write a term paper where they explain and summarize (potentially criticize) the main contribution of that paper.

**Target audience/readers:** your fellow students.
Requirements and timeline.

Credit requirements: term paper and presentation.

Approximate schedule:

- **(by 24 April)** topic selection;
- **(3 – 7 June)** individual meeting (discuss progress made so far);
- **(8 – 12 July)** individual meeting (outline of presentation and term paper);
- **(15 – 19 July)** depends on number of participants presentation in a plenary session (≈ 30 minutes plus 15 minutes discussion).
- **(by 30 September)** term paper
Topics / papers.

Extension of exponential random graph models to multivariate networks (more than one relation on the same set of actors).
(2) ERGM for valued edges I.


Extension of exponential random graph models to valued networks (edges have associated weights).

Compare to Krivitsky (2012).

Extension of exponential random graph models to valued networks (edges have associated weights).

Compare to Desmarais and Cranmer (2012).

Extension of exponential random graph models to time-dependent networks (a network observed at two or more points in time).

Compare to Krivitsky and Handcock (2012).
(5) Separable temporal ERGM.


Extension of exponential random graph models to time-dependent networks (a network observed at two or more points in time).

Compare to Hanneke, Fu, and Xing (2010).
(6) Adjusting ERGM for network size.


Extension of stochastic actor-oriented models to networks where the set of actors (nodes) changes over time.
(8) Multivariate SAOM.


Extension of stochastic actor-oriented models to multiple networks on the same set of actors.

Models networks given by dyadic, time-stamped interaction events, such as emails or phone calls.

Models event networks in a different way than Butts (2008).

Model for collections of several networks.
(12) Algebraic constraints.


Analyzes networks with multiple relations by searching for *algebraic constraints* among given or compound relations.