Modeling ordinal, ranked and censored relational data

Peter Hoff

Statistics and Biostatistics, University of Washington
Joint work with Bailey Fosdick, Alex Volfovsky and Kate Stovel
Supported by NICHD grant R01 HD-67509.
Inference from relational data

Relational data:

Nodeset: a collection of objects;
Relation: a quantitative relation between pairs of objects;
Sociomatrix: a square matrix of relations.

\[
Y = \begin{pmatrix}
    \text{NA} & y_{1,2} & \cdots & y_{1,n} \\
    y_{2,1} & \text{NA} & \cdots & y_{2,n} \\
    \vdots & \vdots & \ddots & \vdots \\
    y_{n,1} & \cdots & y_{n,n-1} & \text{NA}
\end{pmatrix}
\]

Statistical inference:

\[
P = \{ P_\theta : \theta \in \Theta \}
\]

\(P_\theta\) is a distribution over \(n \times n\) sociomatrices for each \(\theta \in \Theta\).

- assume \(Y \sim P_\theta\) for some \(\theta \in \Theta\);
- infer \(\theta\) from \(Y\).
Inference from relational data

Relational data:

Nodeset: a collection of objects;
Relation: a quantitative relation between pairs of objects;
Sociomatrix: a square matrix of relations.

\[
Y = \begin{pmatrix}
    \text{NA} & y_{1,2} & \cdots & y_{1,n} \\
y_{2,1} & \text{NA} & \cdots & y_{2,n} \\
    \vdots & \vdots & \ddots & \vdots \\
y_{n,1} & \cdots & y_{n,n-1} & \text{NA}
\end{pmatrix}
\]

Statistical inference:

\[
P = \{ P_\theta : \theta \in \Theta \}
\]

\(P_\theta\) is a distribution over \(n \times n\) sociomatrices for each \(\theta \in \Theta\).

- assume \(Y \sim P_\theta\) for some \(\theta \in \Theta\);
- infer \(\theta\) from \(Y\).
Inference from relational data

Relational data:

Nodeset: a collection of objects;
Relation: a quantitative relation between pairs of objects;
Sociomatrix: a square matrix of relations.

\[ Y = \begin{pmatrix}
  \text{NA} & y_{1,2} & \cdots & y_{1,n} \\
y_{2,1} & \text{NA} & \cdots & y_{2,n} \\
  \vdots & \vdots & \ddots & \vdots \\
y_{n,1} & \cdots & y_{n,n-1} & \text{NA}
\end{pmatrix} \]

Statistical inference:

\[ \mathcal{P} = \{ P_\theta : \theta \in \Theta \} \]

\( P_\theta \) is a distribution over \( n \times n \) sociomatrices for each \( \theta \in \Theta \).

- assume \( Y \sim P_\theta \) for some \( \theta \in \Theta \);
- infer \( \theta \) from \( Y \).
Inference from relational data

Relational data:

Nodeset: a collection of objects;
Relation: a quantitative relation between pairs of objects;
Sociomatrix: a square matrix of relations.

\[
Y = \begin{pmatrix}
  \text{NA} & y_{1,2} & \cdots & y_{1,n} \\
y_{2,1} & \text{NA} & \cdots & y_{2,n} \\
  \vdots & \vdots & \ddots & \vdots \\
y_{n,1} & \cdots & y_{n,n-1} & \text{NA}
\end{pmatrix}
\]

Statistical inference:

\[ P = \{ P_\theta : \theta \in \Theta \} \]

\( P_\theta \) is a distribution over \( n \times n \) sociomatrices for each \( \theta \in \Theta \).

- assume \( Y \sim P_\theta \) for some \( \theta \in \Theta \);
- infer \( \theta \) from \( Y \).
Inference from relational data

Relational data:

Nodeset: a collection of objects;
Relation: a quantitative relation between pairs of objects;
Sociomatrix: a square matrix of relations.

\[
\mathbf{Y} = \begin{pmatrix}
\text{NA} & y_{1,2} & \cdots & y_{1,n} \\
y_{2,1} & \text{NA} & \cdots & y_{2,n} \\
\vdots & \vdots & \ddots & \vdots \\
y_{n,1} & \cdots & y_{n,n-1} & \text{NA}
\end{pmatrix}
\]

Statistical inference:

\[ \mathcal{P} = \{ P_\theta : \theta \in \Theta \} \]

\( P_\theta \) is a distribution over \( n \times n \) sociomatrices for each \( \theta \in \Theta \).

- assume \( \mathbf{Y} \sim P_\theta \) for some \( \theta \in \Theta \);
- infer \( \theta \) from \( \mathbf{Y} \).
Inference from relational data

**Relational data:**

Nodeset: a collection of objects;
Relation: a quantitative relation between pairs of objects;
Sociomatrix: a square matrix of relations.

\[
Y = \begin{pmatrix}
  \text{NA} & y_{1,2} & \cdots & y_{1,n} \\
y_{2,1} & \text{NA} & \cdots & y_{2,n} \\
  \vdots & \vdots & & \vdots \\
y_{n,1} & \cdots & y_{n,n-1} & \text{NA}
\end{pmatrix}
\]

**Statistical inference:**

\[ P = \{ P_\theta : \theta \in \Theta \} \]

\( P_\theta \) is a distribution over \( n \times n \) sociomatrices for each \( \theta \in \Theta \).

- assume \( Y \sim P_\theta \) for some \( \theta \in \Theta \);
- infer \( \theta \) from \( Y \).
Inference from relational data

Relational data:

Nodeset: a collection of objects;
Relation: a quantitative relation between pairs of objects;
Sociomatrix: a square matrix of relations.

\[ Y = \begin{pmatrix} NA & y_{1,2} & \cdots & y_{1,n} \\ y_{2,1} & NA & \cdots & y_{2,n} \\ \vdots & \vdots & \ddots & \vdots \\ y_{n,1} & \cdots & y_{n,n-1} & NA \end{pmatrix} \]

Statistical inference:

\[ \mathcal{P} = \{ P_\theta : \theta \in \Theta \} \]

\( P_\theta \) is a distribution over \( n \times n \) sociomatrices for each \( \theta \in \Theta \).

- assume \( Y \sim P_\theta \) for some \( \theta \in \Theta \);
- infer \( \theta \) from \( Y \).
Sample spaces

\[ \mathcal{Y} = \text{all possible values of } \mathbf{Y} \]
\[ = \text{the “sample space”} \]

What is \( \mathcal{Y} \) for a given relational dataset?
Sample spaces

\[ \mathcal{Y} = \text{all possible values of } Y \]
\[ = \text{the “sample space”} \]

What is \( \mathcal{Y} \) for a given relational dataset?
Networks, links and graphs
The vast majority of statistical models assume

- \( y_{i,j} \in \{0, 1\} \)
- \( \mathcal{Y} = \{0, 1\}^{n(n-1)} \), i.e. the set of (directed) graphs with no loops.

**Example** (latent variable models):

\[
Pr(\mathbf{Y} = \mathbf{y} | \theta) = \prod_{i \neq j} f(\{y_{i,j}, y_{j,i}\} | a_i, b_j, \mu)
\]


**Example** (ERGMs):

\[
Pr(\mathbf{Y} = \mathbf{y} | \theta) = c(\theta) \exp (t(\mathbf{y}) \cdot \theta)
\]

The vast majority of statistical models assume

- $y_{i,j} \in \{0, 1\}$
- $\mathcal{Y} = \{0, 1\}^{n(n-1)}$, i.e. the set of (directed) graphs with no loops.

**Example** (latent variable models):

$$\Pr(Y = y | \theta) = \prod_{i \neq j} f(y_{i,j}, y_{j,i} | a_i, b_j, \mu)$$


**Example** (ERGMs):

$$\Pr(Y = y | \theta) = c(\theta) \exp (t(y) \cdot \theta)$$

Standard statistical models

The vast majority of statistical models assume

- \( y_{i,j} \in \{0, 1\} \)
- \( \mathcal{Y} = \{0, 1\}^{n(n-1)} \), i.e. the set of (directed) graphs with no loops.

**Example** (latent variable models):

\[
Pr(\mathbf{Y} = \mathbf{y}|\theta) = \prod_{i \neq j} f(\{y_{i,j}, y_{j,i}\}|a_i, b_j, \mu)
\]


**Example** (ERGMs):

\[
Pr(\mathbf{Y} = \mathbf{y}|\theta) = c(\theta) \exp(t(\mathbf{y}) \cdot \theta)
\]

Standard statistical models

The vast majority of statistical models assume

- $y_{i,j} \in \{0, 1\}$
- $\mathcal{Y} = \{0, 1\}^{n(n-1)}$, i.e. the set of (directed) graphs with no loops.

**Example** (latent variable models):

$$\Pr(Y = y | \theta) = \prod_{i \neq j} f(y_{i,j}, y_{j,i} | a_i, b_j, \mu)$$


**Example** (ERGMs):

$$\Pr(Y = y | \theta) = c(\theta) \exp(t(y) \cdot \theta)$$

Standard statistical models

The vast majority of statistical models assume

- \( y_{i,j} \in \{0, 1\} \)
- \( \mathcal{Y} = \{0, 1\}^{n(n-1)} \), i.e. the set of (directed) graphs with no loops.

**Example** (latent variable models):

\[
\Pr(\mathbf{Y} = \mathbf{y} | \theta) = \prod_{i \neq j} f(\{y_{i,j}, y_{j,i}\} | a_i, b_j, \mu)
\]


**Example** (ERGMs):

\[
\Pr(\mathbf{Y} = \mathbf{y} | \theta) = c(\theta) \exp( t(\mathbf{y}) \cdot \theta)
\]

The vast majority of statistical models assume

- \( y_{i,j} \in \{0, 1\} \)
- \( \mathcal{Y} = \{0, 1\}^{n(n-1)} \), i.e. the set of (directed) graphs with no loops.

**Example** (latent variable models):

\[
\Pr(\mathbf{Y} = \mathbf{y}|\theta) = \prod_{i \neq j} f(\{y_{i,j}, y_{j,i}\}|a_i, b_j, \mu)
\]


**Example** (ERGMs):

\[
\Pr(\mathbf{Y} = \mathbf{y}|\theta) = c(\theta) \exp(t(\mathbf{y}) \cdot \theta)
\]

Binary versus valued network relations

Literature suggests that most data consist of unconstrained binary relations:

- \( \mathcal{Y} = \{0, 1\}^{n(n-1)} \);
- \( P(\mathcal{Y}|\theta) = 1 \) for all \( \theta \in \Theta \);
- \( P(A|\theta) < 1 \) for any proper subset \( A \) of \( \mathcal{Y} \).

Raw relational data sources suggest otherwise:

http://moreno.ss.uci.edu/data.html

A substantial fraction, if not majority, of relational data are “valued.”
Binary versus valued network relations

Literature suggests that most data consist of unconstrained binary relations:

- \( \mathcal{Y} = \{0, 1\}^{n(n-1)} \);
- \( P(\mathcal{Y}|\theta) = 1 \) for all \( \theta \in \Theta \);
- \( P(A|\theta) < 1 \) for any proper subset \( A \) of \( \mathcal{Y} \).

Raw relational data sources suggest otherwise:

http://moreno.ss.uci.edu/data.html

A substantial fraction, if not majority, of relational data are “valued.”
Literature suggests that most data consist of unconstrained binary relations:

- $\mathcal{Y} = \{0, 1\}^{n(n-1)}$;
- $P(\mathcal{Y}|\theta) = 1$ for all $\theta \in \Theta$;
- $P(A|\theta) < 1$ for any proper subset $A$ of $\mathcal{Y}$.

Raw relational data sources suggest otherwise:
http://moreno.ss.uci.edu/data.html

A substantial fraction, if not majority, of relational data are “valued.”
Binary versus valued network relations

Literature suggests that most data consist of unconstrained binary relations:

- $\mathcal{Y} = \{0, 1\}^{n(n-1)}$;
- $P(\mathcal{Y}|\theta) = 1$ for all $\theta \in \Theta$;
- $P(A|\theta) < 1$ for any proper subset $A$ of $\mathcal{Y}$.

Raw relational data sources suggest otherwise:
http://moreno.ss.uci.edu/data.html

A substantial fraction, if not majority, of relational data are “valued.”
Binary versus valued network relations

Literature suggests that most data consist of unconstrained binary relations:

- $\mathcal{Y} = \{0, 1\}^{n(n-1)}$;
- $P(\mathcal{Y} | \theta) = 1$ for all $\theta \in \Theta$;
- $P(A | \theta) < 1$ for any proper subset $A$ of $\mathcal{Y}$.

Raw relational data sources suggest otherwise:
http://moreno.ss.uci.edu/data.html

A substantial fraction, if not majority, of relational data are “valued.”
Literature suggests that most data consist of unconstrained binary relations:

- \( \mathcal{Y} = \{0, 1\}^{n(n-1)} \);
- \( P(\mathcal{Y}|\theta) = 1 \) for all \( \theta \in \Theta \);
- \( P(A|\theta) < 1 \) for any proper subset \( A \) of \( \mathcal{Y} \).

Raw relational data sources suggest otherwise: http://moreno.ss.uci.edu/data.html

A substantial fraction, if not majority, of relational data are “valued.”
A popular solution

If your observed data are non-binary, and your model is for binary data, then

▷ transform your data, so that it is binary:

\[ y_{i,j} = 1(y_{i,j}^{\text{obs}} > c) \]

**WARNING:** Dichotomizing can lead to information loss!
A popular solution

If your observed data are non-binary, and your model is for binary data, then

▷ transform your data, so that it is binary:

$$y_{i,j} = 1(y_{i,j}^{\text{obs}} > c)$$

**WARNING:** Dichotomizing can lead to information loss!
A popular solution

If your observed data are non-binary, and your model is for binary data, then

▷ transform your data, so that it is binary:

\[ y_{i,j} = 1(y_{i,j}^{\text{obs}} > c) \]

WARNING: Dichotomizing can lead to information loss!
A popular solution

If your observed data are non-binary, and your model is for binary data, then transform your data, so that it is binary:

\[ y_{i,j} = 1(y_{i,j}^{\text{obs}} > c) \]

**WARNING:** Dichotomizing can lead to information loss!
A popular solution

If your observed data are non-binary, and your model is for binary data, then
▷ transform your data, so that it is binary:

\[ y_{i,j} = 1(y_{i,j}^{\text{obs}} > c) \]

**WARNING:** Dichotomizing can lead to information loss!
If your observed data are non-binary, and your model is for binary data, then transform your data, so that it is binary:

\[ y_{i,j} = 1(y_{i,j}^{\text{obs}} > c) \]

**WARNING:** Dichotomizing can lead to information loss!
Example: High-school friendship network among females

Friendship relations among 646 female study participants.

Also have information on smoking, drinking, gpa, etc.
Fixed rank nomination scheme

Each study participant was asked to

- nominate up to 5 friends from a roster of students;
- rank-order their nominated friends.

Data coding:

\[ s_{i,j} = \text{student } i\text{'s score for student } j \]
\[ = 0 \text{ if } j \text{ is not nominated by } i \]
\[ = 1 \text{ if } j \text{ is } i\text{'s least favorite nominated friend} \]
\[ \vdots \]
\[ = 5^* \text{ if } j \text{ is } i\text{'s favorite nominated friend} \]

*if 5 friends are nominated.

Note: Relations are neither binary nor completely observed.
Fixed rank nomination scheme

Each study participant was asked to
  • nominate up to 5 friends from a roster of students;
  • rank-order their nominated friends.

Data coding:

\[ s_{i,j} = \begin{cases} 
  \text{student } i\text{'s score for student } j \\
  0 & \text{if } j \text{ is not nominated by } i \\
  1 & \text{if } j \text{ is } i\text{'s least favorite nominated friend} \\
  \vdots \\
  5 & \text{if } j \text{ is } i\text{'s favorite nominated friend}
\end{cases} \]

*if 5 friends are nominated.

Note: Relations are neither binary nor completely observed.
Fixed rank nomination scheme

Each study participant was asked to

- nominate up to 5 friends from a roster of students;
- rank-order their nominated friends.

**Data coding:**

\[ s_{i,j} = \begin{cases} 
    \text{student } i \text{'s score for student } j \\
    0 & \text{if } j \text{ is not nominated by } i \\
    1 & \text{if } j \text{ is } i \text{'s least favorite nominated friend} \\
    2 & \text{if } j \text{ is } i \text{'s second least favorite nominated friend} \\
    3 & \text{if } j \text{ is } i \text{'s third least favorite nominated friend} \\
    4 & \text{if } j \text{ is } i \text{'s fourth least favorite nominated friend} \\
    5 & \text{if } j \text{ is } i \text{'s favorite nominated friend} \\
\end{cases} \]

*if 5 friends are nominated.

**Note:** Relations are neither binary nor completely observed.
Fixed rank nomination scheme

Each study participant was asked to
- nominate up to 5 friends from a roster of students;
- rank-order their nominated friends.

Data coding:

\[ s_{i,j} = \text{student } i\text{’s score for student } j \]
\[ = 0 \text{ if } j \text{ is not nominated by } i \]
\[ = 1 \text{ if } j \text{ is } i\text{’s least favorite nominated friend} \]
\[ = 2 \]
\[ = 3 \]
\[ = 4 \]
\[ = 5^* \text{ if } j \text{ is } i\text{’s favorite nominated friend} \]

*if 5 friends are nominated.

Note: Relations are neither binary nor completely observed.
Each study participant was asked to

- nominate up to 5 friends from a roster of students;
- rank-order their nominated friends.

**Data coding:**

\[ s_{i,j} = \text{student } i's \text{ score for student } j \]

= 0 if \( j \) is not nominated by \( i \)

= 1 if \( j \) is \( i \)'s least favorite nominated friend

\[ \vdots \]

= 5* if \( j \) is \( i \)'s favorite nominated friend

* if 5 friends are nominated.

**Note:** Relations are neither binary nor completely observed.
Fixed rank nomination scheme

Each study participant was asked to

- nominate up to 5 friends from a roster of students;
- rank-order their nominated friends.

**Data coding:**

\[ s_{i,j} = \text{student } i \text{'s score for student } j \]

\[ = 0 \text{ if } j \text{ is not nominated by } i \]

\[ = 1 \text{ if } j \text{ is } i \text{'s least favorite nominated friend} \]

\[ \vdots \]

\[ = 5^* \text{ if } j \text{ is } i \text{'s favorite nominated friend} \]

*if 5 friends are nominated.

**Note:** Relations are neither binary nor completely observed.
Fixed rank nomination schemes

“list up to \( m \) of your best friends”

Such surveys are often used in studies of schools, work environments, etc.

- AddHealth (Harris et al., 2009)
- PROSPER (Moody et al., 2011)
- Netherlands School Study (Weerman and Smeenk, 2005)
- Sampson’s monastery (Sampson, 1969, Breiger et al., 1975)
Fixed rank nomination schemes

“list up to $m$ of your best friends”

Such surveys are often used in studies of schools, work environments, etc.

- AddHealth (Harris et al., 2009)
- PROSPER (Moody et al., 2011)
- Netherlands School Study (Weerman and Smeenk, 2005)
- Sampson’s monastery (Sampson, 1969, Breiger et al., 1975)
Fixed rank nomination schemes

“list up to $m$ of your best friends”

Such surveys are often used in studies of schools, work environments, etc.

- AddHealth (Harris et al., 2009)
- PROSPER (Moody et al., 2011)
- Netherlands School Study (Weerman and Smeenk, 2005)
- Sampson’s monastery (Sampson, 1969, Breiger et al., 1975)
Fixed rank nomination schemes

“list up to $m$ of your best friends”

Such surveys are often used in studies of schools, work environments, etc.

- AddHealth (Harris et al., 2009)
- PROSPER (Moody et al., 2011)
- Netherlands School Study (Weerman and Smeenk, 2005)
- Sampson’s monastery (Sampson, 1969, Breiger et al., 1975)
Fixed rank nomination schemes

“list up to $m$ of your best friends”

Such surveys are often used in studies of schools, work environments, etc.

