Relational event models for the analysis of social networks

an overview of network models

Alessandro Lomi   Jürgen Lerner

University of Exeter   University of Konstanz

University of Exeter, Business School   3–4 June 2019

On these slides we discuss

- what is special about **network** data;
- a range of statistical network models
  - for time-independent network data;
  - for longitudinal data given by snapshots;
  - for networks of relational events.

Outline.

Network data.

Models for time-independent networks and network snapshots.

Relational event models.

Outline.

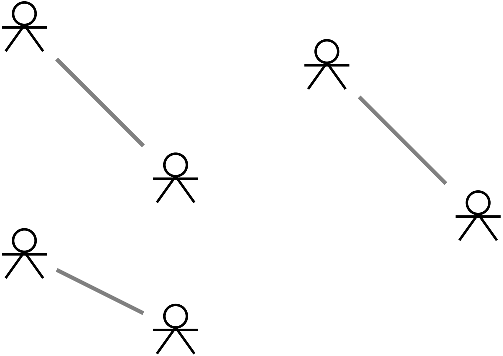
Network data.

Models for time-independent networks and network snapshots.

Relational event models.

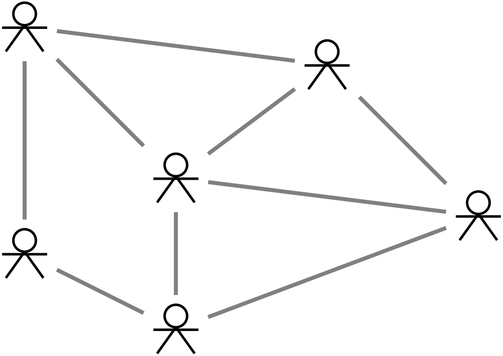
Network data.

Observations are associated with **dyads**



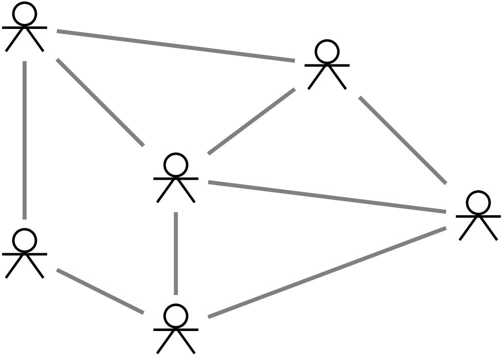
Network data.

Observations are associated with **overlapping dyads**



Network data.

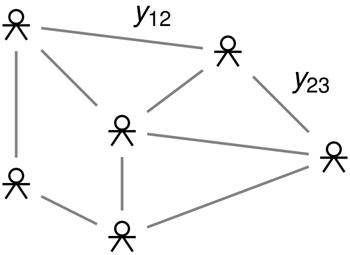
Observations are associated with **overlapping dyads**



$\Rightarrow$ ***independence of observations is nearly unthinkable.***

Networks as realizations of random variables.

Observations are associated with **overlapping dyads.**



$$\begin{bmatrix} \cdot & 1 & 0 & 1 & 1 & 0 \\ 1 & \cdot & 1 & 0 & 1 & 0 \\ 0 & 1 & \cdot & 0 & 1 & 1 \\ 1 & 0 & 0 & \cdot & 0 & 1 \\ 1 & 1 & 1 & 0 & \cdot & 1 \\ 0 & 0 & 1 & 1 & 1 & \cdot \end{bmatrix}$$

Observed network $y = (y_{ij})$ is realization of a matrix of random variables $Y = (Y_{ij})$.

Outline.

Network data.

Models for time-independent networks and network snapshots.

Relational event models.

Cross-sectional data: time-independent networks.

A simple statistical model for networks.

Observed network $y = (y_{ij})$ is realization of a
matrix of random variables $Y = (Y_{ij})$.

Mutually independent model

$$P(Y = y) \;=\; \prod_{ij} P(Y_{ij} = y_{ij}) \; .$$

Dyadic tie probabilities are independent of each other.

- This model is completely unrealistic.

