Latent factor models for relational data

Peter Hoff

Statistics, Biostatistics and the CSSS
University of Washington

Outline

Relational data

Relational data consist of

- a set of units or nodes A, and
- a set of measurements $\mathbf{Y} \equiv \{y_{i,j}\}$ specific to pairs of nodes $(i,j) \in A \times A$.

Examples:

International relations

- A = countries,
- $y_{i,j}$ = indicator of a dispute initiated by i with target j.

Needle-sharing network

- A = IV drug users,
- $y_{i,j}$ = needle-sharing activity between i and j.

Protein-protein interactions

- A = proteins,
- $y_{i,j}$ = the interaction between i and j.

Not an example:

Dependence graph

- A = variables,
- $y_{i,j}$ = presence of a high correlation between i and j.

Relational data

Relational data consist of

- a set of units or nodes A, and
- a set of measurements $\mathbf{Y} \equiv \{y_{i,j}\}$ specific to pairs of nodes $(i,j) \in A \times A$.

Examples:

International relations

- A = countries,
- $y_{i,j}$ = indicator of a dispute initiated by i with target j.

Needle-sharing network

- A = IV drug users,
- $y_{i,j}$ = needle-sharing activity between i and j.

Protein-protein interactions

- A = proteins,
- $y_{i,j}$ = the interaction between i and j.

Not an example:

Dependence graph

- A = variables,
- $y_{i,j}$ = presence of a high correlation between i and j.

Relational data

Relational data consist of

- a set of units or nodes A, and
- a set of measurements $\mathbf{Y} \equiv \{y_{i,j}\}$ specific to pairs of nodes $(i,j) \in A \times A$.

Examples:

International relations

- A = countries,
- $y_{i,j}$ = indicator of a dispute initiated by i with target j.

Needle-sharing network

- A = IV drug users,
- $y_{i,j}$ = needle-sharing activity between i and j.

Protein-protein interactions

- A = proteins,
- $y_{i,j}$ = the interaction between i and j.

Not an example:

Dependence graph

- A = variables,
- $y_{i,j}$ = presence of a high correlation between i and j.

Relational data

Relational data consist of

- a set of units or nodes A, and
- a set of measurements $\mathbf{Y} \equiv \{y_{i,j}\}$ specific to pairs of nodes $(i, j) \in A \times A$.

Examples:

International relations

- $A =$ countries,
- $y_{i,j} =$ indicator of a dispute initiated by i with target j.

Needle-sharing network

- $A =$ IV drug users,
- $y_{i,j} =$ needle-sharing activity between i and j.

Protein-protein interactions

- $A =$ proteins,
- $y_{i,j} =$ the interaction between i and j.

Not an example:

Dependence graph

- $A =$ variables,
- $y_{i,j} =$ presence of a high correlation between i and j.

Relational data

Relational data consist of

- a set of units or nodes A, and
- a set of measurements $\mathbf{Y} \equiv \{y_{i,j}\}$ specific to pairs of nodes $(i,j) \in A \times A$.

Examples:

International relations

- A = countries,
- $y_{i,j}$ = indicator of a dispute initiated by i with target j.

Needle-sharing network

- A = IV drug users,
- $y_{i,j}$ = needle-sharing activity between i and j.

Protein-protein interactions

- A = proteins,
- $y_{i,j}$ = the interaction between i and j.

Not an example:

Dependence graph

- A = variables,
- $y_{i,j}$ = presence of a high correlation between i and j.

Relational data

Relational data consist of

- a set of units or nodes A, and
- a set of measurements $\mathbf{Y} \equiv \{y_{i,j}\}$ specific to pairs of nodes $(i,j) \in A \times A$.

Examples:

International relations

- $A =$ countries,
- $y_{i,j} =$ indicator of a dispute initiated by i with target j.

Needle-sharing network

- $A =$ IV drug users,
- $y_{i,j} =$ needle-sharing activity between i and j.

Protein-protein interactions

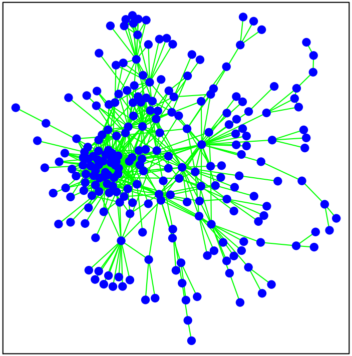
- $A =$ proteins,
- $y_{i,j} =$ the interaction between i and j.

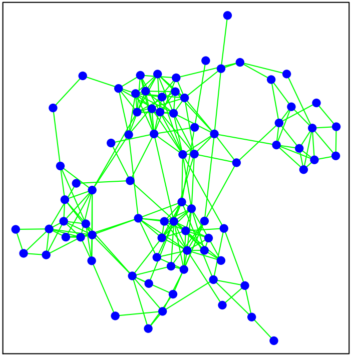
Not an example:

Dependence graph

- $A =$ variables,
- $y_{i,j} =$ presence of a high correlation between i and j.

Descriptive goals

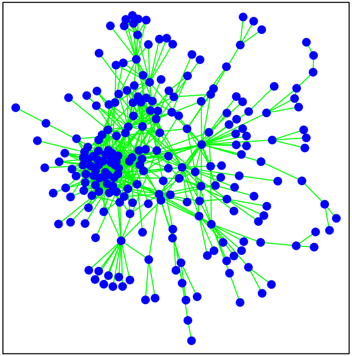


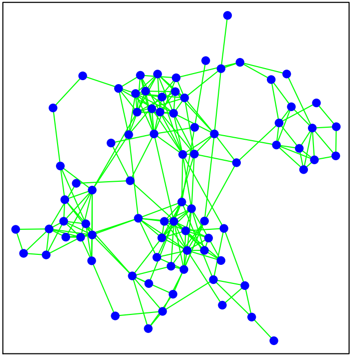


How can we summarize network patterns?