- AddHealth (Harris et al., 2009)
- PROSPER (Moody et al., 2011)
- Netherlands School Study (Weerman and Smeenk, 2005)
- Sampson’s monastery (Sampson, 1969, Breiger et al., 1975)
Fixed rank nomination schemes

“list up to \( m \) of your best friends”

Such surveys are often used in studies of schools, work environments, etc.

- AddHealth (Harris et al., 2009)
- PROSPER (Moody et al., 2011)
- Netherlands School Study (Weerman and Smeenk, 2005)
- Sampson’s monastery (Sampson, 1969, Breiger et al., 1975)
Ordinal censored data

Frequently the survey design and rank information is ignored: For $m = 3$, 

$$
\begin{pmatrix}
\text{NA} & C & 2 & 1 & C & 3 \\
1 & \text{NA} & 0 & 2 & 0 & 0 \\
C & 3 & \text{NA} & 2 & C & 1 \\
0 & 0 & 0 & \text{NA} & 1 & 0 \\
C & 2 & 1 & C & \text{NA} & 3 \\
3 & C & C & 2 & 1 & \text{NA}
\end{pmatrix}
\rightarrow
\begin{pmatrix}
\text{NA} & 0 & 1 & 1 & 0 & 1 \\
1 & \text{NA} & 0 & 1 & 0 & 0 \\
0 & 1 & \text{NA} & 1 & 0 & 1 \\
0 & 0 & 0 & \text{NA} & 1 & 0 \\
0 & 1 & 1 & 0 & \text{NA} & 1 \\
1 & 0 & 0 & 1 & 1 & \text{NA}
\end{pmatrix}
$$

This is problematic for two reasons:

- ignoring the rank information is inefficient;
- ignoring the censoring can result in misleading inference.

What could go wrong?

**Intuition:**

$$y_{i,j} \sim \mu + \alpha x_i + \beta x_j + \gamma x_{i,j}$$

Ignoring censoring underestimates the outdegree $\rightarrow$ biased estimates of $\mu, \alpha$.

Throwing away ranks is throwing away information $\rightarrow$ less precise estimates.
Ordinal censored data

Frequently the survey design and rank information is ignored: For $m = 3$,

\[
\begin{pmatrix}
NA & C & 2 & 1 & C & 3 \\
1 & NA & 0 & 2 & 0 & 0 \\
C & 3 & NA & 2 & C & 1 \\
0 & 0 & 0 & NA & 1 & 0 \\
C & 2 & 1 & C & NA & 3 \\
3 & C & C & 2 & 1 & NA \\
\end{pmatrix}
\rightarrow
\begin{pmatrix}
NA & 0 & 1 & 1 & 0 & 1 \\
1 & NA & 0 & 1 & 0 & 0 \\
0 & 1 & NA & 1 & 0 & 1 \\
0 & 0 & 0 & NA & 1 & 0 \\
0 & 1 & 1 & 0 & NA & 1 \\
1 & 0 & 0 & 1 & 1 & NA \\
\end{pmatrix}
\]

This is problematic for two reasons:

- ignoring the rank information is inefficient;
- ignoring the censoring can result in misleading inference.

What could go wrong?

**Intuition:**

\[ y_{i,j} \sim \mu + \alpha x_i + \beta x_j + \gamma x_{i,j} \]

Ignoring censoring underestimates the outdegree $\rightarrow$ biased estimates of $\mu, \alpha$.

Throwing away ranks is throwing away information $\rightarrow$ less precise estimates.
Ordinal censored data

Frequently the survey design and rank information is ignored: For \( m = 3 \),

\[
\begin{pmatrix}
NA & C & 2 & 1 & C & 3 \\
1 & NA & 0 & 2 & 0 & 0 \\
C & 3 & NA & 2 & C & 1 \\
0 & 0 & 0 & NA & 1 & 0 \\
C & 2 & 1 & C & NA & 3 \\
3 & C & C & 2 & 1 & NA
\end{pmatrix}
\rightarrow
\begin{pmatrix}
NA & 0 & 1 & 1 & 0 & 1 \\
1 & NA & 0 & 1 & 0 & 0 \\
0 & 1 & NA & 1 & 0 & 1 \\
0 & 0 & 0 & NA & 1 & 0 \\
0 & 1 & 1 & 0 & NA & 1 \\
1 & 0 & 0 & 1 & 1 & NA
\end{pmatrix}
\]

This is problematic for two reasons:

- ignoring the rank information is inefficient;
- ignoring the censoring can result in misleading inference.

What could go wrong?

**Intuition:**

\[ y_{i,j} \sim \mu + \alpha x_i + \beta x_j + \gamma x_{i,j} \]

Ignoring censoring underestimates the outdegree \( \rightarrow \) biased estimates of \( \mu, \alpha \).

Throwing away ranks is throwing away information \( \rightarrow \) less precise estimates.
**Ordinal censored data**

Frequently the survey design and rank information is ignored: For $m = 3$,

\[
\begin{pmatrix}
NA & C & 2 & 1 & C & 3 \\
1 & NA & 0 & 2 & 0 & 0 \\
C & 3 & NA & 2 & C & 1 \\
0 & 0 & 0 & NA & 1 & 0 \\
C & 2 & 1 & C & NA & 3 \\
3 & C & C & 2 & 1 & NA
\end{pmatrix}
\rightarrow
\begin{pmatrix}
NA & 0 & 1 & 1 & 0 & 1 \\
1 & NA & 0 & 1 & 0 & 0 \\
0 & 1 & NA & 1 & 0 & 1 \\
0 & 0 & 0 & NA & 1 & 0 \\
0 & 1 & 1 & 0 & NA & 1 \\
1 & 0 & 0 & 1 & 1 & NA
\end{pmatrix}
\]

This is problematic for two reasons:

- ignoring the rank information is inefficient;
- ignoring the censoring can result in misleading inference.

What could go wrong?

**Intuition:**

\[y_{i,j} \sim \mu + \alpha x_i + \beta x_j + \gamma x_{i,j}\]

Ignoring censoring underestimates the outdegree \(\rightarrow\) biased estimates of $\mu, \alpha$.

Throwing away ranks is throwing away information \(\rightarrow\) less precise estimates.
Ordinal censored data

Frequently the survey design and rank information is ignored: For \( m = 3 \),

\[
\begin{pmatrix}
NA & C & 2 & 1 & C & 3 \\
1 & NA & 0 & 2 & 0 & 0 \\
C & 3 & NA & 2 & C & 1 \\
0 & 0 & 0 & NA & 1 & 0 \\
C & 2 & 1 & C & NA & 3 \\
3 & C & C & 2 & 1 & NA
\end{pmatrix}
\rightarrow
\begin{pmatrix}
NA & 0 & 1 & 1 & 0 & 0 \\
1 & NA & 0 & 1 & 0 & 0 \\
o & 1 & NA & 1 & 0 & 1 \\
o & 0 & 0 & NA & 1 & 0 \\
o & 1 & 1 & 0 & NA & 1 \\
1 & 0 & 0 & 1 & 1 & NA
\end{pmatrix}
\]

This is problematic for two reasons:

- ignoring the rank information is inefficient;
- ignoring the censoring can result in misleading inference.

What could go wrong?

Intuition:

\[ y_{i,j} \sim \mu + \alpha x_i + \beta x_j + \gamma x_{i,j} \]

Ignoring censoring underestimates the outdegree \( \Rightarrow \) biased estimates of \( \mu, \alpha \).

Throwing away ranks is throwing away information \( \Rightarrow \) less precise estimates.
Ordinal censored data

Frequently the survey design and rank information is ignored: For $m = 3$,

$\begin{pmatrix}
NA & C & 2 & 1 & C & 3 \\
1 & NA & 0 & 2 & 0 & 0 \\
C & 3 & NA & 2 & C & 1 \\
0 & 0 & 0 & NA & 1 & 0 \\
C & 2 & 1 & C & NA & 3 \\
3 & C & C & 2 & 1 & NA
\end{pmatrix} \rightarrow 
\begin{pmatrix}
NA & 0 & 1 & 1 & 0 & 1 \\
1 & NA & 0 & 1 & 0 & 0 \\
0 & 1 & NA & 1 & 0 & 1 \\
0 & 0 & 0 & NA & 1 & 0 \\
0 & 1 & 1 & 0 & NA & 1 \\
1 & 0 & 0 & 1 & 1 & NA
\end{pmatrix}$

This is problematic for two reasons:
- ignoring the rank information is inefficient;
- ignoring the censoring can result in misleading inference.

What could go wrong?

**Intuition:**

$y_{i,j} \sim \mu + \alpha x_i + \beta x_j + \gamma x_{i,j}$

Ignoring censoring underestimates the outdegree $\rightarrow$ biased estimates of $\mu, \alpha$.

Throwing away ranks is throwing away information $\rightarrow$ less precise estimates.
### Ordinal censored data

Frequently the survey design and rank information is ignored: For $m = 3$,

$$
\begin{pmatrix}
NA & C & 2 & 1 & C & 3 \\
1 & NA & 0 & 2 & 0 & 0 \\
C & 3 & NA & 2 & C & 1 \\
0 & 0 & 0 & NA & 1 & 0 \\
C & 2 & 1 & C & NA & 3 \\
3 & C & C & 2 & 1 & NA \\
\end{pmatrix}
\rightarrow
\begin{pmatrix}
NA & 0 & 1 & 1 & 0 & 1 \\
1 & NA & 0 & 1 & 0 & 0 \\
0 & 0 & C & 1 & 0 & 1 \\
0 & 0 & 0 & NA & 1 & 0 \\
0 & 1 & 1 & 0 & NA & 1 \\
1 & 0 & 0 & 1 & 1 & NA \\
\end{pmatrix}
$$

This is problematic for two reasons:

- ignoring the rank information is inefficient;
- ignoring the censoring can result in misleading inference.

What could go wrong?

**Intuition:**

$$
y_{i,j} \sim \mu + \alpha x_i + \beta x_j + \gamma x_{i,j}
$$

Ignoring censoring underestimates the outdegree $\rightarrow$ biased estimates of $\mu, \alpha$.

Throwing away ranks is throwing away information $\rightarrow$ less precise estimates.
FRN likelihood

\[ Y = \{ y_{i,j} : i \neq j \} \text{ a sociomatrix of ordinal relations (unobserved)} \]
\[ S = \{ s_{i,j} : i \neq j \} \text{ a sociomatrix of ranked nomination scores (observed)} \]

For the (positive) FRN survey design,

\[ s_{i,j} = [(m - \text{rank}_i(y_{i,j}) + 1) \wedge 0] \times 1(y_{i,j} > 0). \]

Perhaps more intuitively,

\[ s_{i,j} > 0 \implies y_{i,j} > 0 \tag{1} \]
\[ s_{i,j} = 0 \text{ and } d_i < m \implies y_{i,j} \leq 0 \tag{2} \]
\[ s_{i,j} > s_{i,k} \implies y_{i,j} > y_{i,k}. \tag{3} \]
FRN likelihood

\[ Y = \{y_{i,j} : i \neq j\} \text{ a sociomatrix of ordinal relations (unobserved)} \]
\[ S = \{s_{i,j} : i \neq j\} \text{ a sociomatrix of ranked nomination scores (observed)} \]

For the (positive) FRN survey design,

\[ s_{i,j} = [(m - \text{rank}_i(y_{i,j}) + 1) \wedge 0] \times 1(y_{i,j} > 0). \]

Perhaps more intuitively,

\[ s_{i,j} > 0 \Rightarrow y_{i,j} > 0 \quad (1) \]
\[ s_{i,j} = 0 \text{ and } d_i < m \Rightarrow y_{i,j} \leq 0 \quad (2) \]
\[ s_{i,j} > s_{i,k} \Rightarrow y_{i,j} > y_{i,k}. \quad (3) \]
FRN likelihood

\[ Y = \{y_{i,j} : i \neq j\} \text{ a sociomatrix of ordinal relations (unobserved)} \]
\[ S = \{s_{i,j} : i \neq j\} \text{ a sociomatrix of ranked nomination scores (observed)} \]

For the (positive) FRN survey design,

\[ s_{i,j} = [(m - \text{rank}_i(y_{i,j}) + 1) \land 0] \times 1(y_{i,j} > 0). \]

Perhaps more intuitively,

\[ s_{i,j} > 0 \iff y_{i,j} > 0 \quad (1) \]
\[ s_{i,j} = 0 \text{ and } d_i < m \iff y_{i,j} \leq 0 \quad (2) \]
\[ s_{i,j} > s_{i,k} \iff y_{i,j} > y_{i,k}. \quad (3) \]
FRN likelihood

\[ F(S) = \{Y : (1) \text{ (2) and (3) hold } \} \]

Observation of \( S \) implies \( Y \in F(S) \).

Given a model \( \{p_\theta : \theta \in \Theta\} \) for \( Y \), our likelihood is then

\[
L_F(S : \theta) = \Pr(Y \in F(S)|\theta) \\
= \int_{F(S)} p(Y|\theta) \, d\mu(Y).
\]
FRN likelihood

\[ F(S) = \{ Y : (1) \text{ (2) and (3) hold } \} \]

Observation of \( S \) implies \( Y \in F(S) \).

Given a model \( \{p_\theta : \theta \in \Theta\} \) for \( Y \), our likelihood is then

\[
L_F(S : \theta) = \Pr(Y \in F(S)|\theta) \\
= \int_{F(S)} p(Y|\theta) \, d\mu(Y).
\]
Other likelihoods

**Rank likelihood**

\[
R(S) = \{ Y : s_{i,j} > s_{i,k} \Rightarrow y_{i,j} > y_{i,k} \}
\]

\[
L_R(\theta : S) = \Pr(Y \in R(S)|\theta)
\]

**Binary likelihood**

\[
B(S) = \{ Y : s_{i,j} > 0 \Rightarrow y_{i,j} > 0, s_{i,j} = 0 \Rightarrow y_{i,j} \leq 0 \}
\]

\[
L_B(\theta : S) = \Pr(Y \in B(S)|\theta)
\]
Other likelihoods

**Rank likelihood**

\[ R(S) = \{ Y : s_{i,j} > s_{i,k} \Rightarrow y_{i,j} > y_{i,k} \} \]

\[ L_R(\theta : S) = \Pr(Y \in R(S)|\theta) \]

**Binary likelihood**

\[ B(S) = \{ Y : s_{i,j} > 0 \Rightarrow y_{i,j} > 0, s_{i,j} = 0 \Rightarrow y_{i,j} \leq 0 \} \]

\[ L_B(\theta : S) = \Pr(Y \in B(S)|\theta) \]
Use of the rank likelihood is
- “valid”, but
- not fully informative.
Use of the rank likelihood is

- “valid”, but
- not fully informative.
Use of the rank likelihood is

- "valid", but
- not fully informative.
Use of the rank likelihood is

- "valid", but
- not fully informative.
Relationship among likelihoods

\[ F(S) \subset R(S) \]

Use of the rank likelihood is
- “valid”, but
- not fully informative.
Multiplicative effects - higher order dependence

Relationship among likelihoods

$$F(S) \nsubseteq B(S)$$

Use of the binary likelihood is not valid

- not valid
Relationship among likelihoods

\[ F(S) \not\subset B(S) \]

Use of the binary likelihood is

- not valid
Estimation for set-based likelihoods

**Model:** $Y \sim p(Y | \theta), \ \theta \in \Theta$

**Data:** $Y \in F(S)$

**Estimation:** Given $p(\theta)$, $p(\theta | Y \in F(S))$ can be approximated with MCMC.

**Gibbs sampler:**

1. Simulate $\theta \sim p(\theta | Y)$.
2. For each pair $(i, j)$, simulate $y_{i,j} \sim p(y_{i,j} | \theta, Y_{-(i,j)}, Y \in F(S))$ as follows:
   
   2.1 if $s_{i,j} > 0$ simulate
   
   $$y_{i,j} \sim p(y_{i,j} | Y_{-(i,j)}, \theta) \times 1(\max\{y_{i,k} : s_{i,k} < s_{i,j}\} \leq y_{i,j} \leq \min\{y_{i,k} : s_{i,k} > s_{i,j}\});$$

   2.2 if $s_{i,j} = 0$ and $d_i < m$, simulate $y_{i,j} \sim p(y_{i,j} | Y_{-(i,j)}, \theta) \times 1(y_{i,j} \leq 0)$;

   2.3 if $s_{i,j} = 0$ and $d_i = m$, simulate
   
   $$y_{i,j} \sim p(y_{i,j} | Y_{-(i,j)}, \theta) \times 1(y_{i,j} \leq \min\{y_{i,k} : s_{i,k} > 0\}).$$
Estimation for set-based likelihoods

Model: $Y \sim p(Y|\theta), \ \theta \in \Theta$

Data: $Y \in F(S)$

Estimation: Given $p(\theta)$, $p(\theta|Y \in F(S))$ can be approximated with MCMC.

Gibbs sampler:

1. Simulate $\theta \sim p(\theta|Y)$.
2. For each pair $(i, j)$, simulate $y_{i,j} \sim p(y_{i,j}|\theta, Y_{-(i,j)}, Y \in F(S))$ as follows:
   2.1 if $s_{i,j} > 0$ simulate
      $$y_{i,j} \sim p(y_{i,j}|Y_{-(i,j)}, \theta) \times 1(\max\{y_{i,k}: s_{i,k} < s_{i,j}\} \leq y_{i,j} \leq \min\{y_{i,k}: s_{i,k} > s_{i,j}\});$$
   2.2 if $s_{i,j} = 0$ and $d_i < m$, simulate $y_{i,j} \sim p(y_{i,j}|Y_{-(i,j)}, \theta) \times 1(y_{i,j} \leq 0)$;
   2.3 if $s_{i,j} = 0$ and $d_i = m$, simulate
      $$y_{i,j} \sim p(y_{i,j}|Y_{-(i,j)}, \theta) \times 1(y_{i,j} \leq \min\{y_{i,k}: s_{i,k} > 0\}).$$
Estimation for set-based likelihoods

**Model:** \( \mathbf{Y} \sim p(\mathbf{Y}|\theta), \ \theta \in \Theta \)

**Data:** \( \mathbf{Y} \in F(S) \)

**Estimation:** Given \( p(\theta), p(\theta|\mathbf{Y} \in F(S)) \) can be approximated with MCMC.

**Gibbs sampler:**

1. Simulate \( \theta \sim p(\theta|\mathbf{Y}) \).
2. For each pair \((i,j)\), simulate \( y_{i,j} \sim p(y_{i,j}|\theta, \mathbf{Y}_{-(i,j)}, \mathbf{Y} \in F(S)) \) as follows:
   2.1 if \( s_{i,j} > 0 \) simulate
   
   \[ y_{i,j} \sim p(y_{i,j}|\mathbf{Y}_{-(i,j)}, \theta) \times 1(\max\{y_{i,k}: s_{i,k} < s_{i,j}\} \leq y_{i,j} \leq \min\{y_{i,k}: s_{i,k} > s_{i,j}\}); \]
   
   2.2 if \( s_{i,j} = 0 \) and \( d_i < m \), simulate \( y_{i,j} \sim p(y_{i,j}|\mathbf{Y}_{-(i,j)}, \theta) \times 1(y_{i,j} \leq 0); \)
   
   2.3 if \( s_{i,j} = 0 \) and \( d_i = m \), simulate
   
   \[ y_{i,j} \sim p(y_{i,j}|\mathbf{Y}_{-(i,j)}, \theta) \times 1(y_{i,j} \leq \min\{y_{i,k}: s_{i,k} > 0\}); \]
Estimation for set-based likelihoods

Model: \( \mathbf{Y} \sim p(\mathbf{Y}|\theta), \ \theta \in \Theta \)

Data: \( \mathbf{Y} \in F(\mathbf{S}) \)

Estimation: Given \( p(\theta), p(\theta|\mathbf{Y} \in F(\mathbf{S})) \) can be approximated with MCMC.