Exponential random graph models (ERGM).

very general framework for network models

Probability of a network y has the functional form

$$P_\theta(Y = y) = \frac{1}{z} \exp\left(\sum_k \theta_k \cdot s_k(y) \right) \quad , \text{ where}$$

- the θ_k are **parameters**
- the $s_k(y)$ are **statistics** of the network y, such as
  - number of edges (**density**)
  - number of edges connecting nodes with similar characteristics (**homophily**)
  - number of triangles (**triadic closure**)
  - number of stars (**activity** or **popularity**)
  - ...
- z is a normalizing constant.

Exponential random graph models (ERGM).

very general framework for network models

Probability of a network y has the functional form

$$P_\theta(Y = y) \quad = \quad \frac{1}{z} \exp\left(\sum_k \theta_k \cdot s_k(y)\right) \quad .$$

Given an observed network, parameters θ are estimated s. t.

- the expected values of all statistics are equal to
- the statistics of the observed network.

Estimated parameters reveal network effects (homophily, ...).

Some points about ERGMs.

- Can deal with complex dependence among dyadic observations;
- might lead to degenerate models;
- parameter estimation is quite involved (but software exists)

$\Rightarrow$ e. g., R package `ergm` (`https://statnet.org/`).

Is a model for cross-sectional networks without time information (does not apply to our setting).

Longitudinal data given by network snapshots.

Longitudinal data given by network snapshots.

Networks, observed at given points in time $t_1, \ldots, t_h$

$$y = (y^{(1)}, \ldots, y^{(h)}) \ .$$

Each $y^{(\ell)}$ is an observed adjacency matrix.

Often results from repeated application of a questionnaire:
"*list all your friends within your school class*."

Stochastic actor-oriented models (SAOM).

Define probability distributions on

- **latent (unobserved)** sequences of micro-steps,
- where each micro-step changes the value of one dyad
  (deleting a tie or creating a tie),
- transforming a network $y^{(t)}$ into the next $y^{(t+1)}$.

Stochastic actor-oriented models (SAOM).

Define probability distributions on

- **latent (unobserved)** sequences of micro-steps,
- where each micro-step changes the value of one dyad (deleting a tie or creating a tie),
- transforming a network $y^{(t)}$ into the next $y^{(t+1)}$.

A micro-step is defined by two processes determining

1. the actor i who becomes active (gets the opportunity to change one out-going tie);
2. the target actor j (so that the tie y_{ij} will be flipped).

The active actor chooses micro-steps leading to desirable configurations (reciprocated ties, transitive closure, ...).

Distribution of micro-steps depends on current (latent) state.

Some points about SAOMs.

Define probability distributions on **latent (unobserved)** sequences of micro-steps,...

- Model network **dynamics** (rather than network state).
- Parameter estimation is quite involved (but software exists)

$\Rightarrow$ `RSiena` (`www.stats.ox.ac.uk/~snijders/siena`).

Is a model for network snapshots at given points in time (does not apply to our setting).

Outline.

Network data.

Models for time-independent networks and network snapshots.

Relational event models.

Networks of relational events.

Given by sequences of time-stamped dyadic events

$$E = (e_1, \ldots, e_N), \text{ where } e_i = (a_i, b_i, t_i, x_i)$$

- a_i **source** (sender) of the event;
- b_i **target** (receiver) of the event;
- t_i **time** of the event (potentially: ordinal);
- x_i **type** of the event (potentially: type and weight).

Who does when what to whom?

Relational event models (REM).

Given a sequence of time-stamped dyadic events

$$E = (e_1, \ldots, e_N) \ , \text{ where } e_i = (a_i, b_i, t_i, x_i) \ ,$$

define $G[E_{<t}] = G[\{e_i \in E\ ;\ t_i < t\}]$ **(network of past events).**

Relational event models (REM).

Given a sequence of time-stamped dyadic events

$$E = (e_1, \ldots, e_N) \ , \text{ where } e_i = (a_i, b_i, t_i, x_i) \ ,$$

define $G[E_{<t}] = G[\{e_i \in E \, ; \ t_i < t\}]$ **(network of past events).**

Relational event models specify probability distributions

$$P(E) = \prod_{i=1}^{N} P(e_i \,|\, G[E_{<t_i}]) \ .$$

- Events are assumed to be conditionally independent, given the network of past events.