1. Are there categories of nodes corresponding to network roles? (stochastic equivalence)
2. Are there clusters of nodes with large within-cluster density? (clustering/homophily/transitivity)

Descriptive goals

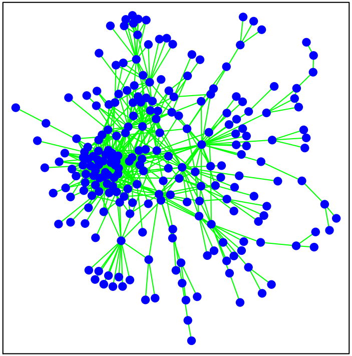


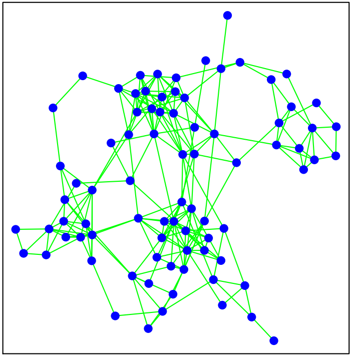


How can we summarize network patterns?

1. Are there categories of nodes corresponding to network roles? (stochastic equivalence)
2. Are there clusters of nodes with large within-cluster density? (clustering/homophily/transitivity)

Descriptive goals

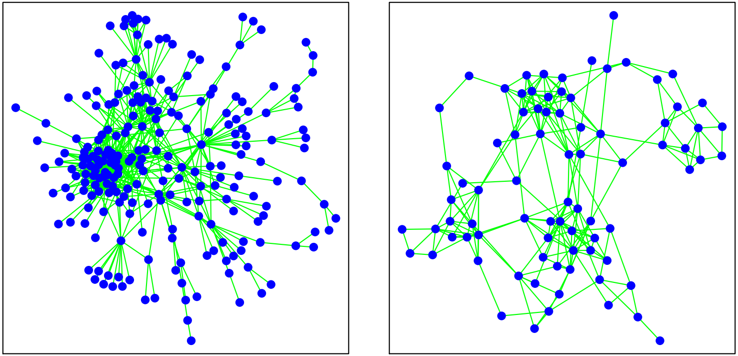




How can we summarize network patterns?

1. Are there categories of nodes corresponding to network roles? (stochastic equivalence)
2. Are there clusters of nodes with large within-cluster density? (clustering/homophily/transitivity)

Descriptive goals



How can we summarize network patterns?

1. Are there categories of nodes corresponding to network roles?
   (stochastic equivalence)
2. Are there clusters of nodes with large within-cluster density?
   (clustering/homophily/transitivity)

Inferential goals in the regression framework

$y_{i,j}$ measures $i \to j$, $\mathbf{x}_{i,j}$ is a vector of explanatory variables.

$$\mathbf{Y} = \begin{pmatrix} y_{1,1} & y_{1,2} & y_{1,3} & \text{NA} & y_{1,5} & \cdots \\ y_{2,1} & y_{2,2} & y_{2,3} & y_{2,4} & y_{2,5} & \cdots \\ y_{3,1} & \text{NA} & y_{3,3} & y_{3,4} & \text{NA} & \cdots \\ y_{4,1} & y_{4,2} & y_{4,3} & y_{4,4} & y_{4,5} & \cdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \end{pmatrix} \quad \mathbf{X} = \begin{pmatrix} \mathbf{x}_{1,1} & \mathbf{x}_{1,2} & \mathbf{x}_{1,3} & \mathbf{x}_{1,4} & \mathbf{x}_{1,5} & \cdots \\ \mathbf{x}_{2,1} & \mathbf{x}_{2,2} & \mathbf{x}_{2,3} & \mathbf{x}_{2,4} & \mathbf{x}_{2,5} & \cdots \\ \mathbf{x}_{3,1} & \mathbf{x}_{3,2} & \mathbf{x}_{3,3} & \mathbf{x}_{3,4} & \mathbf{x}_{3,5} & \cdots \\ \mathbf{x}_{4,1} & \mathbf{x}_{4,2} & \mathbf{x}_{4,3} & \mathbf{x}_{4,4} & \mathbf{x}_{4,5} & \cdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \end{pmatrix}$$

Consider a basic (generalized) linear model

$$y_{i,j} \sim \boldsymbol{\beta}^T \mathbf{x}_{i,j} + e_{i,j}$$

A model can provide

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

Inferential goals in the regression framework

$y_{i,j}$ measures $i \to j$, $\mathbf{x}_{i,j}$ is a vector of explanatory variables.

$$\mathbf{Y} = \begin{pmatrix} y_{1,1} & y_{1,2} & y_{1,3} & \text{NA} & y_{1,5} & \cdots \\ y_{2,1} & y_{2,2} & y_{2,3} & y_{2,4} & y_{2,5} & \cdots \\ y_{3,1} & \text{NA} & y_{3,3} & y_{3,4} & \text{NA} & \cdots \\ y_{4,1} & y_{4,2} & y_{4,3} & y_{4,4} & y_{4,5} & \cdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \end{pmatrix} \quad \mathbf{X} = \begin{pmatrix} \mathbf{x}_{1,1} & \mathbf{x}_{1,2} & \mathbf{x}_{1,3} & \mathbf{x}_{1,4} & \mathbf{x}_{1,5} & \cdots \\ \mathbf{x}_{2,1} & \mathbf{x}_{2,2} & \mathbf{x}_{2,3} & \mathbf{x}_{2,4} & \mathbf{x}_{2,5} & \cdots \\ \mathbf{x}_{3,1} & \mathbf{x}_{3,2} & \mathbf{x}_{3,3} & \mathbf{x}_{3,4} & \mathbf{x}_{3,5} & \cdots \\ \mathbf{x}_{4,1} & \mathbf{x}_{4,2} & \mathbf{x}_{4,3} & \mathbf{x}_{4,4} & \mathbf{x}_{4,5} & \cdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \end{pmatrix}$$

Consider a basic (generalized) linear model

$$y_{i,j} \sim \boldsymbol{\beta}^T \mathbf{x}_{i,j} + e_{i,j}$$

A model can provide

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

Inferential goals in the regression framework

$y_{i,j}$ measures $i \to j$, $\mathbf{x}_{i,j}$ is a vector of explanatory variables.