Gibbs sampler:

1. Simulate \( \theta \sim p(\theta|\mathbf{Y}) \).
2. For each pair \((i, j)\), simulate \( y_{i,j} \sim p(y_{i,j}|\theta, \mathbf{Y}_{-(i,j)}, \mathbf{Y} \in F(\mathbf{S})) \) as follows:
   2.1 if \( s_{i,j} > 0 \) simulate
   \[
   y_{i,j} \sim p(y_{i,j}|\mathbf{Y}_{-(i,j)}, \theta) \times 1(\max\{y_{i,k}: s_{i,k} < s_{i,j}\} \leq y_{i,j} \leq \min\{y_{i,k}: s_{i,k} > s_{i,j}\});
   \]
   2.2 if \( s_{i,j} = 0 \) and \( d_i < m \), simulate \( y_{i,j} \sim p(y_{i,j}|\mathbf{Y}_{-(i,j)}, \theta) \times 1(y_{i,j} \leq 0) \);
   2.3 if \( s_{i,j} = 0 \) and \( d_i = m \), simulate
   \[
   y_{i,j} \sim p(y_{i,j}|\mathbf{Y}_{-(i,j)}, \theta) \times 1(y_{i,j} \leq \min\{y_{i,k}: s_{i,k} > 0\});
   \]
Social relations regression model

\[ y_{i,j} = \beta^T x_{i,j} + a_i + b_j + \epsilon_{i,j} \]

\[
\left( \begin{array}{c}
\epsilon_{i,j} \\
\epsilon_{j,i}
\end{array} \right), i \neq j \sim \text{i.i.d. normal}(0, \Sigma) \\
\left( \begin{array}{c}
a_i \\
b_i
\end{array} \right), i = 1, \ldots, n \sim \text{i.i.d. normal}(0, \Sigma_{ab})
\]

- \( \beta \) represents covariate effects;
- \((a_i, b_i)\) represents additive sender and receiver effects;
- \(\text{cov}(\epsilon_{i,j}, \epsilon_{j,i})\) represents reciprocity.

Multiplicative effects - higher order dependence

**Social relations regression model**

\[ y_{i,j} = \beta^T x_{i,j} + a_i + b_j + \epsilon_{i,j} \]

\[ (\begin{pmatrix} e_{i,j} \\ e_{j,i} \end{pmatrix}, i \neq j) \sim \text{i.i.d. normal}(0, \Sigma_\epsilon) \]
\[ (\begin{pmatrix} a_i \\ b_i \end{pmatrix}, i = 1, \ldots, n) \sim \text{i.i.d. normal}(0, \Sigma_{ab}) \]

- \( \beta \) represents covariate effects;
- \((a_i, b_i)\) represents additive sender and receiver effects;
- \( \text{cov}(\epsilon_{i,j}, \epsilon_{j,i}) \) represents reciprocity.

Social relations regression model

\[ y_{i,j} = \beta^T x_{i,j} + a_i + b_j + \epsilon_{i,j} \]

\[
\begin{align*}
((\epsilon_{i,j}^i, i \neq j) &\sim \text{i.i.d. normal}(0, \Sigma_\epsilon) \\
((a_i, b_i), i = 1, \ldots, n) &\sim \text{i.i.d. normal}(0, \Sigma_{ab})
\end{align*}
\]

- \( \beta \) represents covariate effects;
- \((a_i, b_i)\) represents additive sender and receiver effects;
- \(\text{cov}(\epsilon_{i,j}, \epsilon_{j,i})\) represents reciprocity.

Multiplicative effects - higher order dependence

Social relations regression model

$$y_{i,j} = \beta^T x_{i,j} + a_i + b_j + \epsilon_{i,j}$$

$$
\left(\begin{array}{c}
\epsilon_{i,j} \\
\epsilon_{j,i}
\end{array}\right), i \neq j \sim \text{i.i.d. normal}(0, \Sigma_\epsilon) \\
\left(\begin{array}{c}
a_i \\
b_i
\end{array}\right), i = 1, \ldots, n \sim \text{i.i.d. normal}(0, \Sigma_{ab})
$$

- $\beta$ represents covariate effects;
- $(a_i, b_i)$ represents additive sender and receiver effects;
- $\text{cov}(\epsilon_{i,j}, \epsilon_{j,i})$ represents reciprocity.

Social relations regression model

\[ y_{i,j} = \beta^T x_{i,j} + a_i + b_j + \epsilon_{i,j} \]

\[
\begin{align*}
((\epsilon_{i,j}, i \neq j) & \sim \text{i.i.d. normal}(0, \Sigma_\epsilon) \\
((a_i, b_i), i = 1, \ldots, n) & \sim \text{i.i.d. normal}(0, \Sigma_{ab})
\end{align*}
\]

- $\beta$ represents covariate effects;
- $(a_i, b_i)$ represents additive sender and receiver effects;
- $\text{cov}(\epsilon_{i,j}, \epsilon_{j,i})$ represents reciprocity.

Social relations regression model

\[ y_{i,j} = \beta^T x_{i,j} + a_i + b_j + \epsilon_{i,j} \]

\[
\left( \begin{array}{c}
\epsilon_{i,j} \\
\epsilon_{j,i}
\end{array} \right), \ i \neq j \sim \text{i.i.d. normal}(0, \Sigma_{\epsilon})
\]

\[
\left( \begin{array}{c}
a_i \\
b_i
\end{array} \right), \ i = 1, \ldots, n \sim \text{i.i.d. normal}(0, \Sigma_{ab})
\]

- \( \beta \) represents covariate effects;
- \( (a_i, b_i) \) represents additive sender and receiver effects;
- \( \text{cov}(\epsilon_{i,j}, \epsilon_{j,i}) \) represents reciprocity.

Simulation study

\[ \mathbf{\beta}^T \mathbf{x}_{i,j} = \beta_0 + \beta_r x_{i,r} + \beta_c x_{j,c} + \beta_{d1} x_{i,j,1} + \beta_{d2} x_{i,j,2} \]

- \( x_{i,r} \), \( x_{j,c} \): individual-level variables (smoking behavior, gpa)
- \( x_{i,j,1} \): pair-specific variable (time spent together)
- \( x_{i,j,2} \): co-membership in a group (in the same sport or club)

How well do the different likelihoods estimate \( \mathbf{\beta} \)?
Simulation study

\[ \beta^T x_{i,j} = \beta_0 + \beta_r x_{i,r} + \beta_c x_{j,c} + \beta_{d1} x_{i,j,1} + \beta_{d2} x_{i,j,2} \]

- \( x_{i,r}, x_{j,c} \): individual-level variables (smoking behavior, gpa)
- \( x_{i,j,1} \): pair-specific variable (time spent together)
- \( x_{i,j,2} \): co-membership in a group (in the same sport or club)

How well do the different likelihoods estimate \( \beta \)?
Multiplicative effects - higher order dependence

Simulation study

\[ \mathbf{\beta}^T \mathbf{x}_{i,j} = \beta_0 + \beta_r x_{i,r} + \beta_c x_{j,c} + \beta_{d_1} x_{i,j,1} + \beta_{d_2} x_{i,j,2} \]

- \( x_{i,r}, x_{j,c} \): individual-level variables (smoking behavior, gpa)
- \( x_{i,j,1} \): pair-specific variable (time spent together)
- \( x_{i,j,2} \): co-membership in a group (in the same sport or club)

How well do the different likelihoods estimate \( \mathbf{\beta} \)?
Simulation study

$$\beta^T x_{i,j} = \beta_0 + \beta_r x_{i,r} + \beta_c x_{j,c} + \beta_{d1} x_{i,j,1} + \beta_{d2} x_{i,j,2}$$

- $x_{i,r}, x_{j,c}$: individual-level variables (smoking behavior, gpa)
- $x_{i,j,1}$: pair-specific variable (time spent together)
- $x_{i,j,2}$: co-membership in a group (in the same sport or club)

How well do the different likelihoods estimate $\beta$?
Simulation study

\[ \mathbf{\beta}^T \mathbf{x}_{i,j} = \beta_0 + \beta_r x_{i,r} + \beta_c x_{j,c} + \beta_{d1} x_{i,j,1} + \beta_{d2} x_{i,j,2} \]

- \( x_{i,r}, x_{j,c} \): individual-level variables (smoking behavior, gpa)
- \( x_{i,j,1} \): pair-specific variable (time spent together)
- \( x_{i,j,2} \): co-membership in a group (in the same sport or club)

How well do the different likelihoods estimate \( \mathbf{\beta} \)?
Simulation study

**Simulation:** 16 100-node sociomatrices, 8 under \( m = 5 \), 8 under \( m = 15 \).
Simulation study

Results:

- $L_F$ and $L_R$ provide accurate estimation for $\beta$;
- $L_B$ provides “reasonable” inference for $\beta_c$ and $\beta_{d1}$;
- $L_B$ provides poor inference for $\beta_0$, $\beta_r$ and $\beta_{d2}$.

Explanation: When the amount of censoring is large,

- heterogeneity of censored outdegrees is low;
- effects based on row heterogeneity will be underestimated by $L_B$;
- $L_B$ estimates are “confidently” underestimated.
Simulation study

**Results:**

- $L_F$ and $L_R$ provide accurate estimation for $\beta$;
- $L_B$ provides “reasonable” inference for $\beta_c$ and $\beta_d$;
- $L_B$ provides poor inference for $\beta_0$, $\beta_r$ and $\beta_d$.

**Explanation:** When the amount of censoring is large,

- heterogeneity of censored outdegrees is low;
- effects based on row heterogeneity will be underestimated by $L_B$;
- $L_B$ estimates are “confidently” underestimated.
Simulation study

Results:

- $L_F$ and $L_R$ provide accurate estimation for $\beta$;
- $L_B$ provides “reasonable” inference for $\beta_c$ and $\beta_{d1}$;
- $L_B$ provides poor inference for $\beta_0$, $\beta_r$ and $\beta_{d2}$.

Explanation: When the amount of censoring is large,
- heterogeneity of censored outdegrees is low;
- effects based on row heterogeneity will be underestimated by $L_B$;
- $L_B$ estimates are “confidently” underestimated.
Simulation study

Results:

- $L_F$ and $L_R$ provide accurate estimation for $\beta$;
- $L_B$ provides “reasonable” inference for $\beta_c$ and $\beta_{d1}$;
- $L_B$ provides poor inference for $\beta_0$, $\beta_r$ and $\beta_{d2}$.

Explanation: When the amount of censoring is large,

- heterogeneity of censored outdegrees is low;
- effects based on row heterogeneity will be underestimated by $L_B$;
- $L_B$ estimates are “confidently” underestimated.
Results:

- $L_F$ and $L_R$ provide accurate estimation for $\beta$;
- $L_B$ provides “reasonable” inference for $\beta_c$ and $\beta_{d1}$;
- $L_B$ provides poor inference for $\beta_0$, $\beta_r$ and $\beta_{d2}$.

Explanation: When the amount of censoring is large,

- heterogeneity of censored outdegrees is low;
- effects based on row heterogeneity will be underestimated by $L_B$;
- $L_B$ estimates are “confidently” underestimated.
Simulation study

Results:

- $L_F$ and $L_R$ provide accurate estimation for $\beta$;
- $L_B$ provides “reasonable” inference for $\beta_c$ and $\beta_{d1}$;
- $L_B$ provides poor inference for $\beta_0$, $\beta_r$ and $\beta_{d2}$.

Explanation: When the amount of censoring is large,

- heterogeneity of censored outdegrees is low;
- effects based on row heterogeneity will be underestimated by $L_B$;
- $L_B$ estimates are “confidently” underestimated.
Simulation study

Results:

- $L_F$ and $L_R$ provide accurate estimation for $\beta$;
- $L_B$ provides “reasonable” inference for $\beta_c$ and $\beta_d$;
- $L_B$ provides poor inference for $\beta_0$, $\beta_r$ and $\beta_{d_2}$.

Explanation: When the amount of censoring is large,

- heterogeneity of censored outdegrees is low;
- effects based on row heterogeneity will be underestimated by $L_B$;
- $L_B$ estimates are “confidently” underestimated.
Simulation study

Results:

- $L_F$ and $L_R$ provide accurate estimation for $\beta$;
- $L_B$ provides “reasonable” inference for $\beta_c$ and $\beta_{d1}$;
- $L_B$ provides poor inference for $\beta_0$, $\beta_r$ and $\beta_{d2}$.

Explanation: When the amount of censoring is large,

- heterogeneity of censored outdegrees is low;
- effects based on row heterogeneity will be underestimated by $L_B$;
- $L_B$ estimates are “confidently” underestimated.
Simulation study

Results:

- $L_F$ and $L_R$ provide accurate estimation for $\beta$;
- $L_B$ provides “reasonable” inference for $\beta_c$ and $\beta_{d1}$;
- $L_B$ provides poor inference for $\beta_0$, $\beta_r$ and $\beta_{d2}$.

Explanation: When the amount of censoring is large,

- heterogeneity of censored outdegrees is low;
- effects based on row heterogeneity will be underestimated by $L_B$;
- $L_B$ estimates are “confidently” underestimated.
Simulation study

Results:

- $L_F$ and $L_R$ provide accurate estimation for $\beta$;
- $L_B$ provides "reasonable" inference for $\beta_c$ and $\beta_{d1}$;
- $L_B$ provides poor inference for $\beta_0$, $\beta_r$ and $\beta_{d2}$.

Explanation: When the amount of censoring is large,

- heterogeneity of censored outdegrees is low;
- effects based on row heterogeneity will be underestimated by $L_B$;
- $L_B$ estimates are "confidently" underestimated.
Information in the ranks

We can construct a modified binary likelihood to account for censoring:

\[ s_{i,j} > 0 \Rightarrow y_{i,j} > 0 \]  \hspace{1cm} (1)
\[ s_{i,j} = 0 \text{ and } d_i < m \Rightarrow y_{i,j} \leq 0 \]  \hspace{1cm} (2)
\[ \min\{y_{i,j} : s_{i,j} > 0\} \geq \max\{y_{i,j} : s_{i,j} = 0\}. \]  \hspace{1cm} (4)

Censored binary likelihood:

\[ C(S) = \{Y : (1), (2) \text{ and } (4) \text{ hold}\} \]
\[ L_C(S) = \Pr(Y \in C(S) | \theta) \]

Comparison to \( L_F \):

(1) recognizes scored relations as positive
(2) recognizes the censoring
(4) recognizes scored relations as larger than unscored

Condition (4) does not differentiate among the scored relations.
Information in the ranks

We can construct a modified binary likelihood to account for censoring:

\[ s_{i,j} > 0 \Rightarrow y_{i,j} > 0 \]  \hspace{1cm} (1)
\[ s_{i,j} = 0 \text{ and } d_i < m \Rightarrow y_{i,j} \leq 0 \]  \hspace{1cm} (2)
\[ \min\{y_{i,j} : s_{i,j} > 0\} \geq \max\{y_{i,j} : s_{i,j} = 0\} \]  \hspace{1cm} (4)

**Censored binary likelihood:**

\[ C(S) = \{Y : (1), (2) \text{ and } (4) \text{ hold}\} \]
\[ L_C(S) = Pr(Y \in C(S)|\theta) \]

**Comparison to \( L_F \):**

(1) recognizes scored relations as positive
(2) recognizes the censoring
(4) recognizes scored relations as larger than unscored

Condition (4) does not differentiate among the scored relations.
Information in the ranks

We can construct a modified binary likelihood to account for censoring:

\[ s_{i,j} > 0 \Rightarrow y_{i,j} > 0 \]  \hspace{1cm} (1)
\[ s_{i,j} = 0 \text{ and } d_i < m \Rightarrow y_{i,j} \leq 0 \]  \hspace{1cm} (2)
\[ \min\{y_{i,j} : s_{i,j} > 0\} \geq \max\{y_{i,j} : s_{i,j} = 0\}. \]  \hspace{1cm} (4)

**Censored binary likelihood:**

\[ C(S) = \{ Y : (1), (2) \text{ and } (4) \text{ hold} \} \]
\[ L_C(S) = \Pr(Y \in C(S)|\theta) \]

**Comparison to \( L_F \):**

(1) recognizes scored relations as positive
(2) recognizes the censoring
(4) recognizes scored relations as larger than unscored

Condition (4) does not differentiate among the scored relations.
Information in the ranks

We can construct a modified binary likelihood to account for censoring:

\[ s_{i,j} > 0 \Rightarrow y_{i,j} > 0 \]  \hspace{1cm} (1)

\[ s_{i,j} = 0 \text{ and } d_i < m \Rightarrow y_{i,j} \leq 0 \]  \hspace{1cm} (2)

\[ \min\{y_{i,j} : s_{i,j} > 0\} \geq \max\{y_{i,j} : s_{i,j} = 0\}. \]  \hspace{1cm} (4)

Censored binary likelihood:

\[ C(S) = \{Y : \text{(1), (2) and (4) hold}\} \]

\[ L_C(S) = \Pr(Y \in C(S)|\theta) \]

Comparison to \( L_F \):

(1) recognizes scored relations as positive
(2) recognizes the censoring
(4) recognizes scored relations as larger than unscored

Condition (4) does not differentiate among the scored relations.
Information in the ranks

We can construct a modified binary likelihood to account for censoring:

\[ s_{i,j} > 0 \Rightarrow y_{i,j} > 0 \]  \hspace{1cm} (1)
\[ s_{i,j} = 0 \text{ and } d_i < m \Rightarrow y_{i,j} \leq 0 \]  \hspace{1cm} (2)
\[ \min\{y_{i,j} : s_{i,j} > 0\} \geq \max\{y_{i,j} : s_{i,j} = 0\} \]  \hspace{1cm} (4)

Censored binary likelihood:

\[ C(S) = \{Y : (1), (2) \text{ and } (4) \text{ hold}\} \]
\[ L_C(S) = \Pr(Y \in C(S) | \theta) \]

Comparison to \( L_F \):

(1) recognizes scored relations as positive
(2) recognizes the censoring
(4) recognizes scored relations as larger than unscored

Condition (4) does not differentiate among the scored relations.
Information in the ranks

We can construct a modified binary likelihood to account for censoring:

\[
\begin{align*}
    s_{i,j} > 0 & \Rightarrow y_{i,j} > 0 \\
    s_{i,j} = 0 \text{ and } d_i < m & \Rightarrow y_{i,j} \leq 0 \\
    \min\{y_{i,j} : s_{i,j} > 0\} & \geq \max\{y_{i,j} : s_{i,j} = 0\}.
\end{align*}
\]

Censored binary likelihood:

\[
C(S) = \{Y : (1), (2) \text{ and } (4) \text{ hold}\}
\]

\[
L_C(S) = \Pr(Y \in C(S)|\theta)
\]

Comparison to \( L_F \):

1. recognizes scored relations as positive
2. recognizes the censoring
3. recognizes scored relations as larger than unscored

Condition (4) does not differentiate among the scored relations.
Information in the ranks

We can construct a modified binary likelihood to account for censoring:

\[ s_{i,j} > 0 \Rightarrow y_{i,j} > 0 \]  \hfill (1)
\[ s_{i,j} = 0 \text{ and } d_i < m \Rightarrow y_{i,j} \leq 0 \]  \hfill (2)
\[ \min\{y_{i,j} : s_{i,j} > 0\} \geq \max\{y_{i,j} : s_{i,j} = 0\}. \]  \hfill (4)

**Censored binary likelihood:**

\[ C(S) = \{Y : \text{(1), (2) and (4) hold}\} \]
\[ L_C(S) = \Pr(Y \in C(S)|\theta) \]

**Comparison to** \( L_F \):

(1) recognizes scored relations as positive
(2) recognizes the censoring
(4) recognizes scored relations as larger than unscored