$\Rightarrow$ With sufficient statistics of the network of past events, REMs are regression models(!).

Relational event models (REM).

Relational event models specify probability distributions

$$P(E) = \prod_{i=1}^{N} P(e_i \mid G[E_{<t_i}]) \ .$$

The conditional probabilities $P(e_i \mid G[E_{<t_i}])$ are specified by

$$y(e_i) \sim f(G[E_{<t_i}]; e_i)$$

- Response variables $y(e_i)$, encoding aspects of the event,
- are drawn from a probability distribution f,
- which is a function of the network of past events $G[E_{<t_i}]$.

Explanatory variables: statistics characterizing how the dyad (a_i, b_i) is embedded in the network of past events $G[E_{<t_i}]$.

$$y(e_i) \sim f(G[E_{<t_i}]; e_i)$$

Explanatory variables: event network statistics.

The distribution f in $y(e_i) \sim f(G[E_{<t_i}]; e_i)$ is typically a parametric function of dyad statistics $s = s_1, \ldots, s_k$

$$f(s(G[E_{<t_i}]; e_i); \theta), \text{ where } \theta = \theta_1, \ldots, \theta_k$$

Explanatory variables: event network statistics.

The distribution f in $y(e_i) \sim f(G[E_{<t_i}]; e_i)$ is typically a parametric function of dyad statistics $s = s_1, \ldots, s_k$

$$f(s(G[E_{<t_i}]; e_i); \theta), \text{ where } \theta = \theta_1, \ldots, \theta_k$$

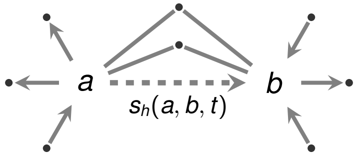
Statistics s_h characterize how dyads (A, B) are embedded into the network of past events.

- Distribution of events on the dyad (A, B) depend on past events
  - from A to B (**repetition**);
  - from B to A (**reciprocation**);
  - to/from A or B (**degree effects**);
  - to/from common third actors (**triadic effects**);
  - ...
- Parameters θ_h (to be estimated from empirical data) determine the effect of statistics on response variables.

Event network statistics $s_h(a, b, t)$.

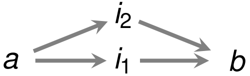
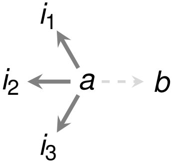
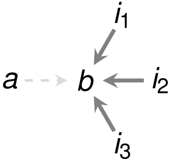
Statistics $s_h(a, b, t)$ assign time-varying values to dyads;

- are the variables which explain events on (a, b) at t;
- are functions of past events happening before t on the same or other dyads;
- introduce dependence among dyadic observations.



Examples: repetition, reciprocation, (in-/out-/mixed-)degrees, triadic effects, four-cycle effects, covariate effects, ...

Examples: common statistics $s_h(a, b, t)$.

statistic	$s_h(a, b, t) =$	$a \dashrightarrow b$ depends on
`repetition`	$att(a, b, t)$	$a \longrightarrow b$
`reciproc.`	$att(b, a, t)$	$a \longleftarrow b$
`transitivity`	$\sum_i att(a, i, t) \cdot att(i, b, t)$	
`outDegSource`	$\sum_i att(a, i, t)$	
`inDegTarget`	$\sum_i att(i, b, t)$	

Families of probability distributions for the next event

$$y(e_i) \sim f(G[E_{<t_i}]; e_i)$$

Here: describe four variants

1. modeling relative event rates;
2. modeling the time to events;
3. modeling conditional event types;
4. modeling conditional event weights.

Families of probability distributions for the next event

$$y(e_i) \sim f(G[E_{<t_i}]; e_i)$$

(I) Modeling relative event rates.

$\Rightarrow$ Cox proportional-hazard models.