$$\mathbf{Y} = \begin{pmatrix} y_{1,1} & y_{1,2} & y_{1,3} & \text{NA} & y_{1,5} & \cdots \\ y_{2,1} & y_{2,2} & y_{2,3} & y_{2,4} & y_{2,5} & \cdots \\ y_{3,1} & \text{NA} & y_{3,3} & y_{3,4} & \text{NA} & \cdots \\ y_{4,1} & y_{4,2} & y_{4,3} & y_{4,4} & y_{4,5} & \cdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \end{pmatrix} \quad \mathbf{X} = \begin{pmatrix} \mathbf{x}_{1,1} & \mathbf{x}_{1,2} & \mathbf{x}_{1,3} & \mathbf{x}_{1,4} & \mathbf{x}_{1,5} & \cdots \\ \mathbf{x}_{2,1} & \mathbf{x}_{2,2} & \mathbf{x}_{2,3} & \mathbf{x}_{2,4} & \mathbf{x}_{2,5} & \cdots \\ \mathbf{x}_{3,1} & \mathbf{x}_{3,2} & \mathbf{x}_{3,3} & \mathbf{x}_{3,4} & \mathbf{x}_{3,5} & \cdots \\ \mathbf{x}_{4,1} & \mathbf{x}_{4,2} & \mathbf{x}_{4,3} & \mathbf{x}_{4,4} & \mathbf{x}_{4,5} & \cdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \end{pmatrix}$$

Consider a basic (generalized) linear model

$$y_{i,j} \sim \boldsymbol{\beta}^T \mathbf{x}_{i,j} + e_{i,j}$$

A model can provide

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

Inferential goals in the regression framework

$y_{i,j}$ measures $i \to j$, $\mathbf{x}_{i,j}$ is a vector of explanatory variables.

$$\mathbf{Y} = \begin{pmatrix} y_{1,1} & y_{1,2} & y_{1,3} & \text{NA} & y_{1,5} & \cdots \\ y_{2,1} & y_{2,2} & y_{2,3} & y_{2,4} & y_{2,5} & \cdots \\ y_{3,1} & \text{NA} & y_{3,3} & y_{3,4} & \text{NA} & \cdots \\ y_{4,1} & y_{4,2} & y_{4,3} & y_{4,4} & y_{4,5} & \cdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \end{pmatrix} \quad \mathbf{X} = \begin{pmatrix} \mathbf{x}_{1,1} & \mathbf{x}_{1,2} & \mathbf{x}_{1,3} & \mathbf{x}_{1,4} & \mathbf{x}_{1,5} & \cdots \\ \mathbf{x}_{2,1} & \mathbf{x}_{2,2} & \mathbf{x}_{2,3} & \mathbf{x}_{2,4} & \mathbf{x}_{2,5} & \cdots \\ \mathbf{x}_{3,1} & \mathbf{x}_{3,2} & \mathbf{x}_{3,3} & \mathbf{x}_{3,4} & \mathbf{x}_{3,5} & \cdots \\ \mathbf{x}_{4,1} & \mathbf{x}_{4,2} & \mathbf{x}_{4,3} & \mathbf{x}_{4,4} & \mathbf{x}_{4,5} & \cdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \end{pmatrix}$$

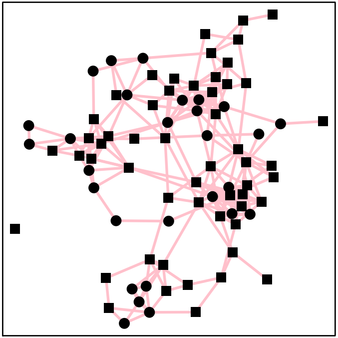
Consider a basic (generalized) linear model

$$y_{i,j} \sim \boldsymbol{\beta}^T \mathbf{x}_{i,j} + e_{i,j}$$

A model can provide

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

Adolescent health social network



Data on 82 12th graders from a single high school:

54 boys, 28 girls

$\hat{\Pr}(y_{i,j} = 1 | \text{same sex}) = 0.077$

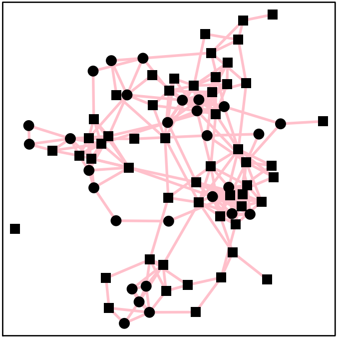
$\hat{\Pr}(y_{i,j} = 1 | \text{opposite sex}) = 0.056$

Model 0: $\{y_{i,j}\} \sim$ iid binary(θ)

Model 1: $\{y_{i,j}\}$ are independent, with

$$y_{i,j} \sim \begin{cases} \text{binary } (\theta_A) \text{ if } i \text{ and } j \text{ of same sex} \\ \text{binary } (\theta_B) \text{ if } i \text{ and } j \text{ of opposite sex} \end{cases}$$

Adolescent health social network



Data on 82 12th graders from a single high school:

54 boys, 28 girls

$\hat{\Pr}(y_{i,j} = 1 | \text{same sex}) = 0.077$

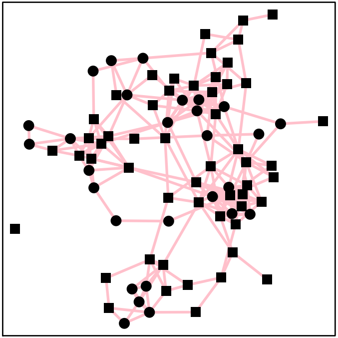
$\hat{\Pr}(y_{i,j} = 1 | \text{opposite sex}) = 0.056$

Model 0: $\{y_{i,j}\} \sim$ iid binary(θ)

Model 1: $\{y_{i,j}\}$ are independent, with

$$y_{i,j} \sim \begin{cases} \text{binary } (\theta_A) \text{ if } i \text{ and } j \text{ of same sex} \\ \text{binary } (\theta_B) \text{ if } i \text{ and } j \text{ of opposite sex} \end{cases}$$

Adolescent health social network



Data on 82 12th graders from a single high school:

54 boys, 28 girls

$\hat{\Pr}(y_{i,j} = 1 | \text{same sex}) = 0.077$

$\hat{\Pr}(y_{i,j} = 1 | \text{opposite sex}) = 0.056$

Model 0: $\{y_{i,j}\} \sim$ iid binary(θ)