Condition (4) does not differentiate among the scored relations.
Simulation study

Posterior concentration around truth across $8 \times 4$ simulated datasets

Conclusion: Accounting for censoring alone may be adequate if $m \ll n$. 
**Simulation study**

Posterior concentration around truth across $8 \times 4$ simulated datasets

**Conclusion:** Accounting for censoring alone may be adequate if $m \ll n$. 
Adolescent health data

\[
E[y_{i,j}|\beta, x_{i,j}] = \beta^T x_{i,j} = \beta_r^T x_{r,i} + \beta_c^T x_{c,j} + \beta_d^T x_{d,i,j}
\]

\[
x_i = (r\text{smoke}_i, r\text{drink}_i, r\text{gpa}_i)
\]
\[
x_j = (c\text{smoke}_j, c\text{drink}_j, c\text{gpa}_j)
\]
\[
x_{i,j} = (d\text{smoke}_{i,j}, d\text{drink}_{i,j}, d\text{gpa}_{i,j},
\text{dacad}_{i,j}, \text{darts}_{i,j}, \text{dsports}_{i,j}, \text{dcivic}_{i,j}, \text{dgrade}_{i,j}, \text{drace}_{i,j})
\]
Adolescent health data

\[ E[y_{i,j} | \beta, x_{i,j}] = \beta^T x_{i,j} = \beta_r^T x_{r,i} + \beta_c^T x_{c,j} + \beta_d^T x_{d,i,j} \]

\[ x_i = (r\text{smoke}_i, r\text{drink}_i, r\text{gpa}_i) \]
\[ x_j = (c\text{smoke}_j, c\text{drink}_j, c\text{gpa}_j) \]
\[ x_{i,j} = (d\text{smoke}_{i,j}, d\text{drink}_{i,j}, d\text{gpa}_{i,j}, \]
\[ d\text{acad}_{i,j}, d\text{darts}_{i,j}, d\text{sports}_{i,j}, d\text{civic}_{i,j}, d\text{grade}_{i,j}, d\text{race}_{i,j} ) \]
Parameter estimation results

- Multiplicative effects - higher order dependence

- Quantitative variables with estimates and confidence intervals:
  - intercept: $\beta = -3.65$
  - rsmoke, rdrink, rgpa:
    - $\beta$ estimates and confidence intervals shown
  - csmoke, cdrink, cgpa:
    - $\beta$ estimates and confidence intervals shown
  - dsmoke, ddrink, dgpa:
    - $\beta$ estimates and confidence intervals shown
  - dacad, darts, dsport, dcivic:
    - $\beta$ estimates and confidence intervals shown
  - dgrade, drace:
    - $\beta$ estimates and confidence intervals shown
Results across 16 different networks

<table>
<thead>
<tr>
<th>Likelihood</th>
<th>intercept</th>
<th>row</th>
<th>column</th>
<th>mean-zero dyadic</th>
<th>other dyadic</th>
</tr>
</thead>
<tbody>
<tr>
<td>binomial</td>
<td>0.89, 1.68</td>
<td>2.22, 2.95</td>
<td>1.02, 1.03</td>
<td>1.06, 1.06</td>
<td>1.20, 1.09</td>
</tr>
<tr>
<td>rank</td>
<td>NA,NA</td>
<td>NA,NA</td>
<td>1.05, 0.98</td>
<td>0.99, 0.99</td>
<td>1.06, 0.98</td>
</tr>
</tbody>
</table>

Average relative magnitudes of parameter estimates (first number) and confidence interval widths (second number) from $L_F$ compared to $L_B$ and $L_R$.

**Summary:** Under $L_B$,
- $\hat{\beta}_0$ is too negative, se($\hat{\beta}_0$) is too small;
- $|\hat{\beta}_r|$ is too small, se($\hat{\beta}_r$) is too small.
### Results across 16 different networks

<table>
<thead>
<tr>
<th>Likelihood</th>
<th>intercept</th>
<th>row</th>
<th>column</th>
<th>mean-zero dyadic</th>
<th>other dyadic</th>
</tr>
</thead>
<tbody>
<tr>
<td>binomial</td>
<td>0.89, 1.68</td>
<td>2.22, 2.95</td>
<td>1.02, 1.03</td>
<td>1.06, 1.06</td>
<td>1.20, 1.09</td>
</tr>
<tr>
<td>rank</td>
<td>NA,NA</td>
<td>NA,NA</td>
<td>1.05, 0.98</td>
<td>0.99, 0.99</td>
<td>1.06, 0.98</td>
</tr>
</tbody>
</table>

Average relative magnitudes of parameter estimates (first number) and confidence interval widths (second number) from $L_F$ compared to $L_B$ and $L_R$.

**Summary:** Under $L_B$,

- $\hat{\beta}_0$ is too negative, $\text{se}(\hat{\beta}_0)$ is too small;
- $|\hat{\beta}_r|$ is too small, $\text{se}(\hat{\beta}_r)$ is too small.
AME fit for fixed rank nomination data

Description:

An MCMC routine providing a fit to an additive and multiplicative effects (AME) regression model for fixed rank nomination data

Usage:

```r
ame_frn(Y, X=NULL, Xrow=NULL, Xcol=NULL, odmax = rep(max(apply(Y > 0, 1, sum, na.rm = TRUE)), nrow(Y)), rvar = TRUE, cvar = TRUE, dcor = TRUE, R = 0, seed = 1, nscan = 50000, burn = 500, odens = 25, plot = TRUE, print = TRUE, intercept=TRUE)
```

Arguments:

- **Y**: an n x n square relational matrix of ranked nominations, where a higher rank indicates a stronger relationship

- **X**: an n x n x p array of covariates

- **Xrow**: an n x pr matrix of nodal row covariates

- **Xcol**: an n x pc matrix of nodal column covariates

- **odmax**: a scalar integer or vector of length n giving the maximum number of nominations that each node may make
Summary

- Design matters: inference requires data and context.
- Recognizing censored and ranked data improves validity and precision.
- Inference for a variety of rank- and censored-data likelihoods:
  *amen* package in R
Summary

- Design matters: inference requires data *and* context.
- Recognizing censored and ranked data improves validity and precision.
- Inference for a variety of rank- and censored-data likelihoods:
  
  `amen` package in R
Summary

- Design matters: inference requires data \textit{and} context.
- Recognizing censored and ranked data improves validity and precision.
- Inference for a variety of rank- and censored-data likelihoods:
  \texttt{amen} package in R
Summary

- Design matters: inference requires data \textit{and} context.
- Recognizing censored and ranked data improves validity and precision.
- Inference for a variety of rank- and censored-data likelihoods:
  \texttt{amen package in R}
Inferential approach

- **Statistical inference** utilizes **probability models**
- **Networks and relational data** are represented by **matrices and arrays**

Social network analysis can utilize probability models of matrices and arrays.

\[ Y \sim M + E \]

- **M** is a low-dimensional matrix of network patterns;
- **E** is patternless noise.

We will describe **M** via

- regression models
- additive random effects (ANOVA/SRM)
- multiplicative random effects (SVD/Eigen)
Inferential approach

- **Statistical inference** utilizes **probability models**
- **Networks and relational data** are represented by **matrices and arrays**

Social network analysis can utilize **probability models of matrices and arrays**.

\[ Y \sim M + E \]

- **M** is a low-dimensional matrix of network patterns;
- **E** is patternless noise.

We will describe **M** via
- regression models
- additive random effects (ANOVA/SRM)
- multiplicative random effects (SVD/Eigen)
Inferential approach

- Statistical inference utilizes probability models
- Networks and relational data are represented by matrices and arrays

Social network analysis can utilize probability models of matrices and arrays.

\[ Y \sim M + E \]

- \( M \) is a low-dimensional matrix of network patterns;
- \( E \) is patternless noise.

We will describe \( M \) via

- regression models
  - additive random effects (ANOVA/SRM)
  - multiplicative random effects (SVD/Eigen)
Inferential approach

- Statistical inference utilizes probability models
- Networks and relational data are represented by matrices and arrays

Social network analysis can utilize probability models of matrices and arrays.

\[ Y \sim M + E \]

- \( M \) is a low-dimensional matrix of network patterns;
- \( E \) is patternless noise.

We will describe \( M \) via
- regression models
- additive random effects (ANOVA/SRM)
- multiplicative random effects (SVD/Eigen)
Inferential approach

- Statistical inference utilizes probability models
- Networks and relational data are represented by matrices and arrays

Social network analysis can utilize probability models of matrices and arrays.

\[ Y \sim M + E \]

- \( M \) is a low-dimensional matrix of network patterns;
- \( E \) is patternless noise.

We will describe \( M \) via
- regression models
- additive random effects (ANOVA/SRM)
- multiplicative random effects (SVD/Eigen)
Inferential approach

- Statistical inference utilizes probability models
- Networks and relational data are represented by matrices and arrays

Social network analysis can utilize probability models of matrices and arrays.

\[ Y \sim M + E \]

- \( M \) is a low-dimensional matrix of network patterns;
- \( E \) is patternless noise.

We will describe \( M \) via
- regression models
- additive random effects (ANOVA/SRM)
- multiplicative random effects (SVD/Eigen)
Multiplicative effects - higher order dependence

Inferential goals in the regression framework

\[ y_{i,j} \text{ measures } i \rightarrow j, \quad x_{i,j} \text{ is a vector of explanatory variables.} \]

\[
Y = \begin{pmatrix}
y_{1,1} & y_{1,2} & y_{1,3} & \text{NA} & y_{1,5} & \cdots \\
y_{2,1} & y_{2,2} & y_{2,3} & y_{2,4} & y_{2,5} & \cdots \\
y_{3,1} & \text{NA} & y_{3,3} & y_{3,4} & \text{NA} & \cdots \\
y_{4,1} & y_{4,2} & y_{4,3} & y_{4,4} & y_{4,5} & \cdots \\
\vdots & \vdots & \vdots & \vdots & \vdots & \vdots
\end{pmatrix}
\]

\[
X = \begin{pmatrix}
x_{1,1} & x_{1,2} & x_{1,3} & x_{1,4} & x_{1,5} & \cdots \\
x_{2,1} & x_{2,2} & x_{2,3} & x_{2,4} & x_{2,5} & \cdots \\
x_{3,1} & x_{3,2} & x_{3,3} & x_{3,4} & x_{3,5} & \cdots \\
x_{4,1} & x_{4,2} & x_{4,3} & x_{4,4} & x_{4,5} & \cdots \\
\vdots & \vdots & \vdots & \vdots & \vdots & \vdots
\end{pmatrix}
\]

Consider a basic (generalized) linear model

\[ y_{i,j} \sim \beta^T x_{i,j} + e_{i,j} \]

A model can provide
\begin{itemize}
  \item a measure of the association between \( X \) and \( Y \): \( \hat{\beta}, \text{se}(\hat{\beta}) \)
  \item imputations of missing observations: \( p(y_{1,4}|Y, X) \)
  \item a probabilistic description of network features: \( g(\tilde{Y}), \tilde{Y} \sim p(\tilde{Y}|Y, X) \)
\end{itemize}
Inferential goals in the regression framework

\[ y_{i,j} \text{ measures } i \rightarrow j, \quad x_{i,j} \text{ is a vector of explanatory variables.} \]

\[
Y = \begin{pmatrix}
y_{1,1} & y_{1,2} & y_{1,3} & \text{NA} & y_{1,5} & \cdots \\
y_{2,1} & y_{2,2} & y_{2,3} & y_{2,4} & y_{2,5} & \cdots \\
y_{3,1} & \text{NA} & y_{3,3} & y_{3,4} & \text{NA} & \cdots \\
y_{4,1} & y_{4,2} & y_{4,3} & y_{4,4} & y_{4,5} & \cdots \\
\vdots & \vdots & \vdots & \vdots & \vdots & \ddots
\end{pmatrix}
\]

\[
X = \begin{pmatrix}
x_{1,1} & x_{1,2} & x_{1,3} & x_{1,4} & x_{1,5} & \cdots \\
x_{2,1} & x_{2,2} & x_{2,3} & x_{2,4} & x_{2,5} & \cdots \\
x_{3,1} & x_{3,2} & x_{3,3} & x_{3,4} & x_{3,5} & \cdots \\
x_{4,1} & x_{4,2} & x_{4,3} & x_{4,4} & x_{4,5} & \cdots \\
\vdots & \vdots & \vdots & \vdots & \vdots & \ddots
\end{pmatrix}
\]

Consider a basic (generalized) linear model

\[ y_{i,j} \sim \beta^T x_{i,j} + e_{i,j} \]

A model can provide

- a measure of the association between \( X \) and \( Y \): \( \hat{\beta}, \text{se}(\hat{\beta}) \)
- imputations of missing observations: \( p(y_{1,4}|Y, X) \)
- a probabilistic description of network features: \( g(\tilde{Y}), \tilde{Y} \sim p(\tilde{Y}|Y, X) \)
Multiplicative effects - higher order dependence

Inferential goals in the regression framework

\[ y_{i,j} \text{ measures } i \to j, \quad x_{i,j} \text{ is a vector of explanatory variables.} \]

\[
Y = \begin{pmatrix}
y_{1,1} & y_{1,2} & y_{1,3} & \text{NA} & y_{1,5} & \cdots \\
y_{2,1} & y_{2,2} & y_{2,3} & y_{2,4} & y_{2,5} & \cdots \\
y_{3,1} & \text{NA} & y_{3,3} & y_{3,4} & \text{NA} & \cdots \\
y_{4,1} & y_{4,2} & y_{4,3} & y_{4,4} & y_{4,5} & \cdots \\
\vdots & \vdots & \vdots & \vdots & \vdots & \ddots
\end{pmatrix}
\]

\[
X = \begin{pmatrix}
x_{1,1} & x_{1,2} & x_{1,3} & x_{1,4} & x_{1,5} & \cdots \\
x_{2,1} & x_{2,2} & x_{2,3} & x_{2,4} & x_{2,5} & \cdots \\
x_{3,1} & x_{3,2} & x_{3,3} & x_{3,4} & x_{3,5} & \cdots \\
x_{4,1} & x_{4,2} & x_{4,3} & x_{4,4} & x_{4,5} & \cdots \\
\vdots & \vdots & \vdots & \vdots & \vdots & \ddots
\end{pmatrix}
\]

Consider a basic (generalized) linear model

\[
y_{i,j} \sim \beta^T x_{i,j} + e_{i,j}
\]

A model can provide

- a measure of the association between \(X\) and \(Y\): \(\hat{\beta}, \text{se}(\hat{\beta})\)
- imputations of missing observations: \(p(y_{1,4}|Y, X)\)
- a probabilistic description of network features: \(g(\tilde{Y}), \tilde{Y} \sim p(\tilde{Y}|Y, X)\)
Inferential goals in the regression framework

\[ y_{i,j} \text{ measures } i \to j, \quad x_{i,j} \text{ is a vector of explanatory variables.} \]

\[
Y = \begin{pmatrix}
y_{1,1} & y_{1,2} & y_{1,3} & \text{NA} & y_{1,5} & \cdots \\
y_{2,1} & y_{2,2} & y_{2,3} & y_{2,4} & y_{2,5} & \cdots \\
y_{3,1} & \text{NA} & y_{3,3} & y_{3,4} & \text{NA} & \cdots \\
y_{4,1} & y_{4,2} & y_{4,3} & y_{4,4} & y_{4,5} & \cdots \\
\vdots & \vdots & \vdots & \vdots & \vdots & \ddots
\end{pmatrix}, \quad
X = \begin{pmatrix}
x_{1,1} & x_{1,2} & x_{1,3} & x_{1,4} & x_{1,5} & \cdots \\
x_{2,1} & x_{2,2} & x_{2,3} & x_{2,4} & x_{2,5} & \cdots \\
x_{3,1} & x_{3,2} & x_{3,3} & x_{3,4} & x_{3,5} & \cdots \\
x_{4,1} & x_{4,2} & x_{4,3} & x_{4,4} & x_{4,5} & \cdots \\
\vdots & \vdots & \vdots & \vdots & \vdots & \ddots
\end{pmatrix}
\]

Consider a basic (generalized) linear model

\[ y_{i,j} \sim \beta^T x_{i,j} + e_{i,j} \]

A model can provide

- a measure of the association between \(X\) and \(Y\): \(\hat{\beta}, se(\hat{\beta})\)
- imputations of missing observations: \(p(y_{1,4}|Y, X)\)
- a probabilistic description of network features: \(g(\tilde{Y}), \tilde{Y} \sim p(\tilde{Y}|Y, X)\)
Network dependence

GLM: \( y_{i,j} \sim \beta^T x_{i,j} + e_{i,j} \)

Networks typically show evidence AGAINST independence of \( \{e_{i,j} : i \neq j\} \).

Not accounting for dependence can lead to

- biased effect estimation;
- uncalibrated confidence intervals;
- poor predictive performance;
- inaccurate description of network phenomenon.

We will test for and model various forms of network dependence.
Network dependence

GLM: \( y_{i,j} \sim \beta^T x_{i,j} + e_{i,j} \)

Networks typically shows evidence AGAINST independence of \( \{e_{i,j} : i \neq j\} \).

Not accounting for dependence can lead to

- biased effect estimation;
- uncalibrated confidence intervals;
- poor predictive performance;
- inaccurate description of network phenomenon.

We will test for and model various forms of network dependence.
Network dependence

GLM: \( y_{i,j} \sim \beta^T x_{i,j} + e_{i,j} \)

Networks typically show evidence AGAINST independence of \( \{e_{i,j} : i \neq j\} \).

Not accounting for dependence can lead to

- biased effect estimation;
- uncalibrated confidence intervals;
- poor predictive performance;
- inaccurate description of network phenomenon.

We will test for and model various forms of network dependence.
Network dependence

GLM: \( y_{i,j} \sim \beta^T x_{i,j} + e_{i,j} \)

Networks typically shows evidence AGAINST independence of \( \{e_{i,j} : i \neq j\} \).

Not accounting for dependence can lead to

- biased effect estimation;
- uncalibrated confidence intervals;
- poor predictive performance;
- inaccurate description of network phenomenon.

We will test for and model various forms of network dependence.
Network dependence

**GLM:** \( y_{i,j} \sim \beta^T x_{i,j} + e_{i,j} \)

Networks typically shows evidence AGAINST independence of \( \{ e_{i,j} : i \neq j \} \).

Not accounting for dependence can lead to

- biased effect estimation;
- uncalibrated confidence intervals;
- poor predictive performance;
- inaccurate description of network phenomenon.

We will test for and model various forms of network dependence.
Network dependence

**GLM:** \( y_{i,j} \sim \beta^T x_{i,j} + e_{i,j} \)

Networks typically show evidence AGAINST independence of \( \{ e_{i,j} : i \neq j \} \).

Not accounting for dependence can lead to
- biased effect estimation;
- uncalibrated confidence intervals;
- poor predictive performance;
- inaccurate description of network phenomenon.

