Stochastic Models for Networks

Mark S. Handcock

handcock@ucla.edu

Statistical Analysis of Networks

October 22, 2024

Statistical Models for Social Networks

A *social network* is defined as a set of n social "actors" and a social relationship between each pair of actors.

Statistical Models for Social Networks

A *social network* is defined as a set of n social “actors” and a social relationship between each pair of actors.

> To fix ideas and easy of presentation:
>
> > label the actors $1, \ldots, n$.
> >
> > focus on a single binary relationship

Statistical Models for Social Networks

A *social network* is defined as a set of n social “actors” and a social relationship between each pair of actors.

To fix ideas and easy of presentation:

label the actors $1, \ldots, n$.

focus on a single binary relationship

$\Rightarrow$ valued and metric relationships ok

the number of actors n is fixed and known

$\Rightarrow$ can be generalized

Statistical Models for Social Networks

A *social network* is defined as a set of n social "actors" and a social relationship between each pair of actors.

To fix ideas and easy of presentation:

label the actors $1, \ldots, n$.

focus on a single binary relationship

$\Rightarrow$ valued and metric relationships ok

the number of actors n is fixed and known

$\Rightarrow$ can be generalized

The presence or absence of a relationship is observed for each pair of actors

$\Rightarrow$ census: no sampling or missing data

Statistical Models for Social Networks

A *social network* is defined as a set of n social "actors" and a social relationship between each pair of actors.

To fix ideas and easy of presentation:

label the actors $1, \ldots, n$.

focus on a single binary relationship

$\Rightarrow$ valued and metric relationships ok

the number of actors n is fixed and known

$\Rightarrow$ can be generalized

The presence or absence of a relationship is observed for each pair of actors

$\Rightarrow$ census: no sampling or missing data

cross-sectional (time aggregate/static) viewpoint

$$Y_{ij} = \begin{cases} 1 & \text{relationship from actor } i \text{ to actor } j \\ 0 & \text{otherwise} \end{cases}$$

$$Y_{ij} = \begin{cases} 1 & \text{relationship from actor } i \text{ to actor } j \\ 0 & \text{otherwise} \end{cases}$$

call $Y \equiv [Y_{ij}]_{n \times n}$ a *sociomatrix*;
call the graphical representation of Y a *sociogram*

– a $N = n(n-1)$ array of binary random variables
– Y represents a random network with nodes the actors and edges the relationship

$$Y_{ij} = \begin{cases} 1 & \text{relationship from actor } i \text{ to actor } j \\ 0 & \text{otherwise} \end{cases}$$

call $Y \equiv [Y_{ij}]_{n\times n}$ a *sociomatrix*;
call the graphical representation of Y a *sociogram*

– a $N = n(n-1)$ array of binary random variables

– Y represents a random network with nodes the actors and edges the relationship

The basic problem of stochastic modeling is to specify a distribution for Y i.e., $P(Y = y)$