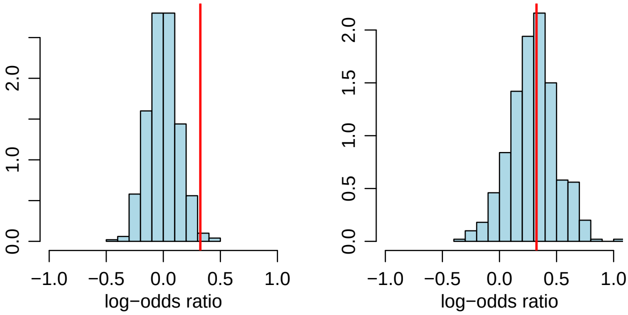
Model 1: $\{y_{i,j}\}$ are independent, with

$$y_{i,j} \sim \begin{cases} \text{binary } (\theta_A) \text{ if } i \text{ and } j \text{ of same sex} \\ \text{binary } (\theta_B) \text{ if } i \text{ and } j \text{ of opposite sex} \end{cases}$$

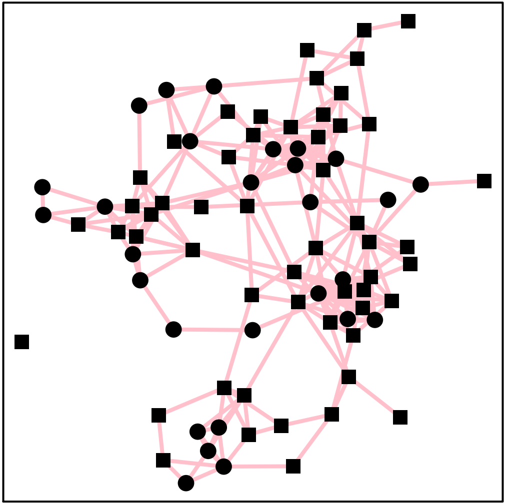
Model fit

```
glm(formula = y ~ x, family = binomial(link = "logit"))

Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)  -2.8332     0.1123  -25.24   <2e-16 ***
x             0.3471     0.1428    2.43   0.0151 *
```

This result says that a model with preferential association is a better description of the data than an i.i.d. binary model.



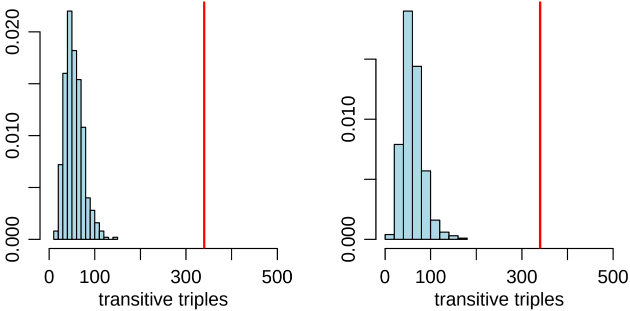
Nodal heterogeneity and independence assumptions



Model lack of fit

Neither of these models do well in terms of representing other features of the data - for example, transitivity:

$$t(\mathbf{Y}) = \sum_{i<j<k} y_{i,j} y_{j,k} y_{k,i}$$



Random effects models

Deviations from ordinary regression models can be represented as

$$y_{i,j} \sim \boldsymbol{\beta}^T \mathbf{x}_{i,j} + \gamma_{i,j}$$

A simple "latent variable" model might include additive node effects:

$$\gamma_{i,j} = a_i + a_j \quad \Rightarrow \quad y_{i,j} \sim \boldsymbol{\beta}^T \mathbf{x}_{i,j} + a_i + a_j$$

$\{a_1, \ldots, a_n\}$ 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

Deviations from ordinary regression models can be represented as

$$y_{i,j} \sim \boldsymbol{\beta}^T \mathbf{x}_{i,j} + \gamma_{i,j}$$

A simple "latent variable" model might include additive node effects:

$$\gamma_{i,j} = a_i + a_j \quad \Rightarrow \quad y_{i,j} \sim \boldsymbol{\beta}^T \mathbf{x}_{i,j} + a_i + a_j$$

$\{a_1, \ldots, a_n\}$ 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

Deviations from ordinary regression models can be represented as

$$y_{i,j} \sim \boldsymbol{\beta}^T \mathbf{x}_{i,j} + \gamma_{i,j}$$

A simple "latent variable" model might include additive node effects:

$$\gamma_{i,j} = a_i + a_j \quad \Rightarrow \quad y_{i,j} \sim \boldsymbol{\beta}^T \mathbf{x}_{i,j} + a_i + a_j$$

$\{a_1, \ldots, a_n\}$ 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

Deviations from ordinary regression models can be represented as

$$y_{i,j} \sim \boldsymbol{\beta}^T \mathbf{x}_{i,j} + \gamma_{i,j}$$

A simple "latent variable" model might include additive node effects:

$$\gamma_{i,j} = a_i + a_j \quad \Rightarrow \quad y_{i,j} \sim \boldsymbol{\beta}^T \mathbf{x}_{i,j} + a_i + a_j$$

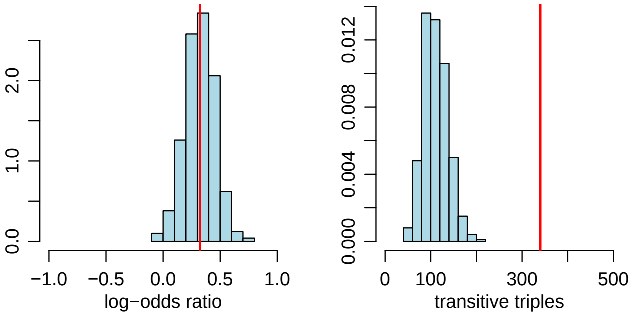
$\{a_1, \ldots, a_n\}$ represent nodal heterogeneity, additive on the regressor scale.

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.