We will test for and model various forms of network dependence.
High School friendship networks

<table>
<thead>
<tr>
<th></th>
<th>$n$</th>
<th>$E[\bar{y}_i]$</th>
<th>$sd[\bar{y}_i]$</th>
<th>$sd[\bar{y}_{ij}]$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>145</td>
<td>3.59</td>
<td>2.95</td>
<td>5.15</td>
</tr>
</tbody>
</table>
Regression modelling

Covariate information: grade, gpa, smoking

- node level regressors: gpa, smoking
- dyad level regressors: same grade, gpa interaction, smoking interaction

```r
glm(c(Y)~ -1+ apply(X,3,c),family=binomial(link=probit))
```

Coefficients:

|                      | Estimate | Std. Error | z value | Pr(>|z|) |
|----------------------|----------|------------|---------|----------|
| intercept            | -2.1832  | 0.0286     | -76.22  | < 2e-16  ***
| rgpa                 | -0.0062  | 0.0227     | -0.28   | 0.783    |
| rsmoke               | -0.1463  | 0.0256     | -5.71   | 1.15e-08 ***
| cgpa                 | 0.094    | 0.0215     | 4.37    | 1.27e-05 ***
| csmoke               | -0.1065  | 0.0256     | -4.15   | 3.33e-05 ***
| igrade               | 0.422    | 0.0393     | 10.74   | < 2e-16  ***
| igpa                 | 0.0697   | 0.0188     | 3.70    | 0.0002   ***
| ismoke               | 0.0129   | 0.0303     | 0.42    | 0.671    |

The p-values say, for example, that a model with preferential association for same grade is a better description of the data than a model without.
Regression modelling

Covariate information: grade, gpa, smoking

- node level regressors: gpa, smoking
- dyad level regressors: same grade, gpa interaction, smoking interaction

```
glm(c(Y) ~ -1 + apply(X,3,c), family=binomial(link=probit))
```

Coefficients:

|            | Estimate | Std. Error | z value | Pr(>|z|)  |
|------------|----------|------------|---------|-----------|
| intercept  | -2.183209| 0.028641   | -76.226 | < 2e-16   |
| rgpa       | -0.006245| 0.022727   | -0.275  | 0.783475  |
| rsmoke     | -0.146312| 0.025635   | -5.707  | 1.15e-08  |
| cgpa       | 0.094009 | 0.021533   | 4.366   | 1.27e-05  |
| csmoke     | -0.106515| 0.025669   | -4.150  | 3.33e-05  |
| igrade     | 0.422207 | 0.039331   | 10.735  | < 2e-16   |
| igpa       | 0.069667 | 0.018814   | 3.703   | 0.000213  |
| ismoke     | 0.012877 | 0.030351   | 0.424   | 0.671365  |

The p-values say, for example, that a model with preferential association for same grade is a better description of the data than a model without.
Regression modelling

Covariate information: grade, gpa, smoking

- node level regressors: gpa, smoking
- dyad level regressors: same grade, gpa interaction, smoking interaction

glm(c(Y) ~ -1 + apply(X,3,c), family=binomial(link=probit))

Coefficients:

|            | Estimate | Std. Error | z value | Pr(>|z|) |
|------------|----------|------------|---------|----------|
| intercept  | -2.183209| 0.028641   | -76.226 | < 2e-16  *** |
| rgpa       | -0.006245| 0.022727   | -0.275  | 0.783475 |
| rsmoke     | -0.146312| 0.025635   | -5.707  | 1.15e-08 *** |
| cgpa       | 0.094009 | 0.021533   | 4.366   | 1.27e-05 *** |
| csmoke     | -0.106515| 0.025669   | -4.150  | 3.33e-05 *** |
| igrade     | 0.422207 | 0.039331   | 10.735  | < 2e-16  *** |
| igpa       | 0.069667 | 0.018814   | 3.703   | 0.000213 *** |
| ismoke     | 0.012877 | 0.030351   | 0.424   | 0.671365 |

The p-values say, for example, that a model with preferential association for same grade is a better description of the data than a model without.
Model comparison

\{y_{i,j} : i \neq j\} \text{ independent with } \Phi^{-1}(\Pr(y_{i,j} = 1)) = \beta^T x_{i,j}

Model 0: \(x_{i,j}\) lacks same grade indicator
Model 1: \(x_{i,j}\) includes same grade indicator

Which model provides a better representation of the data?

**GOF statistic:**

\[
t(\mathbf{Y}) = \log \frac{\text{odds}(y_{i,j} = 1|x_{g,i} = x_{g,j})}{\text{odds}(y_{i,j} = 1|x_{g,i} \neq x_{g,j})}
\]

\[
t(\mathbf{Y}_{\text{obs}}) = 0.962
\]

What do Model 0 and Model 1 predict in terms of \(t(\mathbf{Y})\)?
Model comparison

\{y_{i,j} : i \neq j\} independent with \Phi^{-1}(\Pr(y_{i,j} = 1)) = \beta^T x_{i,j}

**Model 0:** $x_{i,j}$ lacks same grade indicator

**Model 1:** $x_{i,j}$ includes same grade indicator

Which model provides a better representation of the data?

**GOF statistic:**

\[
t(\mathcal{Y}) = \log \frac{\text{odds}(y_{i,j} = 1|x_{g,i} = x_{g,j})}{\text{odds}(y_{i,j} = 1|x_{g,i} \neq x_{g,j})}
\]

\[
t(\mathcal{Y}_{obs}) = 0.962
\]

What do Model 0 and Model 1 predict in terms of $t(\mathcal{Y})$?
Multiplicative effects - higher order dependence

Model comparison

\{y_{i,j} : i \neq j\} independent with \( \Phi^{-1}(\Pr(y_{i,j} = 1)) = \beta^T x_{i,j} \)

**Model 0:** \( x_{i,j} \) lacks same grade indicator  
**Model 1:** \( x_{i,j} \) includes same grade indicator

Which model provides a better representation of the data?

**GOF statistic:**

\[
\begin{align*}
    t(\mathbf{Y}) &= \log \frac{\text{odds}(y_{i,j} = 1 | x_{g,i} = x_{g,j})}{\text{odds}(y_{i,j} = 1 | x_{g,i} \neq x_{g,j})} \\
    t(\mathbf{Y}_{\text{obs}}) &= 0.962
\end{align*}
\]

What do Model 0 and Model 1 predict in terms of \( t(\mathbf{Y}) \)?
Model comparison

\{y_{i,j} : i \neq j\} \text{ independent with } \Phi^{-1}(\Pr(y_{i,j} = 1)) = \beta^T x_{i,j}

**Model 0:**  $x_{i,j}$ lacks same grade indicator

**Model 1:**  $x_{i,j}$ includes same grade indicator

Which model provides a better representation of the data?

**GOF statistic:**

\[
t(\mathbf{Y}) = \log \frac{\text{odds}(y_{i,j} = 1 | x_{g,i} = x_{g,j})}{\text{odds}(y_{i,j} = 1 | x_{g,i} \neq x_{g,j})}
\]

\[
t(\mathbf{Y}_{\text{obs}}) = 0.962
\]

What do Model 0 and Model 1 predict in terms of $t(\mathbf{Y})$?


Model comparison

\[ \{y_{i,j} : i \neq j\} \text{ independent with } \Phi^{-1}(\Pr(y_{i,j} = 1)) = \beta^T x_{i,j} \]

### Model 0:
- \( x_{i,j} \) lacks same grade indicator

### Model 1:
- \( x_{i,j} \) includes same grade indicator

Which model provides a better representation of the data?

**GOF statistic:**

\[
 t(Y) = \log \frac{\text{odds}(y_{i,j} = 1|x_{g,i} = x_{g,j})}{\text{odds}(y_{i,j} = 1|x_{g,i} \neq x_{g,j})}
\]

\[
 t(Y_{\text{obs}}) = 0.962
\]

What do Model 0 and Model 1 predict in terms of \( t(Y) \)?
Model comparison

1. Fit model $Y_{obs} \sim p(Y|X, \beta)$ and obtain estimate $\hat{\beta}$
2. Simulate $Y_1, \ldots, Y_S \sim p(Y|X, \hat{\beta})$ using $\hat{\beta}$.
3. Compare $t(Y_{obs})$ to $\{t(Y_1), \ldots, t(Y_S)\}$.

Related to: parametric bootstrap, posterior predictive checks.
Model comparison

1. Fit model $Y_{obs} \sim p(Y|X, \beta)$ and obtain estimate $\hat{\beta}$
2. Simulate $Y_1, \ldots, Y_S \sim p(Y|X, \hat{\beta})$ using $\hat{\beta}$.
3. Compare $t(Y_{obs})$ to $\{t(Y_1), \ldots, t(Y_S)\}$.

Related to: parametric bootstrap, posterior predictive checks.
Model comparison

1. Fit model $Y_{obs} \sim p(Y|X, \beta)$ and obtain estimate $\hat{\beta}$
2. Simulate $Y_1, \ldots, Y_S \sim p(Y|X, \hat{\beta})$ using $\hat{\beta}$.
3. Compare $t(Y_{obs})$ to $\{t(Y_1), \ldots, t(Y_S)\}$.

Related to: parametric bootstrap, posterior predictive checks.
Model comparison

1. Fit model $Y_{obs} \sim p(Y|X, \beta)$ and obtain estimate $\hat{\beta}$
2. Simulate $Y_1, \ldots, Y_S \sim p(Y|X, \hat{\beta})$ using $\hat{\beta}$.
3. Compare $t(Y_{obs})$ to $\{t(Y_1), \ldots, t(Y_S)\}$.

Related to: parametric bootstrap, posterior predictive checks.
Other model deficiencies
Reciprocity

\[ t(Y) = \frac{\sum_{i \neq j} y_{i,j} y_{j,i}}{\sum_{i \neq j} y_{i,j}} \]

- \( t(Y) \) is the fraction of ties that are reciprocated;
- \( t(Y) = 1 \) if \( Y \) is symmetric;
- For our data, \( t(Y_{\text{obs}}) = 0.362 \).

Questions:
- Is 0.362 large?
- If so, can it be explained by the regressors?
- If not, what are the implications?
Reciprocity

\[ t(\mathbf{Y}) = \frac{\sum_{i \neq j} y_{i,j} y_{j,i}}{\sum_{i \neq j} y_{i,j}} \]

- \( t(\mathbf{Y}) \) is the fraction of ties that are reciprocated;
- \( t(\mathbf{Y}) = 1 \) if \( \mathbf{Y} \) is symmetric;
- For our data, \( t(\mathbf{Y}_{obs}) = 0.362 \).

Questions:

- Is 0.362 large?
- If so, can it be explained by the regressors?
- If not, what are the implications?
Reciprocity

\[ t(\mathbf{Y}) = \frac{\sum_{i \neq j} y_{i,j} y_{j,i}}{\sum_{i \neq j} y_{i,j}} \]

- \( t(\mathbf{Y}) \) is the fraction of ties that are reciprocated;
- \( t(\mathbf{Y}) = 1 \) if \( \mathbf{Y} \) is symmetric;
- For our data, \( t(\mathbf{Y}_{obs}) = 0.362 \).

Questions:

- Is 0.362 large?
- If so, can it be explained by the regressors?
- If not, what are the implications?
Reciprocity

\[ t(Y) = \frac{\sum_{i \neq j} y_{i,j}y_{j,i}}{\sum_{i \neq j} y_{i,j}} \]

- \( t(Y) \) is the fraction of ties that are reciprocated;
- \( t(Y) = 1 \) if \( Y \) is symmetric;
- For our data, \( t(Y_{\text{obs}}) = 0.362 \).

Questions:
- Is 0.362 large?
  - If so, can it be explained by the regressors?
  - If not, what are the implications?
Reciprocity

\[ t(Y) = \frac{\sum_{i \neq j} y_{i,j} y_{j,i}}{\sum_{i \neq j} y_{i,j}} \]

- \( t(Y) \) is the fraction of ties that are reciprocated;
- \( t(Y) = 1 \) if \( Y \) is symmetric;
- For our data, \( t(Y_{\text{obs}}) = 0.362 \).

Questions:

- Is 0.362 large?
- If so, can it be explained by the regressors?
- If not, what are the implications?
Reciprocity

\[ t(\mathbf{Y}) = \frac{\sum_{i \neq j} y_{i,j} y_{j,i}}{\sum_{i \neq j} y_{i,j}} \]

- \( t(\mathbf{Y}) \) is the fraction of ties that are reciprocated;
- \( t(\mathbf{Y}) = 1 \) if \( \mathbf{Y} \) is symmetric;
- For our data, \( t(\mathbf{Y}_{\text{obs}}) = 0.362 \).

**Questions:**
- Is 0.362 large?
- If so, can it be explained by the regressors?
- If not, what are the implications?
Reciprocity

\[ t(Y) = \frac{\sum_{i \neq j} y_{i,j}y_{j,i}}{\sum_{i \neq j} y_{i,j}} \]

- \( t(Y) \) is the fraction of ties that are reciprocated;
- \( t(Y) = 1 \) if \( Y \) is symmetric;
- For our data, \( t(Y_{obs}) = 0.362 \).

Questions:
- Is 0.362 large?
- If so, can it be explained by the regressors?
- If not, what are the implications?
The model fails in terms of this gof statistic.

- It appears that \( \{y_{i,j}, y_{j,i}\} \) are correlated.
- This suggests the independence assumption is violated.
Other model deficiencies
The probit regression model also says that covariate effects adequately describe
• “friendliness” (heterogeneity in outdegree)
• “popularity” (heterogeneity in indegree)

Let’s evaluate this assumption with the following test statistics:

\[ t_{od}(Y) = \{ \# \{ i : \tilde{n_i} = d \}, d = 0, \ldots, n \} \]
\[ t_{id}(Y) = \{ \# \{ j : \tilde{n_j} = d \}, d = 0, \ldots, n \} \]

These functions are known as the in and out degree distributions.
The probit regression model also says that covariate effects adequately describe

- “friendliness” (heterogeneity in outdegree)
- “popularity” (heterogeneity in indegree)

Let’s evaluate this assumption with the following test statistics:

\[ t_{od}(Y) = \left\{ \# \{ i : n\tilde{y}_{i.} = d \}, d = 0, \ldots, n \right\} \]

\[ t_{id}(Y) = \left\{ \# \{ j : n\tilde{y}_{.j} = d \}, d = 0, \ldots, n \right\} \]

These functions are known as the in and out degree distributions.
The probit regression model also says that covariate effects adequately describe
- “friendliness” (heterogeneity in outdegree)
- “popularity” (heterogeneity in indegree)

Let’s evaluate this assumption with the following test statistics:

\[
t_{od}(Y) = \{ \#\{ i : \tilde{n}_{yi} = d \}, d = 0, \ldots, n \} \\
t_{id}(Y) = \{ \#\{ j : \tilde{n}_{yj} = d \}, d = 0, \ldots, n \}
\]

These functions are known as the in and out degree distributions.
Degree distributions

The probit regression model also says that covariate effects adequately describe

- “friendliness” (heterogeneity in outdegree)
- “popularity” (heterogeneity in indegree)

Let’s evaluate this assumption with the following test statistics:

\[
\begin{align*}
t_{od}(Y) &= \# \{i : n\hat{y}_{i.} = d\}, \quad d = 0, \ldots, n \\
t_{id}(Y) &= \# \{j : n\hat{y}_{.j} = d\}, \quad d = 0, \ldots, n
\end{align*}
\]

These functions are known as the in and out degree distributions.
Empirical degree distributions

![Empirical degree distributions graph]

- Multiplicative effects - higher order dependence
The model fails here, particularly in terms of indegree.

This suggests within-column differences in popularity.

Statistically, this can be interpreted in terms of within-column dependence.
A primer on exchangeability and de Finetti’s theorem

Let $Y_1, \ldots, Y_n$ be an exchangeable sequence for all $n$:

$$
\Pr(Y_1 = y_1, \ldots, Y_n = y_n) = \Pr(Y_1 = y_{\pi_1}, \ldots, Y_n = y_{\pi_n}) \forall n
$$

de Finetti’s theorem says

$$
Y_i = g(\mu, \epsilon_i), \text{ where } \epsilon_1, \ldots, \epsilon_n \overset{iid}{\sim} p_{\epsilon}
$$

- The parameter $\mu$ represents “global features” of the sequence.
- The $\epsilon_i$’s represent “local features”, specific to individual $Y_i$’s.

(This theorem justifies the ubiquitous “conditionally i.i.d.” assumption of statistical modeling)
A primer on exchangeability and de Finetti’s theorem

Let $Y_1, \ldots, Y_n$ be an exchangeable sequence for all $n$:

$$\Pr(Y_1 = y_1, \ldots, Y_n = y_n) = \Pr(Y_1 = y_{\pi_1}, \ldots, Y_n = y_{\pi_n}) \ \forall n$$

de Finetti’s theorem says

$$Y_i = g(\mu, \epsilon_i), \text{ where}$$

$$\epsilon_1, \ldots, \epsilon_n \overset{iid}{\sim} p_{\epsilon}$$

- The parameter $\mu$ represents “global features” of the sequence.
- The $\epsilon_i$’s represent “local features”, specific to individual $Y_i$’s.

(This theorem justifies the ubiquitous “conditionally i.i.d.” assumption of statistical modeling)
A primer on exchangeability and de Finetti’s theorem

Let $Y_1, \ldots, Y_n$ be an exchangeable sequence for all $n$:

$$\Pr(Y_1 = y_1, \ldots, Y_n = y_n) = \Pr(Y_1 = y_{\pi_1}, \ldots, Y_n = y_{\pi_n}) \forall n$$

de Finetti’s theorem says

$$Y_i = g(\mu, \epsilon_i), \text{ where}$$

$$\epsilon_1, \ldots, \epsilon_n \overset{iid}{\sim} p_\epsilon$$

- The parameter $\mu$ represents “global features” of the sequence.
- The $\epsilon_i$’s represent “local features”, specific to individual $Y_i$’s.

(This theorem justifies the ubiquitous “conditionally i.i.d.” assumption of statistical modeling)
A primer on exchangeability and de Finetti’s theorem

Let $Y_1, \ldots, Y_n$ be an exchangeable sequence for all $n$:

$$\Pr(Y_1 = y_1, \ldots, Y_n = y_n) = \Pr(Y_1 = y_{\pi_1}, \ldots, Y_n = y_{\pi_n}) \quad \forall n$$

de Finetti’s theorem says

$$Y_i = g(\mu, \epsilon_i) \quad \text{where}$$

$$\epsilon_1, \ldots, \epsilon_n \overset{iid}{\sim} p_{\epsilon}$$

- The parameter $\mu$ represents “global features” of the sequence.
- The $\epsilon_i$’s represent “local features”, specific to individual $Y_i$’s.

(This theorem justifies the ubiquitous “conditionally i.i.d.” assumption of statistical modeling)
A primer on exchangeability and de Finetti’s theorem

Let $Y_1, \ldots, Y_n$ be an exchangeable sequence for all $n$:

$$
\Pr(Y_1 = y_1, \ldots, Y_n = y_n) = \Pr(Y_1 = y_{\pi_1}, \ldots, Y_n = y_{\pi_n}) \quad \forall n
$$

De Finetti’s theorem says

$$
Y_i = g(\mu, \epsilon_i), \text{ where } \epsilon_1, \ldots, \epsilon_n \overset{iid}{\sim} p_{\epsilon}
$$

- The parameter $\mu$ represents “global features” of the sequence.
- The $\epsilon_i$’s represent “local features”, specific to individual $Y_i$’s.