Simple model-classes for social networks

Model 1: Homogeneous Bernoulli graph *Rényi and Erdős model*

Y_{ij} are independent and equally likely

$$P(Y_{ij} = 1) = p \qquad \forall \;\; i, j = 1, \ldots, n$$

for some $0 \leq p \leq 1$

Simple model-classes for social networks

Model 1: Homogeneous Bernoulli graph *Rényi and Erdős model*

Y_{ij} are independent and equally likely

$$P(Y_{ij} = 1) = p \qquad \forall \;\; i, j = 1, \ldots, n$$

for some $0 \leq p \leq 1$

Equivalently

$$\log \text{odds}(Y_{ij} = 1) = \eta \qquad \forall \;\; i, j = 1, \ldots, n$$

where

$$\eta = \text{logit}[P(Y_{ij} = 1)] = \log\left(\frac{p}{1-p}\right)$$

Simple model-classes for social networks

Model 1: Homogeneous Bernoulli graph *Rényi and Erdős model*

Y_{ij} are independent and equally likely

$$P(Y_{ij} = 1) = p \qquad \forall \;\; i, j = 1, \ldots, n$$

for some $0 \leq p \leq 1$
Equivalently

$$\log \text{odds}(Y_{ij} = 1) = \eta \qquad \forall \;\; i, j = 1, \ldots, n$$

where

$$\eta = \text{logit}[P(Y_{ij} = 1)] = \log(\frac{p}{1-p})$$

Independent means, e.g.,

$$P(Y_{12} = 1, Y_{23} = 1) = p \times p$$

Simple model-classes for social networks

Model 1: Homogeneous Bernoulli graph *Rényi and Erdős model*

Y_{ij} are independent and equally likely

$$P(Y_{ij} = 1) = p \qquad \forall \;\; i, j = 1, \ldots, n$$

for some $0 \leq p \leq 1$

Equivalently

$$\log \text{odds}(Y_{ij} = 1) = \eta \qquad \forall \;\; i, j = 1, \ldots, n$$

where

$$\eta = \text{logit}[P(Y_{ij} = 1)] = \log(\frac{p}{1-p})$$

Independent means, e.g.,

$$P(Y_{12} = 1, Y_{23} = 1) = p \times p = \left(\frac{e^\eta}{1+e^\eta}\right)^2$$

Simple model-classes for social networks

Model 1: Homogeneous Bernoulli graph *Rényi and Erdős model*

Y_{ij} are independent and equally likely

$$P(Y_{ij} = 1) = p \qquad \forall \; i, j = 1, \ldots, n$$

for some $0 \leq p \leq 1$

Equivalently

$$\log \text{odds}(Y_{ij} = 1) = \eta \qquad \forall \; i, j = 1, \ldots, n$$

where

$$\eta = \text{logit}[P(Y_{ij} = 1)] = \log\left(\frac{p}{1-p}\right)$$

Independent means, e.g.,

$$P(Y_{12} = 1, Y_{23} = 1) = p \times p = \left(\frac{e^{\eta}}{1 + e^{\eta}}\right)^2$$

Multivariate distribution

More abstractly:

$$P(Y_{12} = y_{12}, Y_{21} = y_{21}) = p^{y_{12}}(1-p)^{1-y_{12}} \times p^{y_{21}}(1-p)^{1-y_{21}}$$

Multivariate distribution

More abstractly:

$$P(Y_{12} = y_{12}, Y_{21} = y_{21}) = p^{y_{12}}(1-p)^{1-y_{12}} \times p^{y_{21}}(1-p)^{1-y_{21}}$$

$$= \frac{e^{\eta(y_{12}+y_{21})}}{(1+e^{\eta})^2}$$

Multivariate distribution

More abstractly:

$$P(Y_{12} = y_{12}, Y_{21} = y_{21}) = p^{y_{12}}(1-p)^{1-y_{12}} \times p^{y_{21}}(1-p)^{1-y_{21}}$$

$$= \frac{e^{\eta(y_{12}+y_{21})}}{(1+e^{\eta})^2}$$

In general:

$$P(Y = y) = \frac{e^{\eta \sum_{i,j} y_{ij}}}{c(\eta)} \qquad y \in \mathcal{Y}$$

where $c(\eta) = [1 + \exp(\eta)]^N$

Properties of the model

What is the network density?

Properties of the model

What is the network density?

What is the probability that a randomly chosen possible tie exists?

Properties of the model

What is the network density?

What is the probability that a randomly chosen possible tie exists?

What is the probability that a dyad is mutual?

Properties of the model

What is the network density?

What is the probability that a randomly chosen possible tie exists?

What is the probability that a dyad is mutual?

What is the probability that a dyad is asymmetric?

Properties of the model

What is the network density?

What is the probability that a randomly chosen possible tie exists?

What is the probability that a dyad is mutual?

What is the probability that a dyad is asymmetric?

Expression as Logistic regression

As the Y_{ij} are independent we recognize this as a logistic regression with just a constant term:

$$\log \text{odds}(Y_{ij} = 1) = \eta \qquad\qquad \forall \;\; i, j = 1, \ldots, n$$

```
ergm(net1 ~ edges)
```

Model 2: Two type of actors

Suppose $1, \ldots, n_1$ are "red" and $n_1 + 1, \ldots, n$ are "blue" and the relationship is directed

Y_{ij} are independent but depend on the color of the actors

Suppose actors of each color differ in their propensity to send ties (regardless of the color of the alters).