Fit of additive effects model



Model building goals

Descriptions of local network structure

- identification of important nodes
- identification of groups of nodes
  - stochastically equivalent groups
  - high density clusters

Descriptions of global network structure

- relationship to explanatory variables
- global measures of density, transitivity, degree distribution

Inference

- prediction and imputation
- confidence intervals for regression effects
- hypothesis testing and model comparison

Model building goals

Descriptions of local network structure

- identification of important nodes
- identification of groups of nodes
  - stochastically equivalent groups
  - high density clusters

Descriptions of global network structure

- relationship to explanatory variables
- global measures of density, transitivity, degree distribution

Inference

- prediction and imputation
- confidence intervals for regression effects
- hypothesis testing and model comparison

Model building goals

Descriptions of local network structure

- identification of important nodes
- identification of groups of nodes
  - stochastically equivalent groups
  - high density clusters

Descriptions of global network structure

- relationship to explanatory variables
- global measures of density, transitivity, degree distribution

Inference

- prediction and imputation
- confidence intervals for regression effects
- hypothesis testing and model comparison

Model building goals

Descriptions of local network structure

- identification of important nodes
- identification of groups of nodes
  - stochastically equivalent groups
  - high density clusters

Descriptions of global network structure

- relationship to explanatory variables
- global measures of density, transitivity, degree distribution

Inference

- prediction and imputation
- confidence intervals for regression effects
- hypothesis testing and model comparison

Model building goals

Descriptions of local network structure

- identification of important nodes
- identification of groups of nodes
  - stochastically equivalent groups
  - high density clusters

Descriptions of global network structure

- relationship to explanatory variables
- global measures of density, transitivity, degree distribution

Inference

- prediction and imputation
- confidence intervals for regression effects
- hypothesis testing and model comparison

Model building goals

Descriptions of local network structure

- identification of important nodes
- identification of groups of nodes
  - stochastically equivalent groups
  - high density clusters

Descriptions of global network structure

- relationship to explanatory variables
- global measures of density, transitivity, degree distribution

Inference

- prediction and imputation
- confidence intervals for regression effects
- hypothesis testing and model comparison

Model building goals

Descriptions of local network structure

- identification of important nodes
- identification of groups of nodes
  - stochastically equivalent groups
  - high density clusters

Descriptions of global network structure

- relationship to explanatory variables
- global measures of density, transitivity, degree distribution

Inference

- prediction and imputation
- confidence intervals for regression effects
- hypothesis testing and model comparison

Model building goals

Descriptions of local network structure

- identification of important nodes
- identification of groups of nodes
    - stochastically equivalent groups
    - high density clusters

Descriptions of global network structure

- relationship to explanatory variables
- global measures of density, transitivity, degree distribution

Inference

- prediction and imputation
- confidence intervals for regression effects
- hypothesis testing and model comparison

Model building goals

Descriptions of local network structure

- identification of important nodes
- identification of groups of nodes
  - stochastically equivalent groups
  - high density clusters

Descriptions of global network structure

- relationship to explanatory variables
- global measures of density, transitivity, degree distribution

Inference

- prediction and imputation
- confidence intervals for regression effects
- hypothesis testing and model comparison

Model building goals

Descriptions of local network structure

- identification of important nodes
- identification of groups of nodes
  - stochastically equivalent groups
  - high density clusters

Descriptions of global network structure

- relationship to explanatory variables
- global measures of density, transitivity, degree distribution

Inference

- prediction and imputation
- confidence intervals for regression effects
- hypothesis testing and model comparison

Model building goals

Descriptions of local network structure

- identification of important nodes
- identification of groups of nodes
  - stochastically equivalent groups
  - high density clusters

Descriptions of global network structure

- relationship to explanatory variables
- global measures of density, transitivity, degree distribution

Inference

- prediction and imputation
- confidence intervals for regression effects
- hypothesis testing and model comparison

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.

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

$$\begin{aligned} Y_i &= g(\theta, \epsilon_i) \text{ , where} \\ \epsilon_1, \ldots, \epsilon_n &\stackrel{iid}{\sim} p_\epsilon \end{aligned}$$

- The parameter θ represents "global features" of the sequence.
- The ϵ_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

$$\begin{aligned} Y_i &= g(\theta, \epsilon_i) \text{ , where} \\ \epsilon_1, \ldots, \epsilon_n &\overset{iid}{\sim} p_\epsilon \end{aligned}$$

- The parameter θ represents "global features" of the sequence.
- The ϵ_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

$$\begin{aligned} Y_i &= g(\theta, \epsilon_i) \text{ , where} \\ \epsilon_1, \ldots, \epsilon_n &\overset{iid}{\sim} p_\epsilon \end{aligned}$$

- The parameter θ represents "global features" of the sequence.
- The ϵ_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

$$\begin{aligned} Y_i &= g(\theta, \epsilon_i) \text{ , where} \\ \epsilon_1, \ldots, \epsilon_n &\overset{iid}{\sim} p_\epsilon \end{aligned}$$

- The parameter θ represents "global features" of the sequence.
- The ϵ_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

$$\begin{aligned} Y_i &= g(\theta, \epsilon_i) \text{ , where} \\ \epsilon_1, \ldots, \epsilon_n &\stackrel{iid}{\sim} p_\epsilon \end{aligned}$$

- The parameter θ represents "global features" of the sequence.
- The ϵ_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 :

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

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

$$\begin{aligned} \Pr(\mathbf{Y}_1 = \mathbf{y}_1, \ldots, \mathbf{Y}_m = \mathbf{y}_m) &= \Pr(\mathbf{Y}_1 = \mathbf{y}_{\pi_1}, \ldots, \mathbf{Y}_m = \mathbf{y}_{\pi m}) \\ \Pr(Y_{i,1} = y_{i,1}, \ldots, Y_{i,n} = y_{i,n}) &= \Pr(Y_{i,1} = y_{i,\omega_1}, \ldots, Y_{i,n} = y_{i,\omega_n}) \end{aligned}$$

A double application of de Finetti's theorem implies

$$\begin{aligned} Y_{i,j} &= g(\theta, a_i, \epsilon_{i,j}) \\ a_1, \ldots, a_n &\stackrel{iid}{\sim} p_a \\ \{\epsilon_{i,j}\} &\stackrel{iid}{\sim} p_\epsilon \end{aligned}$$

- The parameter θ represents global features of the data.
- Heterogeneity in the a_i's represents across-group heterogeneity.
- Heterogeneity in the $\epsilon_{i,j}$'s represents within-group heterogeneity.