(This theorem justifies the ubiquitous “conditionally i.i.d.” assumption of statistical modeling)
Exchangeability for nested data

Now consider an $m \times n$ data matrix:

$$
Y = \begin{pmatrix}
Y_{1,1} & Y_{1,2} & \cdots & Y_{1,n} \\
Y_{2,1} & Y_{2,2} & \cdots & Y_{2,n} \\
\vdots & \vdots & \ddots & \vdots \\
Y_{m,1} & Y_{m,2} & \cdots & Y_{m,n}
\end{pmatrix}
$$

Suppose $\Pr(Y)$ is exchangeable across rows and within rows:

- $\Pr(Y_1 = y_1, \ldots, Y_m = y_m) = \Pr(Y_1 = y_{\pi_1}, \ldots, Y_m = y_{\pi_m})$
- $\Pr(Y_{i,1} = y_{i,1}, \ldots, Y_{i,n} = y_{i,n}) = \Pr(Y_{i,1} = y_{i,\omega_1}, \ldots, Y_{i,n} = y_{i,\omega_n})$

A double application of de Finetti’s theorem implies

$$
Y_{i,j} = g(\mu, a_i, \epsilon_{i,j})
$$

$$
a_1, \ldots, a_n \overset{iid}{\sim} p_a
$$

$$
\{\epsilon_{i,j}\} \overset{iid}{\sim} p_\epsilon
$$

- The parameter $\mu$ represents global features of the data.
- Heterogeneity in the $a_i$’s represents across-row heterogeneity.
- Heterogeneity in the $\epsilon_{i,j}$’s represents within-row heterogeneity.
Exchangeability for nested data

Now consider an $m \times n$ data matrix:

$$
Y = \begin{pmatrix}
Y_{1,1} & Y_{1,2} & \cdots & Y_{1,n} \\
Y_{2,1} & Y_{2,2} & \cdots & Y_{2,n} \\
\vdots & \vdots & \ddots & \vdots \\
Y_{m,1} & Y_{m,2} & \cdots & Y_{m,n}
\end{pmatrix}
$$

Suppose $\Pr(Y)$ is exchangeable across rows and within rows:

$$
\Pr(Y_1 = y_1, \ldots, Y_m = y_m) = \Pr(Y_1 = y_{\pi_1}, \ldots, Y_m = y_{\pi_m})
$$

$$
\Pr(Y_{i,1} = y_{i,1}, \ldots, Y_{i,n} = y_{i,n}) = \Pr(Y_{i,1} = y_{i,\omega_1}, \ldots, Y_{i,n} = y_{i,\omega_n})
$$

A double application of de Finetti’s theorem implies

$$
Y_{i,j} = g(\mu, a_i, \epsilon_{i,j})
$$

$$
a_1, \ldots, a_n \overset{iid}{\sim} p_a
$$

$$
\{\epsilon_{i,j}\} \overset{iid}{\sim} p_{\epsilon}
$$

- The parameter $\mu$ represents **global features** of the data.
- Heterogeneity in the $a_i$’s represents **across-row** heterogeneity.
- Heterogeneity in the $\epsilon_{i,j}$’s represents **within-row** heterogeneity.
Exchangeability for nested data

Now consider an \( m \times n \) data matrix:

\[
Y = \begin{pmatrix}
Y_{1,1} & Y_{1,2} & \cdots & Y_{1,n} \\
Y_{2,1} & Y_{2,2} & \cdots & Y_{2,n} \\
\vdots & \vdots & \ddots & \vdots \\
Y_{m,1} & Y_{m,2} & \cdots & Y_{m,n}
\end{pmatrix}
\]

Suppose \( \Pr(Y) \) is exchangeable across rows and within rows:

\[
\Pr(Y_1 = y_1, \ldots, Y_m = y_m) = \Pr(Y_1 = y_{\pi_1}, \ldots, Y_m = y_{\pi_m})
\]

\[
\Pr(Y_{i,1} = y_{i,1}, \ldots, Y_{i,n} = y_{i,n}) = \Pr(Y_{i,1} = y_{i,\omega_1}, \ldots, Y_{i,n} = y_{i,\omega_n})
\]

A double application of de Finetti’s theorem implies

\[
Y_{i,j} = g(\mu, a_i, \epsilon_{i,j})
\]

\[
a_1, \ldots, a_n \overset{iid}{\sim} p_a
\]

\[
\{\epsilon_{i,j}\} \overset{iid}{\sim} p_\epsilon
\]

- The parameter \( \mu \) represents global features of the data.
- Heterogeneity in the \( a_i \)'s represents across-row heterogeneity.
- Heterogeneity in the \( \epsilon_{i,j} \)'s represents within-row heterogeneity.
Exchangeability for nested data

Now consider an \( m \times n \) data matrix:

\[
\mathbf{Y} = \begin{pmatrix}
Y_{1,1} & Y_{1,2} & \cdots & Y_{1,n} \\
Y_{2,1} & Y_{2,2} & \cdots & Y_{2,n} \\
\vdots & \vdots & \ddots & \vdots \\
Y_{m,1} & Y_{m,2} & \cdots & Y_{m,n}
\end{pmatrix}
\]

Suppose \( \Pr(\mathbf{Y}) \) is exchangeable across rows and within rows:

\[
\Pr(\mathbf{Y}_1 = \mathbf{y}_1, \ldots, \mathbf{Y}_m = \mathbf{y}_m) = \Pr(\mathbf{Y}_1 = \mathbf{y}_{\pi_1}, \ldots, \mathbf{Y}_m = \mathbf{y}_{\pi_m})
\]

\[
\Pr(Y_{i,1} = y_{i,1}, \ldots, Y_{i,n} = y_{i,n}) = \Pr(Y_{i,1} = y_{i,\omega_1}, \ldots, Y_{i,n} = y_{i,\omega_n})
\]

A double application of de Finetti’s theorem implies

\[
Y_{i,j} = g(\mu, a_i, \epsilon_{i,j})
\]

\[
a_1, \ldots, a_n \stackrel{iid}{\sim} p_a
\]

\[
\{\epsilon_{i,j}\} \stackrel{iid}{\sim} p_\epsilon
\]

- The parameter \( \mu \) represents global features of the data.
- Heterogeneity in the \( a_i \)’s represents across-row heterogeneity.
- Heterogeneity in the \( \epsilon_{i,j} \)’s represents within-row heterogeneity.
Exchangeability for nested data

Now consider an $m \times n$ data matrix:

$$\mathbf{Y} = \begin{pmatrix}
Y_{1,1} & Y_{1,2} & \cdots & Y_{1,n} \\
Y_{2,1} & Y_{2,2} & \cdots & Y_{2,n} \\
\vdots & \vdots & \ddots & \vdots \\
Y_{m,1} & Y_{m,2} & \cdots & Y_{m,n}
\end{pmatrix}$$

Suppose $\Pr(\mathbf{Y})$ is exchangeable across rows and within rows:

$$\Pr(\mathbf{Y}_1 = \mathbf{y}_1, \ldots, \mathbf{Y}_m = \mathbf{y}_m) = \Pr(\mathbf{Y}_1 = \mathbf{y}_{\pi_1}, \ldots, \mathbf{Y}_m = \mathbf{y}_{\pi_m})$$
$$\Pr(Y_{i,1} = y_{i,1}, \ldots, Y_{i,n} = y_{i,n}) = \Pr(Y_{i,1} = y_{i,\omega_1}, \ldots, Y_{i,n} = y_{i,\omega_n})$$

A double application of de Finetti’s theorem implies

$$Y_{i,j} = g(\mu, a_i, \epsilon_{i,j})$$
$$a_1, \ldots, a_n \overset{iid}{\sim} p_a$$
$$\{\epsilon_{i,j}\} \overset{iid}{\sim} p_\epsilon$$

- The parameter $\mu$ represents global features of the data.
- Heterogeneity in the $a_i$’s represents across-row heterogeneity.
- Heterogeneity in the $\epsilon_{i,j}$’s represents within-row heterogeneity.
Exchangeability for nested data

Now consider an $m \times n$ data matrix:

$$
Y = \begin{pmatrix}
Y_{1,1} & Y_{1,2} & \ldots & Y_{1,n} \\
Y_{2,1} & Y_{2,2} & \ldots & Y_{2,n} \\
\vdots & \vdots & \ddots & \vdots \\
Y_{m,1} & Y_{m,2} & \ldots & Y_{m,n}
\end{pmatrix}
$$

Suppose $Pr(Y)$ is exchangeable across rows and within rows:

$$
Pr(Y_1 = y_1, \ldots, Y_m = y_m) = Pr(Y_1 = y_{\pi_1}, \ldots, Y_m = y_{\pi_m})
$$

$$
Pr(Y_{i,1} = y_{i,1}, \ldots, Y_{i,n} = y_{i,n}) = Pr(Y_{i,1} = y_{i,\omega_1}, \ldots, Y_{i,n} = y_{i,\omega_n})
$$

A double application of de Finetti's theorem implies

$$
Y_{i,j} = g(\mu, a_i, \epsilon_{i,j})
$$

$$
a_1, \ldots, a_n \overset{iid}{\sim} p_a
$$

$$
\{\epsilon_{i,j}\} \overset{iid}{\sim} p_\epsilon
$$

- The parameter $\mu$ represents **global features** of the data.
- Heterogeneity in the $a_i$’s represents **across-row** heterogeneity.
- Heterogeneity in the $\epsilon_{i,j}$’s represents **within-row** heterogeneity.
Exchangeability for nested data

Now consider an $m \times n$ data matrix:

$$
Y = \begin{pmatrix}
Y_{1,1} & Y_{1,2} & \cdots & Y_{1,n} \\
Y_{2,1} & Y_{2,2} & \cdots & Y_{2,n} \\
\vdots & \vdots & \ddots & \vdots \\
Y_{m,1} & Y_{m,2} & \cdots & Y_{m,n}
\end{pmatrix}
$$

Suppose $\Pr(Y)$ is exchangeable across rows and within rows:

$$
\Pr(Y_1 = y_1, \ldots, Y_m = y_m) = \Pr(Y_1 = y_{\pi_1}, \ldots, Y_m = y_{\pi_m})
$$
$$
\Pr(Y_{i,1} = y_{i,1}, \ldots, Y_{i,n} = y_{i,n}) = \Pr(Y_{i,1} = y_{i,\omega_1}, \ldots, Y_{i,n} = y_{i,\omega_n})
$$

A double application of de Finetti’s theorem implies

$$
Y_{i,j} = g(\mu, a_i, \epsilon_{i,j})
$$
$$
a_1, \ldots, a_n \overset{iid}{\sim} p_a
$$
$$
\{\epsilon_{i,j}\} \overset{iid}{\sim} p_{\epsilon}
$$

- The parameter $\mu$ represents global features of the data.
- Heterogeneity in the $a_i$’s represents across-row heterogeneity.
- Heterogeneity in the $\epsilon_{i,j}$’s represents within-row heterogeneity.
Exchangeability for relational matrices

Let $\mathbf{Y}$ be a binary matrix with no explanatory variables. What properties should a probability model $\Pr(\mathbf{Y} = \mathbf{y})$ have?

$$
\mathbf{y}_A = \begin{pmatrix}
\cdot & 0 & 1 & 1 \\
0 & \cdot & 0 & 1 \\
1 & 0 & \cdot & 0 \\
1 & 1 & 0 & \cdot \\
\end{pmatrix}
\quad \text{and} \quad
\mathbf{y}_B = \begin{pmatrix}
\cdot & 1 & 0 & 0 \\
1 & \cdot & 1 & 0 \\
0 & 1 & \cdot & 1 \\
0 & 0 & 1 & \cdot \\
\end{pmatrix}
$$

$\mathbf{y}_B$ is just $\mathbf{y}_A$ with the nodes relabeled: $y_{B,i,j} = y_{A,\pi_i,\pi_j}$, $\pi = (3, 1, 4, 2)$

$$
\Pr(\mathbf{Y} = \mathbf{y}_A) \overset{?}{=} \Pr(\mathbf{Y} = \mathbf{y}_B)
$$

\textbf{RCE model:} $\Pr(\cdot)$ is RCE if $\Pr(\mathbf{Y} = \mathbf{y}) = \Pr(\mathbf{Y} = \mathbf{y}_\pi)$ for all $\mathbf{y}$ and $\pi$.

(Hoover 1982, Aldous 1983)
Exchangeability for relational matrices

Let $\mathbf{Y}$ be a binary matrix with no explanatory variables. What properties should a probability model $\Pr(\mathbf{Y} = \mathbf{y})$ have?

$$\mathbf{y}_A = \begin{pmatrix} \cdot & 0 & 1 & 1 \\ 0 & \cdot & 0 & 1 \\ 1 & 0 & \cdot & 0 \\ 1 & 1 & 0 & \cdot \end{pmatrix} \quad \mathbf{y}_B = \begin{pmatrix} \cdot & 1 & 0 & 0 \\ 1 & \cdot & 1 & 0 \\ 0 & 1 & \cdot & 1 \\ 0 & 0 & 1 & \cdot \end{pmatrix}$$

$\mathbf{y}_B$ is just $\mathbf{y}_A$ with the nodes relabeled: $\mathbf{y}_{B,i,j} = \mathbf{y}_{A,\pi_i,\pi_j}$, $\pi = (3, 1, 4, 2)$

$$\Pr(\mathbf{Y} = \mathbf{y}_A) \overset{?}{=} \Pr(\mathbf{Y} = \mathbf{y}_B)$$

**RCE model:** $\Pr(\cdot)$ is RCE if $\Pr(\mathbf{Y} = \mathbf{y}) = \Pr(\mathbf{Y} = \mathbf{y}_\pi)$ for all $\mathbf{y}$ and $\pi$.

(Hoover 1982, Aldous 1983)
Exchangeability for relational matrices

Let $\mathbf{Y}$ be a binary matrix with no explanatory variables. What properties should a probability model $\Pr(\mathbf{Y} = \mathbf{y})$ have?

\[
\mathbf{y}_A = \begin{pmatrix}
\cdot & 0 & 1 & 1 \\
0 & \cdot & 0 & 1 \\
1 & 0 & \cdot & 0 \\
1 & 1 & 0 & \cdot 
\end{pmatrix} \\
\mathbf{y}_B = \begin{pmatrix}
\cdot & 1 & 0 & 0 \\
1 & \cdot & 1 & 0 \\
0 & 1 & \cdot & 1 \\
0 & 0 & 1 & \cdot 
\end{pmatrix}
\]

$\mathbf{y}_B$ is just $\mathbf{y}_A$ with the nodes relabeled: $\mathbf{y}_{B,i,j} = \mathbf{y}_{A,\pi_i,\pi_j}$, $\pi = (3, 1, 4, 2)$

\[\Pr(\mathbf{Y} = \mathbf{y}_A) \equiv \Pr(\mathbf{Y} = \mathbf{y}_B)\]

**RCE model:** $\Pr(\cdot)$ is RCE if $\Pr(\mathbf{Y} = \mathbf{y}) = \Pr(\mathbf{Y} = \mathbf{y}_\pi)$ for all $\mathbf{y}$ and $\pi$.

(Hoover 1982, Aldous 1983)
Exchangeability for asymmetric relational matrices

Suppose our model $\Pr()$ for $Y = \{Y_{i,j}, i = 1, \ldots, n, j = 1, \ldots, n\}$ is RCE:

$$\Pr(Y = \{y_{i,j}, i = 1, \ldots, n, j = 1, \ldots, n\}) = \Pr(Y = \{y_{\pi_i, \pi_j}, i = 1, \ldots, n, j = 1, \ldots, n\})$$

Then

$$Y_{i,j} = g(\mu, a_i, b_j, \epsilon_{i,j})$$

$$(a_1, b_1), \ldots, (a_n, b_n) \overset{iid}{\sim} p_{ab}$$

$$\{(\epsilon_{i,j}, \epsilon_{j,i})\} \overset{iid}{\sim} p_{\epsilon}$$

- The parameter $\mu$ represents global features of the matrix.
- The $a_i$’s represent nodal sender features.
- The $b_j$’s represent nodal receiver features.
- The $(\epsilon_{i,j}, \epsilon_{j,i})$’s represent heterogeneity among ordered dyads.
Multiplicative effects - higher order dependence

Exchangeability for asymmetric relational matrices

Suppose our model $Pr()$ for $Y = \{ Y_{i,j}, i = 1, \ldots, n, j = 1, \ldots, n \}$ is RCE:

$$Pr(Y = \{ y_{i,j}, i = 1, \ldots, n, j = 1, \ldots, n \}) = Pr(Y = \{ y_{\pi_i, \pi_j}, i = 1, \ldots, n, j = 1, \ldots, n \})$$

Then

$$Y_{i,j} = g(\mu, a_i, b_j, \epsilon_{i,j})$$

$$(a_1, b_1), \ldots, (a_n, b_n) \overset{iid}{\sim} p_{ab}$$

$$\{(\epsilon_{i,j}, \epsilon_{j,i})\} \overset{iid}{\sim} p_\epsilon$$

- The parameter $\mu$ represents global features of the matrix.
- The $a_i$’s represent nodal sender features.
  - The $b_j$’s represent nodal receiver features.
  - The $(\epsilon_{i,j}, \epsilon_{j,i})$’s represent heterogeneity among ordered dyads.
Exchangeability for asymmetric relational matrices

Suppose our model \( \Pr() \) for \( Y = \{Y_{i,j}, i = 1, \ldots, n, j = 1, \ldots, n\} \) is RCE:

\[
\Pr(Y = \{y_{i,j}, i = 1, \ldots, n, j = 1, \ldots, n\}) = \Pr(Y = \{y_{\pi_{i}, \pi_{j}}, i = 1, \ldots, n, j = 1, \ldots, n\})
\]

Then

\[
Y_{i,j} = g(\mu, a_i, b_j, \epsilon_{i,j})
\]

\[
(a_1, b_1), \ldots, (a_n, b_n) \overset{iid}{\sim} p_{ab}
\]

\[
\{(\epsilon_{i,j}, \epsilon_{j,i})\} \overset{iid}{\sim} p_{\epsilon}
\]

- The parameter \( \mu \) represents global features of the matrix.
- The \( a_i \)'s represent nodal sender features.
- The \( b_j \)'s represent nodal receiver features.
- The \( (\epsilon_{i,j}, \epsilon_{j,i}) \)'s represent heterogeneity among ordered dyads.
Exchangeability for asymmetric relational matrices

Suppose our model $\Pr()$ for $\mathbf{Y} = \{Y_{i,j}, i = 1, \ldots, n, j = 1, \ldots, n\}$ is RCE:

$$\Pr(\mathbf{Y} = \{y_{i,j}, i = 1, \ldots, n, j = 1, \ldots, n\}) = \Pr(\mathbf{Y} = \{y_{\pi_i, \pi_j}, i = 1, \ldots, n, j = 1, \ldots, n\})$$

Then

$$Y_{i,j} = g(\mu, a_i, b_j, \epsilon_{i,j})$$

$$(a_1, b_1), \ldots, (a_n, b_n) \overset{iid}{\sim} p_{ab}$$

$$\{(\epsilon_{i,j}, \epsilon_{j,i})\} \overset{iid}{\sim} p_\epsilon$$

- The parameter $\mu$ represents global features of the matrix.
- The $a_i$'s represent nodal sender features.
- The $b_j$'s represent nodal receiver features.
- The $(\epsilon_{i,j}, \epsilon_{j,i})$'s represent heterogeneity among ordered dyads.
Exchangeability for asymmetric relational matrices

Suppose our model \( \Pr() \) for \( Y = \{ Y_{i,j}, i = 1, \ldots, n, j = 1, \ldots, n \} \) is RCE:

\[
\Pr(Y = \{ y_{i,j}, i = 1, \ldots, n, j = 1, \ldots, n \}) = \Pr(Y = \{ y_{\pi_i,\pi_j}, i = 1, \ldots, n, j = 1, \ldots, n \})
\]

Then

\[
Y_{i,j} = g(\mu, a_i, b_j, \epsilon_{i,j})
\]

\[
(a_1, b_1), \ldots, (a_n, b_n) \sim iid p_{ab}
\]

\[
\{(\epsilon_{i,j}, \epsilon_{j,i})\} \sim iid p_\epsilon
\]

- The parameter \( \mu \) represents global features of the matrix.
- The \( a_i \)'s represent nodal sender features.
- The \( b_j \)'s represent nodal receiver features.
- The \( (\epsilon_{i,j}, \epsilon_{j,i}) \)'s represent heterogeneity among ordered dyads.
The social relations model

\[ y_{i,j} = \mu + e_{i,j}, \; i \neq j \]
\[ e_{i,j} = a_i + b_j + \epsilon_{i,j} \]

(Warner, Kenny and Stoto[1979])

**Original motivation:** Decompose variance around \( \mu \) into parts describing
- heterogeneity across rows means (outdegrees)
- heterogeneity across column means (indegrees)
- correlation between row and column means
- correlation within dyads.
The social relations model

\[ y_{i,j} = \mu + e_{i,j}, \ i \neq j \]
\[ e_{i,j} = a_i + b_j + \epsilon_{i,j} \]