"activity"

$$x_{ij} = \begin{cases} 1 & \text{the color of } i \text{ is "red"} \\ 0 & \text{the color of } i \text{ is "blue"} \end{cases}$$

Model 2: Two type of actors

Suppose $1, \ldots, n_1$ are "red" and $n_1 + 1, \ldots, n$ are "blue" and the relationship is directed

Y_{ij} are independent but depend on the color of the actors

Suppose actors of each color differ in their propensity to send ties (regardless of the color of the alters).

"activity"

$$x_{ij} = \begin{cases} 1 & \text{the color of } i \text{ is "red"} \\ 0 & \text{the color of } i \text{ is "blue"} \end{cases}$$

$$\log \text{odds}(Y_{ij} = 1 | x_{ij}) = \eta_0 + \eta_1 x_{ij} \qquad \forall \; i, j = 1, \ldots, n$$

```
ergm(net1 ~ edges + nodeofactor("color"))
```

Two type of actors and popularity

Suppose $1, \ldots, n_1$ are "red" and $n_1 + 1, \ldots, n$ are "blue" and the relationship is directed

Y_{ij} are independent but depend on the color of the actors

Suppose actors of each color differ in their propensity to receive ties (regardless of the color of the alters).

"popularity"

$$x_{ij} = \begin{cases} 1 & \text{the color of } j \text{ is "red"} \\ 0 & \text{the color of } j \text{ is "blue"} \end{cases}$$

Two type of actors and popularity

Suppose $1, \ldots, n_1$ are "red" and $n_1 + 1, \ldots, n$ are "blue" and the relationship is directed

Y_{ij} are independent but depend on the color of the actors

Suppose actors of each color differ in their propensity to receive ties (regardless of the color of the alters).

"popularity"

$$x_{ij} = \begin{cases} 1 & \text{the color of } j \text{ is "red"} \\ 0 & \text{the color of } j \text{ is "blue"} \end{cases}$$

$$\log \text{odds}(Y_{ij} = 1 | x_{ij}) = \eta_0 + \eta_1 x_{ij} \qquad \forall \; i, j = 1, \ldots, n$$


```
ergm(net1 ~ edges + nodeifactor("color"))
```

Two type of actors and homophily

Suppose $1, \ldots, n_1$ are "red" and $n_1 + 1, \ldots, n$ are "blue" and the relationship is directed

Y_{ij} are independent but depend on the color of the actors

Suppose actors of the same color prefer each other "homophily"

$$x_{ij} = \begin{cases} 1 & \text{the color of } i \text{ is the same as the color of } j \\ 0 & \text{the color of } i \text{ is the not the same as the color of} \end{cases}$$

Two type of actors and homophily

Suppose $1, \ldots, n_1$ are "red" and $n_1 + 1, \ldots, n$ are "blue" and the relationship is directed

Y_{ij} are independent but depend on the color of the actors

Suppose actors of the same color prefer each other "homophily"

$$x_{ij} = \begin{cases} 1 & \text{the color of } i \text{ is the same as the color of } j \\ 0 & \text{the color of } i \text{ is the not the same as the color of} \end{cases}$$

$$\log \text{odds}(Y_{ij} = 1 | x_{ij}) = \eta_0 + \eta_1 x_{ij} \qquad \forall\ i, j = 1, \ldots, n$$

```
ergm(net1 ~ edges + nodematch("color"))
```

Modeling Centrality of the actors

In lecture 7 we considered the concept of centrality of an actor and measures of it for a given undirected network.

Can we represent it in a statistical model directly?

$$\text{logit}[P(Y_{ij} = 1)] = \beta_i + \beta_j$$

So β_i is a measure of the centrality of actor i.
The difference $\beta_i - \beta_j$ indicates the log-odds ratio of a tie from i compared to j.
Note the lack of an edge parameter.

Modeling Centrality of the actors

In `ergm`:

```
ergm(flomarriage ~ sociality(base=0))

             Estimate Std. Error  value Pr(>|z|)
sociality1   -1.9871     1.0669 -1.862   0.0625 .
sociality2   -0.5708     0.6991 -0.817   0.4142
sociality3   -1.1455     0.8038 -1.425   0.1541
sociality4   -0.5708     0.6991 -0.817   0.4142
...
sociality12     -Inf     0.0000   -Inf   <1e-04 ***
...
sociality16  -0.5708     0.6991 -0.817   0.4142
```

Determining Individual Centrality

The centrality model provides the a relative measure of centrality, but it needs to be standardized to allow interpretation. Using the model we can predict the probability of a tie between two actors. From that we can calculate the model-based "eigenvalue" centrality.

$$c_i = \sum_{j=1}^{g} P(Y_{ij} = 1) c_j$$

$$\sum_{j=1}^{g} c_j = 0$$

Modeling the popularity of the actors

In lecture 4 we consider the concept of popularity of an actor and measures of it for a given directed graph.