Exchangeability for nested data

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

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

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

$$\begin{aligned} \Pr(\mathbf{Y}_1 = \mathbf{y}_1, \ldots, \mathbf{Y}_m = \mathbf{y}_m) &= \Pr(\mathbf{Y}_1 = \mathbf{y}_{\pi_1}, \ldots, \mathbf{Y}_m = \mathbf{y}_{\pi m}) \\ \Pr(Y_{i,1} = y_{i,1}, \ldots, Y_{i,n} = y_{i,n}) &= \Pr(Y_{i,1} = y_{i,\omega_1}, \ldots, Y_{i,n} = y_{i,\omega_n}) \end{aligned}$$

A double application of de Finetti's theorem implies

$$\begin{aligned} Y_{i,j} &= g(\theta, a_i, \epsilon_{i,j}) \\ a_1, \ldots, a_n &\stackrel{iid}{\sim} p_a \\ \{\epsilon_{i,j}\} &\stackrel{iid}{\sim} p_\epsilon \end{aligned}$$

- The parameter θ represents global features of the data.
- Heterogeneity in the a_i's represents across-group heterogeneity.
- Heterogeneity in the $\epsilon_{i,j}$'s represents within-group heterogeneity.

Exchangeability for nested data

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

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

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

$$\begin{aligned} \Pr(\mathbf{Y}_1 = \mathbf{y}_1, \ldots, \mathbf{Y}_m = \mathbf{y}_m) &= \Pr(\mathbf{Y}_1 = \mathbf{y}_{\pi_1}, \ldots, \mathbf{Y}_m = \mathbf{y}_{\pi m}) \\ \Pr(Y_{i,1} = y_{i,1}, \ldots, Y_{i,n} = y_{i,n}) &= \Pr(Y_{i,1} = y_{i,\omega_1}, \ldots, Y_{i,n} = y_{i,\omega_n}) \end{aligned}$$

A double application of de Finetti's theorem implies

$$\begin{aligned} Y_{i,j} &= g(\theta, a_i, \epsilon_{i,j}) \\ a_1, \ldots, a_n &\overset{iid}{\sim} p_a \\ \{\epsilon_{i,j}\} &\overset{iid}{\sim} p_\epsilon \end{aligned}$$

- The parameter θ represents global features of the data.
- Heterogeneity in the a_i's represents across-group heterogeneity.
- Heterogeneity in the $\epsilon_{i,j}$'s represents within-group heterogeneity.

Exchangeability for nested data

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

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

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

$$\begin{aligned} \Pr(\mathbf{Y}_1 = \mathbf{y}_1, \ldots, \mathbf{Y}_m = \mathbf{y}_m) &= \Pr(\mathbf{Y}_1 = \mathbf{y}_{\pi_1}, \ldots, \mathbf{Y}_m = \mathbf{y}_{\pi m}) \\ \Pr(Y_{i,1} = y_{i,1}, \ldots, Y_{i,n} = y_{i,n}) &= \Pr(Y_{i,1} = y_{i,\omega_1}, \ldots, Y_{i,n} = y_{i,\omega_n}) \end{aligned}$$

A double application of de Finetti's theorem implies

$$\begin{aligned} Y_{i,j} &= g(\theta, a_i, \epsilon_{i,j}) \\ a_1, \ldots, a_n &\overset{iid}{\sim} p_a \\ \{\epsilon_{i,j}\} &\overset{iid}{\sim} p_\epsilon \end{aligned}$$

- The parameter θ represents global features of the data.
- Heterogeneity in the a_i's represents across-group heterogeneity.
- Heterogeneity in the $\epsilon_{i,j}$'s represents within-group heterogeneity.

Exchangeability for nested data

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

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

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

$$\begin{aligned} \Pr(\mathbf{Y}_1 = \mathbf{y}_1, \ldots, \mathbf{Y}_m = \mathbf{y}_m) &= \Pr(\mathbf{Y}_1 = \mathbf{y}_{\pi_1}, \ldots, \mathbf{Y}_m = \mathbf{y}_{\pi m}) \\ \Pr(Y_{i,1} = y_{i,1}, \ldots, Y_{i,n} = y_{i,n}) &= \Pr(Y_{i,1} = y_{i,\omega_1}, \ldots, Y_{i,n} = y_{i,\omega_n}) \end{aligned}$$

A double application of de Finetti's theorem implies

$$\begin{aligned} Y_{i,j} &= g(\theta, a_i, \epsilon_{i,j}) \\ a_1, \ldots, a_n &\overset{iid}{\sim} p_a \\ \{\epsilon_{i,j}\} &\overset{iid}{\sim} p_\epsilon \end{aligned}$$

- The parameter θ represents global features of the data.
- Heterogeneity in the a_i's represents across-group heterogeneity.
- Heterogeneity in the $\epsilon_{i,j}$'s represents within-group heterogeneity.

Exchangeability for nested data

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

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

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

$$\begin{aligned} \Pr(\mathbf{Y}_1 = \mathbf{y}_1, \ldots, \mathbf{Y}_m = \mathbf{y}_m) &= \Pr(\mathbf{Y}_1 = \mathbf{y}_{\pi_1}, \ldots, \mathbf{Y}_m = \mathbf{y}_{\pi m}) \\ \Pr(Y_{i,1} = y_{i,1}, \ldots, Y_{i,n} = y_{i,n}) &= \Pr(Y_{i,1} = y_{i,\omega_1}, \ldots, Y_{i,n} = y_{i,\omega_n}) \end{aligned}$$

A double application of de Finetti's theorem implies

$$\begin{aligned} Y_{i,j} &= g(\theta, a_i, \epsilon_{i,j}) \\ a_1, \ldots, a_n &\overset{iid}{\sim} p_a \\ \{\epsilon_{i,j}\} &\overset{iid}{\sim} p_\epsilon \end{aligned}$$

- The parameter θ represents global features of the data.
- Heterogeneity in the a_i's represents across-group heterogeneity.
- Heterogeneity in the $\epsilon_{i,j}$'s represents within-group heterogeneity.