(Warner, Kenny and Stoto[1979])

**Original motivation:** Decompose variance around \( \mu \) into parts describing
- heterogeneity across rows means (outdegrees)
- heterogeneity across column means (indegrees)
- correlation between row and column means
- correlation within dyads.
The social relations model

\[ y_{i,j} = \mu + e_{i,j}, \quad i \neq j \]
\[ e_{i,j} = a_i + b_j + \epsilon_{i,j} \]

(Warner, Kenny and Stoto[1979])

**Original motivation:** Decompose variance around \( \mu \) into parts describing
- heterogeneity across rows means (outdegrees)
- heterogeneity across column means (indegrees)
- correlation between row and column means
- correlation within dyads.
The social relations model

\[ y_{i,j} = \mu + e_{i,j}, \quad i \neq j \]
\[ e_{i,j} = a_i + b_j + \epsilon_{i,j} \]

(Warner, Kenny and Stoto[1979])

**Original motivation:** Decompose variance around \( \mu \) into parts describing
- heterogeneity across rows means (outdegrees)
- heterogeneity across column means (indegrees)
- correlation between row and column means
- correlation within dyads.
The social relations model

\[ y_{i,j} = \mu + e_{i,j}, \quad i \neq j \]
\[ e_{i,j} = a_i + b_j + \epsilon_{i,j} \]

(Warner, Kenny and Stoto[1979])

**Original motivation:** Decompose variance around \( \mu \) into parts describing
- heterogeneity across rows means (outdegrees)
- heterogeneity across column means (indegrees)
- correlation between row and column means
- correlation within dyads.
The social relations model

\[ y_{i,j} = \mu + e_{i,j}, \quad i \neq j \]
\[ e_{i,j} = a_i + b_j + \epsilon_{i,j} \]

(Warner, Kenny and Stoto[1979])

**Original motivation:** Decompose variance around \( \mu \) into parts describing
- heterogeneity across rows means (outdegrees)
- heterogeneity across column means (indegrees)
- correlation between row and column means
- correlation within dyads.
The social relations model

Random effects representation:

\[ y_{i,j} = \mu + e_{i,j} \]
\[ e_{i,j} = a_i + b_j + \epsilon_{i,j} \]
\[ \{(a_1, b_1), \ldots, (a_n, b_n)\} \sim \text{i.i.d. } N(0, \Sigma_{ab}) \]
\[ \{(\epsilon_{i,j}, \epsilon_{j,i}) i \neq j\} \sim \text{i.i.d. } N(0, \Sigma_{\epsilon}) \]

The \( Y \) matrix is an RCE array.
(Wong [1982], Li and Loken[2002])

Regression modeling:

\[ y_{i,j} = \beta^T x_{i,j} + a_i + b_j + \epsilon_{i,j} \]

The \( Y \) matrix is partially exchangeable, the error matrix \( E \) is an RCE array.
(Hoff[2005])
The social relations model

Random effects representation:

\[ y_{i,j} = \mu + e_{i,j} \]
\[ e_{i,j} = a_i + b_j + \epsilon_{i,j} \]
\[ \{(a_1, b_1), \ldots, (a_n, b_n)\} \sim \text{i.i.d. } N(0, \Sigma_{ab}) \]
\[ \{(\epsilon_{i,j}, \epsilon_{j,i}) | i \neq j\} \sim \text{i.i.d. } N(0, \Sigma_{\epsilon}) \]

The \( Y \) matrix is an RCE array.
(Wong [1982], Li and Loken[2002])

Regression modeling:

\[ y_{i,j} = \beta^T x_{i,j} + a_i + b_j + \epsilon_{i,j} \]

The \( Y \) matrix is partially exchangeable, the error matrix \( E \) is an RCE array.
(Hoff[2005])
Modelling non-normal data

Probit regression:

\[ \epsilon_1, \ldots, \epsilon_n \sim \text{i.i.d.} N(0, 1) \]

\[ z_i = \beta^T x_i + \epsilon_i \]

\[ y_i = 1(z_i > 0) \]

This is the latent variable representation of the probit regression model:

\[ \Pr(y_i = 1) = \Pr(z_i > 0) = \Phi(\beta^T x_i) \]

\[ p(y|\beta, X) = \prod_{i=1}^{n} \Phi(\beta^T x_i)^{y_i} [1 - \Phi(\beta^T x_i)]^{1-y_i} \]
Modelling non-normal data

Probit regression:

\[ \epsilon_1, \ldots, \epsilon_n \sim \text{i.i.d.} N(0, 1) \]
\[ z_i = \beta^T x_i + \epsilon_i \]
\[ y_i = 1(z_i > 0) \]

This is the latent variable representation of the probit regression model:

\[ \Pr(y_i = 1) = \Pr(z_i > 0) = \Phi(\beta^T x_i) \]
\[ p(y|\beta, X) = \prod_{i=1}^n \Phi(\beta^T x_i)^{y_i} [1 - \Phi(\beta^T x_i)]^{1-y_i} \]
Social relations regression for binary network data

**Threshold model:** linking latent $z$ to observed $y$:

$$y_{i,j} = 1(z_{i,j} > 0)$$

$$z_{i,j} = \beta^T x_{i,j} + e_{i,j}$$

**Social relations model:** inducing network covariance

$$e_{i,j} = a_i + b_j + \epsilon_{i,j}$$

\{(a_1, b_1), \ldots, (a_n, b_n)\} \sim \text{i.i.d. } N(0, \Sigma_{ab})

\{(\epsilon_{i,j}, \epsilon_{j,i}) | i \neq j\} \sim \text{i.i.d. } N(0, \Sigma_{\epsilon})$$
Multiplicative effects - higher order dependence

Social relations regression for binary network data

**Threshold model:** linking latent \( z \) to observed \( y \):

\[
y_{i,j} = 1(z_{i,j} > 0) \\
z_{i,j} = \beta^T x_{i,j} + e_{i,j}
\]

**Social relations model:** inducing network covariance

\[
e_{i,j} = a_i + b_j + \epsilon_{i,j} \\
\{(a_1, b_1), \ldots, (a_n, b_n)\} \sim \text{i.i.d. } N(0, \Sigma_{ab}) \\
\{(\epsilon_{i,j}, \epsilon_{j,i}) | i \neq j\} \sim \text{i.i.d. } N(0, \Sigma_{\epsilon})
\]
High School friendship network

SRM model:

\[ z_{i,j} = \beta^T x_{i,j} + a_i + b_j + \epsilon_{i,j} \]

\[ y_{i,j} = 1(z_{i,j} > 0) \]
Multiplicative effects - higher order dependence

GOF: Log odds ratio

![Histograms of log odds ratio](image)

- Left: Distribution of log odds ratio.
- Right: Enhanced distribution with different colors.

52/76
GOF: Reciprocity
GOF: Degree distribution (independence model)
GOF: Degree distribution (SRM)
### Regression estimates

#### Independent probit:

<table>
<thead>
<tr>
<th></th>
<th>rgpa</th>
<th>rsmoke</th>
<th>cgpa</th>
<th>csmoke</th>
<th>igrade</th>
<th>igpa</th>
<th>ismoke</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>est</strong></td>
<td>-0.006</td>
<td>-0.147</td>
<td>0.094</td>
<td>-0.107</td>
<td>0.422</td>
<td>0.070</td>
<td>0.013</td>
</tr>
<tr>
<td><strong>sd</strong></td>
<td>0.023</td>
<td>0.026</td>
<td>0.021</td>
<td>0.026</td>
<td>0.039</td>
<td>0.019</td>
<td>0.031</td>
</tr>
<tr>
<td><strong>t</strong></td>
<td>-0.279</td>
<td>-5.709</td>
<td>4.477</td>
<td>-4.138</td>
<td>10.795</td>
<td>3.749</td>
<td>0.414</td>
</tr>
</tbody>
</table>

#### SRM probit:

<table>
<thead>
<tr>
<th></th>
<th>rgpa</th>
<th>rsmoke</th>
<th>cgpa</th>
<th>csmoke</th>
<th>igrade</th>
<th>igpa</th>
<th>ismoke</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>est</strong></td>
<td>-0.002</td>
<td>-0.188</td>
<td>0.103</td>
<td>-0.133</td>
<td>0.502</td>
<td>0.062</td>
<td>0.035</td>
</tr>
<tr>
<td><strong>sd</strong></td>
<td>0.038</td>
<td>0.045</td>
<td>0.056</td>
<td>0.071</td>
<td>0.050</td>
<td>0.022</td>
<td>0.040</td>
</tr>
<tr>
<td><strong>t</strong></td>
<td>-0.061</td>
<td>-4.211</td>
<td>1.850</td>
<td>-1.886</td>
<td>10.004</td>
<td>2.755</td>
<td>0.872</td>
</tr>
</tbody>
</table>
Any more model deficiencies?
Transitivity

\[ t(Y) = \sum_{i\neq j\neq k} y_{i,j}y_{i,k}y_{j,k} \]

\[ t(Y_{obs}) = 1380 \]
Random effects models

Ordinary regression models can be represented as

$$y_{i,j} \sim \beta^T x_{i,j} + e_{i,j}$$

A simple “latent variable” model might include additive node effects:

$$e_{i,j} = a_i + b_j + \epsilon_{i,j} \Rightarrow y_{i,j} \sim \beta^T x_{i,j} + a_i + b_j + \epsilon_{i,j}$$

$$\{(a_1, b_1), \ldots, (a_n, b_n)\}$$ represent nodal heterogeneity, additive on the regressor scale.

Inclusion of these effects in the model can dramatically improve

- within-sample model fit (measured by $R^2$, likelihood ratio, BIC, etc.);
- out-of-sample predictive performance (measured by cross-validation).

But this model only captures heterogeneity of outdegree/indegree, and can’t represent more complicated structure, such as clustering, transitivity, etc.
Random effects models

Ordinary regression models can be represented as

\[ y_{i,j} \sim \beta^T x_{i,j} + e_{i,j} \]

A simple “latent variable” model might include additive node effects:

\[ e_{i,j} = a_i + b_j + \epsilon_{i,j} \Rightarrow y_{i,j} \sim \beta^T x_{i,j} + a_i + b_j + \epsilon_{i,j} \]

\{ (a_1, b_1), \ldots, (a_n, b_n) \} represent nodal heterogeneity, additive on the regressor scale.

Inclusion of these effects in the model can dramatically improve

- within-sample model fit (measured by \( R^2 \), likelihood ratio, BIC, etc.);
- out-of-sample predictive performance (measured by cross-validation).

But this model only captures heterogeneity of outdegree/indegree, and can’t represent more complicated structure, such as clustering, transitivity, etc.
Ordinary regression models can be represented as

\[ y_{i,j} \sim \beta^T x_{i,j} + e_{i,j} \]

A simple “latent variable” model might include additive node effects:

\[ e_{i,j} = a_i + b_j + \epsilon_{i,j} \quad \Rightarrow \quad y_{i,j} \sim \beta^T x_{i,j} + a_i + b_j + \epsilon_{i,j} \]

\{ (a_1, b_1), \ldots, (a_n, b_n) \} represent nodal heterogeneity, additive on the regressor scale.

Inclusion of these effects in the model can dramatically improve

- within-sample model fit (measured by \( R^2 \), likelihood ratio, BIC, etc.);
- out-of-sample predictive performance (measured by cross-validation).

But this model only captures heterogeneity of outdegree/indegree, and can’t represent more complicated structure, such as clustering, transitivity, etc.
Ordinary regression models can be represented as
\[ y_{i,j} \sim \beta^T x_{i,j} + e_{i,j} \]

A simple “latent variable” model might include additive node effects:
\[ e_{i,j} = a_i + b_j + \epsilon_{i,j} \Rightarrow y_{i,j} \sim \beta^T x_{i,j} + a_i + b_j + \epsilon_{i,j} \]

\( \{(a_1, b_1), \ldots, (a_n, b_n)\} \) represent nodal heterogeneity, additive on the regressor scale.

Inclusion of these effects in the model can dramatically improve
- within-sample model fit (measured by \( R^2 \), likelihood ratio, BIC, etc.);
- out-of-sample predictive performance (measured by cross-validation).

But this model only captures heterogeneity of outdegree/indegree, and can’t represent more complicated structure, such as clustering, transitivity, etc.
Model building principles

- **Statistical inference** utilizes **probability models**
- **Networks and relational data** are represented by **matrices and arrays**

Social network analysis can utilize probability models of matrices and arrays.

We will construct social network models based on these tools:

1. **Probability**: symmetry considerations (exchangeability) will motivate latent variable models generally.
2. **Matrix algebra**: matrix decomposition methods will motivate latent factor models specifically.
Model building principles

- Statistical inference utilizes probability models
- Networks and relational data are represented by matrices and arrays

Social network analysis can utilize probability models of matrices and arrays.

We will construct social network models based on these tools:

1. Probability: symmetry considerations (exchangeability) will motivate latent variable models generally.
2. Matrix algebra: matrix decomposition methods will motivate latent factor models specifically.
Model building principles

- Statistical inference utilizes probability models
- Networks and relational data are represented by matrices and arrays

Social network analysis can utilize probability models of matrices and arrays.

We will construct social network models based on these tools:

1. **Probability**: symmetry considerations (exchangeability) will motivate latent variable models generally.
2. **Matrix algebra**: matrix decomposition methods will motivate latent factor models specifically.
Model building principles

- Statistical inference utilizes probability models
- Networks and relational data are represented by matrices and arrays

Social network analysis can utilize probability models of matrices and arrays.

We will construct social network models based on these tools:

1. **Probability**: symmetry considerations (exchangeability) will motivate latent variable models generally.
2. **Matrix algebra**: matrix decomposition methods will motivate latent factor models specifically.
Recall from linear algebra:

- Every $m \times n$ matrix $Z$ can be written

$$Z = UDV^T$$

$$z_{i,j} = u_i^T D v_j$$

where $D = \text{diag}(d_1, \ldots, d_n)$, $U$ and $V$ are orthonormal.

- If $UDV^T$ is the svd of $Z$, then

$$\hat{Z}_k \equiv U_{[1:k]} D_{[1:k,1:k]} V_{[1:k]}^T$$

is the least-squares rank-$k$ approximation to $Z$. 
Recall from linear algebra:

- Every $m \times n$ matrix $Z$ can be written
  \[ Z = UDV^T \]
  \[ z_{i,j} = u_i^T D v_j \]
  where $D = \text{diag}(d_1, \ldots, d_n)$, $U$ and $V$ are orthonormal.

- If $UDV^T$ is the svd of $Z$, then
  \[ \hat{Z}_k \equiv U_{[1:k]} D_{[1:k,1:k]} V_{[1:k]}^T \]
  is the least-squares rank-$k$ approximation to $Z$. 
Recall from linear algebra:

- Every $m \times n$ matrix $Z$ can be written
  \[ Z = UDV^T \]
  \[ z_{i,j} = u_i^T D v_j \]
  where $D = \text{diag}(d_1, \ldots, d_n)$, $U$ and $V$ are orthonormal.

- If $UDV^T$ is the svd of $Z$, then
  \[ \hat{Z}_k \equiv U_{[1:k]} D_{[1:k,1:k]} V_{[1:k]}^T \]
  is the least-squares rank-$k$ approximation to $Z$. 

Least squares approximations of increasing rank
Model-based SVD

Probit version of the SVD:

\[ y_{i,j} = g(z_{i,j}) , \]
\[ z_{i,j} = u_i^T D v_j + \epsilon_{i,j} , \]
\[ \{ \epsilon_{i,j} \} \sim \text{normal}(0, 1) \]

where \( g \) is a nondecreasing function

where \( u_i, v_i \in \mathbb{R}^K \), \( D=\text{diag}(d_1, \ldots, d_K) \)

Writing \( \{ z_{i,j} \} \) as a matrix,

\[ Z = UDV^T + E \]
\[ Y = g(Z) \]
Model-based SVD

Probit version of the SVD:

\[ y_{i,j} = g(z_{i,j}) , \]

\[ z_{i,j} = u_i^T D v_j + \epsilon_{i,j} , \]

\( \{\epsilon_{i,j}\} \sim \text{normal}(0,1) \)

Writing \( \{z_{i,j}\} \) as a matrix,

\[ Z = UDV^T + \mathbf{E} \]

\[ Y = g(Z) \]
Exchangeability for asymmetric relational matrices

Suppose our model \( \Pr() \) for \( Y = \{ Y_{i,j}, i = 1, \ldots, n, j = 1, \ldots, n \} \) is RCE:

\[
\Pr(Y = \{ y_{i,j}, i = 1, \ldots, n, j = 1, \ldots, n \}) = \Pr(Y = \{ y_{\pi_i, \pi_j}, i = 1, \ldots, n, j = 1, \ldots, n \})
\]

Then

\[
Y_{i,j} = g(\theta, a_i, b_j, \epsilon_{i,j})
\]

\[
(a_1, b_1), \ldots, (a_n, b_n) \sim iid \quad p_{ab}
\]

\[
\{(\epsilon_{i,j}, \epsilon_{j,i})\} \sim iid \quad p_{\epsilon}
\]

- The parameter \( \theta \) represents \textbf{global features} of the matrix.
- The \( a_i \)'s represent \textbf{nodal sender features}.
- The \( b_j \)'s represent \textbf{nodal receiver features}.
- The \( (\epsilon_{i,j}, \epsilon_{j,i}) \)'s represent \textbf{heterogeneity among ordered dyads}. 
Exchangeability for asymmetric relational matrices

Suppose our model $\Pr()$ for $\mathbf{Y} = \{Y_{i,j}, i = 1, \ldots, n, j = 1, \ldots, n\}$ is RCE:

$$\Pr(\mathbf{Y} = \{y_{i,j}, i = 1, \ldots, n, j = 1, \ldots, n\}) = \Pr(\mathbf{Y} = \{y_{\pi_i,\pi_j}, i = 1, \ldots, n, j = 1, \ldots, n\})$$

Then

$$Y_{i,j} = g(\theta, a_i, b_j, \epsilon_{i,j})$$

$$(a_1, b_1), \ldots, (a_n, b_n) \overset{iid}{\sim} p_{ab}$$

$$\{(\epsilon_{i,j}, \epsilon_{j,i})\} \overset{iid}{\sim} p_{\epsilon}$$

- The parameter $\theta$ represents global features of the matrix.
- The $a_i$'s represent nodal sender features.
  - The $b_j$'s represent nodal receiver features.
  - The $(\epsilon_{i,j}, \epsilon_{j,i})$'s represent heterogeneity among ordered dyads.
Exchangeability for asymmetric relational matrices

Suppose our model $\Pr()$ for $Y = \{Y_{i,j}, i = 1, \ldots, n, j = 1, \ldots, n\}$ is RCE:

$$\Pr(Y = \{y_{i,j}, i = 1, \ldots, n, j = 1, \ldots, n\}) = \Pr(Y = \{y_{\pi_i, \pi_j}, i = 1, \ldots, n, j = 1, \ldots, n\})$$

Then

$$Y_{i,j} = g(\theta, a_i, b_j, \epsilon_{i,j})$$

$$(a_1, b_1), \ldots, (a_n, b_n) \overset{iid}{\sim} p_{ab}$$

$$\{(\epsilon_{i,j}, \epsilon_{j,i})\} \overset{iid}{\sim} p_\epsilon$$

- The parameter $\theta$ represents **global features** of the matrix.
- The $a_i$'s represent **nodal sender features**.
- The $b_j$'s represent **nodal receiver features**.
- The $(\epsilon_{i,j}, \epsilon_{j,i})$'s represent **heterogeneity among ordered dyads**.
Suppose our model $Pr()$ for $Y = \{ Y_{i,j}, i = 1, \ldots, n,j = 1, \ldots, n \}$ is RCE:

$$Pr(Y = \{y_{i,j}, i = 1, \ldots, n,j = 1, \ldots, n\}) = Pr(Y = \{y_{\pi,i,\pi,j}, i = 1, \ldots, n,j = 1, \ldots, n\})$$