Can we represent it in a statistical model directly?

$$\text{logit}[P(Y_{ij} = 1)] = \beta_i + \gamma_j$$

So β_i is a measure of the centrality of actor i. and γ_j is a measure of the prestige of actor j.

Modeling the Prestige of the actors

The difference $\beta_i - \beta_j$ indicates the log-odds ratio of an outtie from *i* compared to *j*.
The difference $\gamma_i - \gamma_j$ indicates the log-odds ratio of an intie from *i* compared to *j*.
Note the lack of an edge parameter.

Modeling the Prestige of the actors

In `ergm`:

```
ergm(samplike ~ receiver(base=0) + sender)
```

Examples: Friends and Acquaintances

	Estimate	Std. Error	z-score	Pr(>|z|)
receiver1	1.11537	0.75579	1.476	0.1400
receiver2	0.84327	0.72575	1.162	0.2453
receiver3	-1.10099	0.82058	-1.342	0.1797
...				
receiver5	0.35509	0.71816	0.494	0.6210
...				
receiver18	-1.08076	0.82026	-1.318	0.1876
sender2	-0.31750	0.77797	-0.408	0.6832
sender3	-0.45453	0.78233	-0.581	0.5612
...				
sender18	-0.14531	0.76409	-0.190	0.8492

The General Bernoulli graph model

Y_{ij} are independent but have arbitrary distributions

$$\text{logit}[P(Y_{ij} = 1)] \; = \; \eta_{ij} \qquad\qquad i, j = 1, \ldots, n$$

Based on independence:

$$P(Y = y) = \frac{\exp\left\{\sum_{i,j} \eta_{ij} y_{ij}\right\}}{c(\eta)} \qquad y \in \mathcal{Y}$$

$$c(\eta) = \prod_{i,j} [1 + \exp(\eta_{ij})]$$

one term per ij pair

Dyad-independence models with attributes

Y_{ij} are independent but depend on dyadic covariates $x_{k,ij}, k = 1, \ldots, K$.

$x_{k,ij}$ is the kth covariate for the ijth pair

$$\log\left[P(Y = y)\right] = \sum_{i,j} \eta_{ij} y_{ij} - c(\eta) \qquad y \in \mathcal{Y}$$

where Y_{ij} can depend on dyadic covariates $x_{k,ij}$

$$\eta_{ij} = \eta_1 x_{1,ij} + \eta_2 x_{2,ij} \ldots + \eta_K x_{K,ij}$$

and $c(\eta) = \log[c(\eta)]$.

$$\text{logit}[P(Y_{ij} = 1)] = \eta_1 x_{1,ij} + \eta_2 x_{2,ij} \ldots + \eta_K x_{K,ij}$$

By independence, the joint distribution can be expressed as:

$$P(Y = y) = \frac{\exp\{\eta_1 g_1(y) + \eta_2 g_2(y) \ldots + \eta_K g_K(y)\}}{c(\eta)} \qquad y \in \mathcal{Y}$$

$$g_k(y) = \sum_{i,j} x_{k,ij} y_{ij}, \quad k = 1, \ldots, K \qquad q = K$$

$$c(\eta) = \prod_{i,j} [1 + \exp(\sum_k \eta_k x_{k,ij})]$$

Interpreting the parameters in the model

$$P(Y = y) = \frac{\exp\left\{\sum_{k=1}^{K} \eta_k g_k(y)\right\}}{c(\eta)}$$

where $\eta_{1,2 \ldots k}$ are parameters $g_{1,2 \ldots k}(y)$ are statistics, and $c(\eta)$ is a normalizing constant:

$$c(\eta) = \sum_{y \in \mathcal{Y}} \exp\left\{\sum_{k=1}^{K} \eta_k g_k(y)\right\}$$