Exchangeability for nested data

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

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

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

$$\begin{aligned} \Pr(\mathbf{Y}_1 = \mathbf{y}_1, \ldots, \mathbf{Y}_m = \mathbf{y}_m) &= \Pr(\mathbf{Y}_1 = \mathbf{y}_{\pi_1}, \ldots, \mathbf{Y}_m = \mathbf{y}_{\pi m}) \\ \Pr(Y_{i,1} = y_{i,1}, \ldots, Y_{i,n} = y_{i,n}) &= \Pr(Y_{i,1} = y_{i,\omega_1}, \ldots, Y_{i,n} = y_{i,\omega_n}) \end{aligned}$$

A double application of de Finetti's theorem implies

$$\begin{aligned} Y_{i,j} &= g(\theta, a_i, \epsilon_{i,j}) \\ a_1, \ldots, a_n &\stackrel{iid}{\sim} p_a \\ \{\epsilon_{i,j}\} &\stackrel{iid}{\sim} p_\epsilon \end{aligned}$$

- The parameter θ represents global features of the data.
- Heterogeneity in the a_i's represents across-group heterogeneity.
- Heterogeneity in the $\epsilon_{i,j}$'s represents within-group heterogeneity.

Exchangeability for symmetric relational matrices

Let $\mathbf{Y}$ be a symmetric 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 : $y_{B,i,j} = y_{A,\pi_i,\pi_j}$, $\pi = (3, 1, 4, 2)$

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

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

(Hoover 1982, Aldous 1983)

Exchangeability for symmetric relational matrices

Let $\mathbf{Y}$ be a symmetric 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 : $y_{B,i,j} = y_{A,\pi_i,\pi_j}$, $\pi = (3, 1, 4, 2)$

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

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

(Hoover 1982, Aldous 1983)

Exchangeability for symmetric relational matrices

Let $\mathbf{Y}$ be a symmetric 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 : $y_{B,i,j} = y_{A,\pi_i,\pi_j}$, $\pi = (3,1,4,2)$

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

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

(Hoover 1982, Aldous 1983)

Exchangeability for symmetric 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

$$\begin{aligned} Y_{i,j} &= g(\theta, a_i, a_j, \epsilon_{i,j}) = g(\theta, a_j, a_i, \epsilon_{j,i}) \\ a_1, \ldots, a_n &\overset{iid}{\sim} p_a \\ \{\epsilon_{i,j}\} &\overset{iid}{\sim} p_\epsilon \end{aligned}$$

- The parameter θ represents global features of the matrix.
- The a_i's represent nodal heterogeneity , i.e. nodal features.
- The $\epsilon_{i,j}$ represent dyad heterogeneity.

Exchangeability for symmetric 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

$$\begin{aligned} Y_{i,j} &= g(\theta, a_i, a_j, \epsilon_{i,j}) = g(\theta, a_j, a_i, \epsilon_{j,i}) \\ a_1, \ldots, a_n &\overset{iid}{\sim} p_a \\ \{\epsilon_{i,j}\} &\overset{iid}{\sim} p_\epsilon \end{aligned}$$

- The parameter θ represents global features of the matrix.
- The a_i's represent nodal heterogeneity , i.e. nodal features.
- The $\epsilon_{i,j}$ represent dyad heterogeneity.

Exchangeability for symmetric 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

$$\begin{aligned} Y_{i,j} &= g(\theta, a_i, a_j, \epsilon_{i,j}) = g(\theta, a_j, a_i, \epsilon_{j,i}) \\ a_1, \ldots, a_n &\overset{iid}{\sim} p_a \\ \{\epsilon_{i,j}\} &\overset{iid}{\sim} p_\epsilon \end{aligned}$$

- The parameter θ represents global features of the matrix.
- The a_i's represent nodal heterogeneity , i.e. nodal features.
- The $\epsilon_{i,j}$ represent dyad heterogeneity.

Exchangeability for symmetric 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

$$\begin{aligned} Y_{i,j} &= g(\theta, a_i, a_j, \epsilon_{i,j}) = g(\theta, a_j, a_i, \epsilon_{j,i}) \\ a_1, \ldots, a_n &\overset{iid}{\sim} p_a \\ \{\epsilon_{i,j}\} &\overset{iid}{\sim} p_\epsilon \end{aligned}$$

- The parameter θ represents global features of the matrix.
- The a_i's represent nodal heterogeneity , i.e. nodal features.
- The $\epsilon_{i,j}$ represent dyad heterogeneity.

Exchangeability for symmetric 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

$$\begin{aligned} Y_{i,j} &= g(\theta, a_i, a_j, \epsilon_{i,j}) = g(\theta, a_j, a_i, \epsilon_{j,i}) \\ a_1, \ldots, a_n &\overset{iid}{\sim} p_a \\ \{\epsilon_{i,j}\} &\overset{iid}{\sim} p_\epsilon \end{aligned}$$

- The parameter θ represents global features of the matrix.
- The a_i's represent nodal heterogeneity , i.e. nodal features.
- The $\epsilon_{i,j}$ represent dyad heterogeneity.

Exchangeability for symmetric 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

$$\begin{aligned} Y_{i,j} &= g(\theta, a_i, a_j, \epsilon_{i,j}) = g(\theta, a_j, a_i, \epsilon_{j,i}) \\ a_1, \ldots, a_n &\overset{iid}{\sim} p_a \\ \{\epsilon_{i,j}\} &\overset{iid}{\sim} p_\epsilon \end{aligned}$$

- The parameter θ represents global features of the matrix.
- The a_i's represent nodal heterogeneity , i.e. nodal features.
- The $\epsilon_{i,j}$ represent dyad heterogeneity.