Then

$$Y_{i,j} = g(\theta, a_i, b_j, \epsilon_{i,j})$$

$$(a_1, b_1), \ldots, (a_n, b_n) \overset{iid}{\sim} p_{ab}$$

$$\{(\epsilon_{i,j}, \epsilon_{j,i})\} \overset{iid}{\sim} p_{\epsilon}$$

- The parameter $\theta$ represents global features of the matrix.
- The $a_i$'s represent nodal sender features.
- The $b_j$'s represent nodal receiver features.
- The $(\epsilon_{i,j}, \epsilon_{j,i})$'s represent heterogeneity among ordered dyads.
Exchangeability for asymmetric relational matrices

Suppose our model \( \Pr() \) for \( Y = \{ Y_{i,j}, i = 1, \ldots, n, j = 1, \ldots, n \} \) is RCE:

\[
\Pr(Y = \{ y_{i,j}, i = 1, \ldots, n, j = 1, \ldots, n \}) = \Pr(Y = \{ y_{\pi_{i}, \pi_{j}}, i = 1, \ldots, n, j = 1, \ldots, n \})
\]

Then

\[
Y_{i,j} = g(\theta, a_{i}, b_{j}, \epsilon_{i,j})
\]

\[
(a_{1}, b_{1}), \ldots, (a_{n}, b_{n}) \overset{iid}{\sim} p_{ab}
\]

\[
\{(\epsilon_{i,j}, \epsilon_{j,i})\} \overset{iid}{\sim} p_{\epsilon}
\]

- The parameter \( \theta \) represents **global features** of the matrix.
- The \( a_{i} \)'s represent **nodal sender features**.
- The \( b_{j} \)'s represent **nodal receiver features**.
- The \( (\epsilon_{i,j}, \epsilon_{j,i}) \)'s represent **heterogeneity among ordered dyads**.
Latent factor model: an exchangeable latent variable model

(Hoff, Raftery and Handcock 2002, Hoff 2005, Hoff 2008)

- Each node $i$ has (unknown) row and column latent factors

$$u_i, v_i \in \mathbb{R}^K$$

- The probability of a tie from $i$ to $j$ depends on their latent factors

$$\Pr(Y_{i,j} = 1 | u_i, u_j) \sim \mu + u_i^T D v_j, \quad D = \begin{pmatrix} d_1 & 0 & 0 \\ 0 & d_2 & 0 \\ 0 & 0 & d_3 \end{pmatrix}$$

- The positions are unknown but exchangeable a priori:

$$(u_1, v_1), \ldots, (u_n, v_n) \overset{iid}{\sim} \text{mvnorm}(\mu, \Sigma)$$

Model characteristics:

- nodes with similar factors may have a large or small probability of a tie
- nodes with similar factors are similar in their relations
Latent factor model: an exchangeable latent variable model

(Hoff, Raftery and Handcock 2002, Hoff 2005, Hoff 2008)

- Each node $i$ has (unknown) row and column latent factors

$$u_i, v_i \in \mathbb{R}^K$$

- The probability of a tie from $i$ to $j$ depends on their latent factors

$$\Pr(Y_{i,j} = 1|u_i, u_j) \sim \mu + u_i^T D v_j, \quad D = \begin{pmatrix} d_1 & 0 & 0 \\ 0 & d_2 & 0 \\ 0 & 0 & d_3 \end{pmatrix}$$

- The positions are unknown but exchangeable a priori:

$$(u_1, v_1), \ldots, (u_n, v_n) \overset{iid}{\sim} \text{mvnorm}(\mu, \Sigma)$$

Model characteristics:

- nodes with similar factors may have a large or small probability of a tie
- nodes with similar factors are similar in their relations
Latent factor model: an exchangeable latent variable model

(Hoff, Raftery and Handcock 2002, Hoff 2005, Hoff 2008)

- Each node $i$ has (unknown) row and column latent factors
  \[ u_i, v_i \in \mathbb{R}^K \]

- The probability of a tie from $i$ to $j$ depends on their latent factors
  \[
  \Pr(Y_{i,j} = 1|u_i, u_j) \sim \mu + u_i^T D v_j, \quad D = \begin{pmatrix}
  d_1 & 0 & 0 \\
  0 & d_2 & 0 \\
  0 & 0 & d_3
\end{pmatrix}
  \]

- The positions are unknown but exchangeable a priori:
  \[
  (u_1, v_1), \ldots, (u_n, v_n) \overset{iid}{\sim} \text{mvnorm}(\mu, \Sigma)
  \]

Model characteristics:
- nodes with similar factors may have a large or small probability of a tie
- nodes with similar factors are similar in their relations
Latent factor model: an exchangeable latent variable model

(Hoff, Raftery and Handcock 2002, Hoff 2005, Hoff 2008)

- Each node $i$ has (unknown) row and column latent factors
  
  $$u_i, v_i \in \mathbb{R}^K$$

- The probability of a tie from $i$ to $j$ depends on their latent factors

  $$\Pr(Y_{i,j} = 1 | u_i, u_j) \sim \mu + u_i^T Dv_j, \quad D = \begin{pmatrix} d_1 & 0 & 0 \\ 0 & d_2 & 0 \\ 0 & 0 & d_3 \end{pmatrix}$$

- The positions are unknown but exchangeable a priori:

  $$(u_1, v_1), \ldots, (u_n, v_n) \overset{iid}{\sim} \text{mvnorm}(\mu, \Sigma)$$

Model characteristics:

- nodes with similar factors may have a large or small probability of a tie
- nodes with similar factors are similar in their relations
Network patterns

What structures can such a model represent?

**Two important types of patterns:**

**Homophily:** Similar nodes link to each other
- “similar” may be in terms of unobserved characteristics
- homophily leads to transitive or clustered social networks
- observed transitivity may be due to exogenous or endogenous factors
  (See Shalizi and Thomas 2010 for a more careful discussion)

**Stochastic equivalence:** Similar nodes have similar relational patterns
- similar nodes may or may not link to each other
- equivalent nodes can be thought of as having the same “role”

**Descriptive measures:**
- Transitivity (global measure): \( \sum_{i,j,k} y_{i,j}y_{j,k}y_{k,i} \)
- Stochastic equivalence (local measure): \( \rho_{i,j} = \text{cor}(y_{[i,]}, y_{[j,]}) \)
Network patterns

What structures can such a model represent?

Two important types of patterns:

Homophily: Similar nodes link to each other
- “similar” may be in terms of unobserved characteristics
- homophily leads to transitive or clustered social networks
- observed transitivity may be due to exogenous or endogenous factors
  (See Shalizi and Thomas 2010 for a more careful discussion)

Stochastic equivalence: Similar nodes have similar relational patterns
- similar nodes may or may not link to each other
- equivalent nodes can be thought of as having the same “role”

Descriptive measures:
- Transitivity (global measure): \[ \sum_{i,j,k} y_{i,j} y_{j,k} y_{k,i} \]
- Stochastic equivalence (local measure): \[ \rho_{i,j} = \text{cor}(y_{[i,i]}, y_{[j,j]}) \]
Network patterns

What structures can such a model represent?

**Two important types of patterns:**

**Homophily:** Similar nodes link to each other
- “similar” may be in terms of unobserved characteristics
- homophily leads to transitive or clustered social networks
- observed transitivity may be due to exogenous or endogenous factors
  (See Shalizi and Thomas 2010 for a more careful discussion)

**Stochastic equivalence:** Similar nodes have similar relational patterns
- similar nodes may or may not link to each other
- equivalent nodes can be thought of as having the same “role”

**Descriptive measures:**
- Transitivity (global measure): $\sum_{i,j,k} y_{i,j}y_{j,k}y_{k,i}$
- Stochastic equivalence (local measure): $\rho_{i,j} = \text{cor}(y_{[i,.]}, y_{[j,.]})$
Multiplicative effects - higher order dependence

Homophily and stochastic equivalence

Do these features occur in real networks?

How well can the model represent these networks?
Homophily and stochastic equivalence

Do these features occur in real networks?
How well can the model represent these networks?
Homophily and stochastic equivalence in real networks

- **AddHealth friendships**: friendships among 247 12th-graders
- **Word neighbors in Genesis**: neighboring occurrences among 158 words
- **Protein binding interactions**: binding patterns among 230 proteins
Understanding eigenvectors and eigenvalues

\[ Z = U^T D V + E \]

\[ z_{i,j} = u_i^T D v_j + \epsilon_{i,j} \]

\[ = \sum_{r=1}^{R} d_r u_{i,r} v_{j,r} + \epsilon_{i,j} \]

For example, in a rank-2 model, we have

\[ z_{i,j} = d_1 (u_{i,1} \times v_{j,1}) + d_2 (u_{i,2} \times v_{j,2}) + \epsilon_{i,j} \]

Interpretation

- \( u_{i,r} \approx u_{j,r} \): similarity of latent factors represents equivalence (local)
- \( U \approx V \): positive association represents homophily (global)
- \( U \approx -V \): negative association represents antihomophily (global)

see also Hoff(2008)
Understanding eigenvectors and eigenvalues

\[ Z = U^T D V + E \]

\[ z_{i,j} = u_i^T D v_j + \epsilon_{i,j} \]

\[ = \sum_{r=1}^{R} d_r u_{i,r} v_{j,r} + \epsilon_{i,j} \]

For example, in a rank-2 model, we have

\[ z_{i,j} = d_1 (u_{i,1} \times v_{j,1}) + d_2 (u_{i,2} \times v_{j,2}) + \epsilon_{i,j} \]

Interpretation

- \( u_{i,r} \approx u_{j,r} \): similarity of latent factors represents equivalence (local)
- \( U \approx V \): positive association represents homophily (global)
- \( U \approx -V \): negative association represents antihomophily (global)

see also Hoff(2008)
Understanding eigenvectors and eigenvalues

\[ Z = U^T D V + E \]

\[ z_{i,j} = u_i^T D v_j + \epsilon_{i,j} \]

\[ = \sum_{r=1}^{R} d_r u_{i,r} v_{j,r} + \epsilon_{i,j} \]

For example, in a rank-2 model, we have

\[ z_{i,j} = d_1 (u_{i,1} \times v_{j,1}) + d_2 (u_{i,2} \times v_{j,2}) + \epsilon_{i,j} \]

Interpretation

- \( u_{i,r} \approx u_{j,r} \): similarity of latent factors represents equivalence (local)
- \( U \approx V \): positive association represents homophily (global)
- \( U \approx -V \): negative association represents antihomophily (global)

see also Hoff(2008)
Understanding eigenvectors and eigenvalues

\[ Z = U^T D V + E \]

\[ z_{i,j} = u_i^T D v_j + \epsilon_{i,j} \]
\[ = \sum_{r=1}^{R} d_r u_{i,r} v_{j,r} + \epsilon_{i,j} \]

For example, in a rank-2 model, we have

\[ z_{i,j} = d_1 (u_{i,1} \times v_{j,1}) + d_2 (u_{i,2} \times v_{j,2}) + \epsilon_{i,j} \]

Interpretation

- \( u_{i,r} \approx u_{j,r} \): similarity of latent factors represents equivalence (local)
- \( U \approx V \): positive association represents homophily (global)
- \( U \approx -V \): negative association represents antihomophily (global)

see also Hoff(2008)
Multiplicative effects - higher order dependence

Understanding eigenvectors and eigenvalues

\[
Z = U^T D V + E
\]

\[
z_{i,j} = u_i^T D v_j + \epsilon_{i,j}
\]

\[
= \sum_{r=1}^{R} d_r u_{i,r} v_{j,r} + \epsilon_{i,j}
\]

For example, in a rank-2 model, we have

\[
z_{i,j} = d_1 (u_{i,1} \times v_{j,1}) + d_2 (u_{i,2} \times v_{j,2}) + \epsilon_{i,j}
\]

Interpretation

\begin{itemize}
  \item \( u_{i,r} \approx u_{j,r} \): similarity of latent factors represents equivalence (local)
  \item \( U \approx V \): positive association represents homophily (global)
  \item \( U \approx -V \): negative association represents antihomophily (global)
\end{itemize}

see also Hoff(2008)
Understanding eigenvectors and eigenvalues

$$Z = U^T D V + E$$

$$z_{i,j} = u_i^T D v_j + \epsilon_{i,j}$$

$$= \sum_{r=1}^{R} d_r u_{i,r} v_{j,r} + \epsilon_{i,j}$$

For example, in a rank-2 model, we have

$$z_{i,j} = d_1 (u_{i,1} \times v_{j,1}) + d_2 (u_{i,2} \times v_{j,2}) + \epsilon_{i,j}$$

Interpretation

- $u_{i,r} \approx u_{j,r}$: similarity of latent factors represents equivalence (local)
- $U \approx V$: positive association represents homophily (global)
- $U \approx -V$: negative association represents antihomophily (global)

see also Hoff(2008)
Understanding eigenvectors and eigenvalues

\[ Z = U^T D V + E \]

\[ z_{i,j} = u_i^T D v_j + \epsilon_{i,j} \]

\[ = \sum_{r=1}^{R} d_r u_{i,r} v_{j,r} + \epsilon_{i,j} \]

For example, in a rank-2 model, we have

\[ z_{i,j} = d_1 (u_{i,1} \times v_{j,1}) + d_2 (u_{i,2} \times v_{j,2}) + \epsilon_{i,j} \]

Interpretation

- \( u_{i,r} \approx u_{j,r} \): similarity of latent factors represents equivalence (local)
- \( U \approx V \): positive association represents homophily (global)
- \( U \approx -V \): negative association represents antihomophily (global)

see also Hoff(2008)
What about the SRM?

LFMs can represent row, column and dyadic correlation, but not efficiently.

It may be desireable to combine the LFM and SRM:

\[
z_{i,j} = \beta^T x_{i,j} + u_i^T D v_j + a_i + b_j + \epsilon_{i,j}
\]

\[
\{(a_1, b_1), \ldots, (a_n, b_n)\} \sim \text{i.i.d. } N(0, \Sigma_{ab})
\]

\[
\{(\epsilon_{i,j}, \epsilon_{j,i}) : i \leq j\} \sim \text{i.i.d. } N(0, \Sigma_{\epsilon})
\]

(Hoff (2005))
High School friendship networks

\[ n \quad E[\bar{y}_i.] \quad sd[\bar{y}_i.] \quad sd[\bar{y}_{ij}] \]

\[
\begin{array}{cccc}
145 & 3.59 & 2.95 & 5.15 \\
\end{array}
\]
High School friendship networks

**SRM probit:**

\[ z_{i,j} = \beta^T x_{i,j} + a_i + b_j + \epsilon_{i,j} \]

<table>
<thead>
<tr>
<th></th>
<th>rgpa</th>
<th>rsmoke</th>
<th>cgpa</th>
<th>csmoke</th>
<th>igrade</th>
<th>igpa</th>
<th>ismoke</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>est</strong></td>
<td>-0.002</td>
<td>-0.188</td>
<td>0.103</td>
<td>-0.133</td>
<td>0.502</td>
<td>0.062</td>
<td>0.035</td>
</tr>
<tr>
<td><strong>sd</strong></td>
<td>0.038</td>
<td>0.045</td>
<td>0.056</td>
<td>0.071</td>
<td>0.050</td>
<td>0.022</td>
<td>0.040</td>
</tr>
<tr>
<td><strong>t</strong></td>
<td>-0.061</td>
<td>-4.211</td>
<td>1.850</td>
<td>-1.886</td>
<td>10.004</td>
<td>2.755</td>
<td>0.872</td>
</tr>
</tbody>
</table>
High School friendship networks

- Reciprocity
- Transitive triples
High School friendship networks

**AME probit:**

\[ z_{i,j} = \beta^T x_{i,j} + a_i + b_j + u_i^T Dv_j + \epsilon_{i,j} \]

<table>
<thead>
<tr>
<th></th>
<th>rgpa</th>
<th>rsmoke</th>
<th>cgpa</th>
<th>csmoke</th>
<th>igrade</th>
<th>igpa</th>
<th>ismoke</th>
</tr>
</thead>
<tbody>
<tr>
<td>est</td>
<td>-0.008</td>
<td>-0.230</td>
<td>0.120</td>
<td>-0.157</td>
<td>0.622</td>
<td>0.069</td>
<td>0.049</td>
</tr>
<tr>
<td>sd</td>
<td>0.042</td>
<td>0.050</td>
<td>0.062</td>
<td>0.079</td>
<td>0.060</td>
<td>0.026</td>
<td>0.046</td>
</tr>
<tr>
<td>t</td>
<td>-0.192</td>
<td>-4.628</td>
<td>1.916</td>
<td>-1.974</td>
<td>10.441</td>
<td>2.664</td>
<td>1.06</td>
</tr>
</tbody>
</table>
Summary

- **Exchangeability** implies a latent variable representation
  1. additive effects (SRM) can capture network covariance
  2. multiplicative effects (LFM) can capture higher-order dependence

- The vanilla AME model can be extended in many directions
  1. **Regression**: Relating network behavior to nodal attributes
  2. **Clustering**: Identifying groups of nodes
  3. **Multivariate**: Multivariate, multiway relational data
Summary

- **Exchangeability** implies a latent variable representation
  1. additive effects (SRM) can capture network covariance
  2. multiplicative effects (LFM) can capture higher-order dependence

- The vanilla AME model can be extended in many directions
  1. **Regression**: Relating network behavior to nodal attributes
  2. **Clustering**: Identifying groups of nodes
  3. **Multivariate**: Multivariate, multiway relational data
Summary

- Exchangeability implies a latent variable representation
  1. additive effects (SRM) can capture network covariance
  2. multiplicative effects (LFM) can capture higher-order dependence

- The vanilla AME model can be extended in many directions
  1. Regression: Relating network behavior to nodal attributes
  2. Clustering: Identifying groups of nodes
  3. Multivariate: Multivariate, multiway relational data
Summary

- **Exchangeability** implies a latent variable representation
  1. additive effects (SRM) can capture network covariance
  2. multiplicative effects (LFM) can capture higher-order dependence

- The vanilla AME model can be extended in many directions
  1. **Regression**: Relating network behavior to nodal attributes
  2. **Clustering**: Identifying groups of nodes
  3. **Multivariate**: Multivariate, multiway relational data
Summary

- **Exchangeability** implies a latent variable representation
  1. additive effects (SRM) can capture network covariance
  2. multiplicative effects (LFM) can capture higher-order dependence

- The vanilla AME model can be extended in many directions
  1. **Regression**: Relating network behavior to nodal attributes
  2. **Clustering**: Identifying groups of nodes
  3. **Multivariate**: Multivariate, multiway relational data
Summary

- **Exchangeability** implies a latent variable representation
  1. additive effects (SRM) can capture network covariance
  2. multiplicative effects (LFM) can capture higher-order dependence

- The vanilla AME model can be extended in many directions
  1. **Regression**: Relating network behavior to nodal attributes
  2. **Clustering**: Identifying groups of nodes
  3. **Multivariate**: Multivariate, multiway relational data
Exchangeability implies a latent variable representation
1. additive effects (SRM) can capture network covariance
2. multiplicative effects (LFM) can capture higher-order dependence

The vanilla AME model can be extended in many directions
1. Regression: Relating network behavior to nodal attributes
2. Clustering: Identifying groups of nodes
3. Multivariate: Multivariate, multiway relational data
Summary

- **Exchangeability** implies a latent variable representation
  1. additive effects (SRM) can capture network covariance
  2. multiplicative effects (LFM) can capture higher-order dependence

- The vanilla AME model can be extended in many directions
  1. **Regression**: Relating network behavior to nodal attributes
  2. **Clustering**: Identifying groups of nodes
  3. **Multivariate**: Multivariate, multiway relational data