In other words,

$$P(Y = y) \propto \eta_1 g_1(y) + \eta_2 g_2(y) + \eta_3 g_3(y) + \ldots + \eta_k g_k(y)$$

Intuition: the ERGM places more/less weight on graphs with certain features, as determined by η, g

Interpreting the parameters in the model

We can also re-express it in terms of the odds of tie y_{ij}

$$\frac{Pr(Y_{ij} = 1)}{Pr(Y_{ij} = 0)} = \exp\left\{\sum_{k=1}^{K} \eta_k(g_k(y_{ij}^{+}) - g_k(y_{ij}^{-}))\right\}$$

where y_{ij}^{+} is the graph with $Y_{ij} = 1$,
y_{ij}^{-} is the graph with $Y_{ij} = 0$

Implications:

Log-odds only depend on the “change-score”,
$\Delta_{ij} = g_k(y_{ij}^{+}) - g_k(y_{ij}^{-})$

Interpreting the parameters in the model

We can also re-express it in terms of the odds of tie y_{ij}

$$\frac{Pr(Y_{ij} = 1)}{Pr(Y_{ij} = 0)} = \exp\left\{\sum_{k=1}^{K} \eta_k (g_k(y_{ij}^{+}) - g_k(y_{ij}^{-}))\right\}$$

where y_{ij}^{+} is the graph with $Y_{ij} = 1$,
y_{ij}^{-} is the graph with $Y_{ij} = 0$

Implications:

- Each unit change in g_k for (i, j) tie present (versus absent) increases the log-odds of (i,j) by η_k
- η is the impact of the covariate on the log-odds of a tie

Modeling Mutuality within the graph

In lecture 2 we considered the concept of mutuality of ties between actors and measures of it for a given graph.

Can we represent it in a statistical model directly?

$$\text{logit}[P(Y_{ij} = y_{ij}, Y_{ji} = y_{ji})] = \theta(y_{ij} + y_{ji}) + \rho y_{ij} y_{ji}$$

So ρ is the propensity for ties to match.

As the *dyad pairs* (Y_{ij}, Y_{ji}) are independent of (Y_{kl}, Y_{lk}), $i, j \neq k, l$

The full model is

$$P(Y = y) = \prod_{i<j} P(Y_{ij} = y_{ij}, Y_{ji} = y_{ji})$$

$$= \frac{\exp\{\rho \sum_{i<j} y_{ij} y_{ji} + \theta \sum_{i,j} y_{ij}\}}{c(\rho, \theta)}$$

where

θ controls the expected number of edges

ρ represent the expected tendency toward *reciprocation*

In `ergm`:

```
ergm(samplike ~ edges + mutual)
```

Including covariates in the mutual model

Recall the dyadic covariates $x_{k,ij}, k = 1, \ldots, K$:

$$x_{k,ij} \text{ is the } k\text{th covariate for the } ij\text{th pair}$$

We can include them in the model as:

$$\text{logit}[P(Y_{ij} = y_{ij}, Y_{ji} = y_{ji} | x_{1,ij}, \ldots, x_{K,ij})] = \sum_{k=1}^{K} \beta_k x_{k,ij} y_{ij} + \rho y_{ij} y_{ji}$$

$$P(Y = y) = \frac{\exp\{\sum_{k=1}^{K} \beta_k \sum_{i,j} x_{k,ij} y_{ij} + \rho \sum_{i<j} y_{ij} y_{ji} + \theta \sum_{i,j} y_{ij}\}}{c(\rho, \theta)}$$

where

β are the regression parameters
θ controls the expected number of edges
ρ represent the expected tendency toward

Some history of models for social networks

Holland and Leinhardt (1981) proposed a general dyad independence model

Also an homogeneous version they refer to as the "p^1" model

$$P(Y=y) = \frac{\exp\{\rho \sum_{i<j} y_{ij} y_{ji} + \theta \sum_{i,j} y_{ij} + \sum_i \alpha_i \sum_j y_{ij} + \sum_j \beta_j \sum_i y_{ij}\}}{c(\rho, \alpha, \beta, \theta)}$$

where

θ controls the expected number of edges

ρ represent the expected tendency toward *reciprocation*

α_i *productivity* of node i; β_j *attractiveness* of node j

Much related work and generalizations