Modeling relative event rates.

Cox proportional-hazard models

On which dyad does the next event happen?

Specify time-varying dyadic event rates $\lambda(A, B; t)$
(expected number of events per time unit on (A, B) at t).

Decompose into time-varying baseline rate and a parametric relative part

$$\begin{aligned}\lambda(A, B; t) &= \lambda_0(t) \cdot \lambda_1(A, B; t; \theta) \\ \lambda_1(A, B; t; \theta) &= \exp\left(\sum_h \theta_h \cdot s_h(A, B; G[E_{<t}])\right)\end{aligned}$$

Baseline rate λ_0 may be estimated by non-parametric methods.

Modeling relative event rates.

Cox proportional-hazard models

Parametric relative event rate

$$\lambda_1(A, B; t; \theta) \quad = \quad \exp\left(\sum_h \theta_h \cdot s_h(A, B; G[E_{<t}])\right)$$

Leads to a partial likelihood for the i'th event $e_i = (a_i, b_i, t_i)$

$$P(e_i \mid G[E_{<t_i}]; \theta) = \frac{\lambda_1(a_i, b_i; t_i; \theta)}{\sum_{ab \in R_{t_i}} \lambda_1(a, b; t_i; \theta)}$$

Risk set R_{t_i}: all dyads that could experience an event at t_i;
potentially: sample from R_{t_i}.

Parameters: maximize $P(E; \theta) = \prod_i P(e_i \mid G[E_{<t_i}]; \theta)$.
`coxph` in the R-package **survival**.

Families of probability distributions for the next event

$$y(e_i) \sim f(G[E_{<t_i}]; e_i)$$

(II) Modeling the time to events.

Modeling the time to events.

assuming piecewise-constant event rates, changing only at event times

On which dyad – and when – does the next event happen?

Specify time-varying dyadic event rates $\lambda(A, B; t)$
(expected number of events per time unit on (A, B) at t).

$$\lambda(A, B; t; \theta) \quad = \quad \exp\left(\sum_h \theta_h \cdot s_h(A, B; G[E_{<t}])\right)$$

Leads to likelihood for the i'th event $e_i = (a_i, b_i, t_i)$

$$P(e_i \mid G[E_{<t_i}]; \theta) = \frac{\lambda(a_i, b_i; t_i; \theta)}{\exp\left((t_i - t_{i-1}) \cdot \sum_{ab \in R_{t_i}} \lambda(a, b; t_i; \theta)\right)}$$

Parameter estimation: `survreg` in R-package **survival**.

Families of probability distributions for the next event

$$y(e_i) \sim f(G[E_{<t_i}]; e_i)$$

(III) Modeling conditional event types.

Modeling conditional event types.

How do actors interact, given that they do interact?

Assume that there are two types of events:

cooperative and conflictive.

Specify probability that type x_i of given event $e_i = (a_i, b_i, t_i, x_i)$ is cooperative (e. g., by a logit model)

$$P(x_i = \text{coop} \mid a_i, b_i; G[E_{<t_i}]; \theta) = \text{logit}^{-1}\left(\sum_h \theta_h \cdot s_h(a_i, b_i; G[E_{<t}])\right)$$

Generalizes to, e. g., (ordered) multi-nominal event types.

Families of probability distributions for the next event

$$y(e_i) \sim f(G[E_{<t_i}]; e_i)$$

(VI) Modeling conditional event weights.

Modeling conditional event types.

How do actors interact, given that they do interact?

Assume that events have numeric weights $x \in \mathbb{R}$
e. g., x measuring the performance of interaction.

Specify distribution of weight x_i of given event $e_i = (a_i, b_i, t_i, x_i)$
e. g., by the normal distribution $\Rightarrow$ linear regression.

$$\begin{aligned} x_i \mid a_i, b_i; G[E_{<t_i}]; \theta &\sim \mathcal{N}(\mu, \sigma^2) \\ \mu &= \sum_h \theta_h \cdot s_h(a_i, b_i; G[E_{<t}]) \end{aligned}$$

Generalizes to other distributions of event weights.