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

$$\begin{aligned} Y_{i,j} &= g(\theta, a_i, b_j, \epsilon_{i,j}) \\ (a_1, b_1), \ldots, (a_n, b_n) &\stackrel{iid}{\sim} p_{ab} \\ \{(\epsilon_{i,j}, \epsilon_{j,i})\} &\stackrel{iid}{\sim} p_\epsilon \end{aligned}$$

- The parameter θ 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

$$\begin{aligned} Y_{i,j} &= g(\theta, a_i, b_j, \epsilon_{i,j}) \\ (a_1, b_1), \ldots, (a_n, b_n) &\overset{iid}{\sim} p_{ab} \\ \{(\epsilon_{i,j}, \epsilon_{j,i})\} &\overset{iid}{\sim} p_\epsilon \end{aligned}$$

- The parameter θ 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

$$
\begin{aligned}
Y_{i,j} &= g(\theta, a_i, b_j, \epsilon_{i,j}) \\
(a_1, b_1), \ldots, (a_n, b_n) &\overset{iid}{\sim} p_{ab} \\
\{(\epsilon_{i,j}, \epsilon_{j,i})\} &\overset{iid}{\sim} p_\epsilon
\end{aligned}
$$

- The parameter θ 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

$$\begin{aligned} Y_{i,j} &= g(\theta, a_i, b_j, \epsilon_{i,j}) \\ (a_1, b_1), \ldots, (a_n, b_n) &\stackrel{iid}{\sim} p_{ab} \\ \{(\epsilon_{i,j}, \epsilon_{j,i})\} &\stackrel{iid}{\sim} p_{\epsilon} \end{aligned}$$

- The parameter θ 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

$$\begin{aligned} Y_{i,j} &= g(\theta, a_i, b_j, \epsilon_{i,j}) \\ (a_1, b_1), \ldots, (a_n, b_n) &\overset{iid}{\sim} p_{ab} \\ \{(\epsilon_{i,j}, \epsilon_{j,i})\} &\overset{iid}{\sim} p_\epsilon \end{aligned}$$

- The parameter θ 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 class model: an exchangeable latent variable model

(Nowicki and Snijders 2001, Airoldi et al. 2008)

- Each node i is a member of an (unknown) latent class

$$a_i \in \{1, \ldots, K\}$$

- The probability of a tie between i and j is

$$\Pr(Y_{i,j} = 1 | a_i, a_j) = \theta_{a_i, a_j}$$

- The classes are unknown but exchangeable a priori:

$$a_1, \ldots, a_n \overset{iid}{\sim} \text{multinomial}(p_1, \ldots, p_K)$$

Model characteristics:

Nodes in the same class may have a small or high probability of ties:

$\theta_{k,k}$ may be small or large

Nodes in the same class are *stochastically equivalent:*

$$\Pr(\{Y_{i,1}, \ldots, Y_{i,n}\} = \{y_1, \ldots, y_n\} | a_i = k) = \Pr(\{Y_{j,1}, \ldots, Y_{j,n}\} = \{y_1, \ldots, y_n\} | a_j = k)$$

Latent class model: an exchangeable latent variable model

(Nowicki and Snijders 2001, Airoldi et al. 2008)

- Each node i is a member of an (unknown) latent class

$$a_i \in \{1, \ldots, K\}$$

- The probability of a tie between i and j is

$$\Pr(Y_{i,j} = 1 | a_i, a_j) = \theta_{a_i, a_j}$$

- The classes are unknown but exchangeable a priori:

$$a_1, \ldots, a_n \overset{iid}{\sim} \text{multinomial}(p_1, \ldots, p_K)$$

Model characteristics:

Nodes in the same class may have a small or high probability of ties:
$\theta_{k,k}$ may be small or large

Nodes in the same class are *stochastically equivalent:*

$$\Pr(\{Y_{i,1}, \ldots, Y_{i,n}\} = \{y_1, \ldots, y_n\} | a_i = k) = \Pr(\{Y_{j,1}, \ldots, Y_{j,n}\} = \{y_1, \ldots, y_n\} | a_j = k)$$

Latent class model: an exchangeable latent variable model

(Nowicki and Snijders 2001, Airoldi et al. 2008)

- Each node i is a member of an (unknown) latent class

$$a_i \in \{1, \ldots, K\}$$

- The probability of a tie between i and j is

$$\Pr(Y_{i,j} = 1 | a_i, a_j) = \theta_{a_i,a_j}$$

- The classes are unknown but exchangeable a priori:

$$a_1, \ldots, a_n \overset{iid}{\sim} \text{multinomial}(p_1, \ldots, p_K)$$

Model characteristics:

Nodes in the same class may have a small or high probability of ties:

$\theta_{k,k}$ may be small or large

Nodes in the same class are *stochastically equivalent:*

$$\Pr(\{Y_{i,1}, \ldots, Y_{i,n}\} = \{y_1, \ldots, y_n\} | a_i = k) = \Pr(\{Y_{j,1}, \ldots, Y_{j,n}\} = \{y_1, \ldots, y_n\} | a_j = k)$$

Latent class model: an exchangeable latent variable model

(Nowicki and Snijders 2001, Airoldi et al. 2008)

- Each node i is a member of an (unknown) latent class

$$a_i \in \{1, \ldots, K\}$$

- The probability of a tie between i and j is

$$\Pr(Y_{i,j} = 1 | a_i, a_j) = \theta_{a_i, a_j}$$

- The classes are unknown but exchangeable a priori:

$$a_1, \ldots, a_n \overset{iid}{\sim} \text{multinomial}(p_1, \ldots, p_K)$$

Model characteristics:

Nodes in the same class may have a small or high probability of ties:

$\theta_{k,k}$ may be small or large

Nodes in the same class are *stochastically equivalent:*

$$\Pr(\{Y_{i,1}, \ldots, Y_{i,n}\} = \{y_1, \ldots, y_n\} | a_i = k) = \Pr(\{Y_{j,1}, \ldots, Y_{j,n}\} = \{y_1, \ldots, y_n\} | a_j = k)$$

Latent distance model: an exchangeable latent variable model

(Hoff, Raftery and Handcock 2002, Handcock, Raftery and Tantrum 2007)

- Each node i has an (unknown) latent position

$$a_i \in \mathbb{R}^K$$

- The probability of a tie from i to j depends on the distance between them

$$\log \text{odds} \Pr(Y_{i,j} = 1 | a_i, a_j) = \theta - |a_i - a_j|$$

- The positions are unknown but exchangeable a priori:

$$a_1, \ldots, a_n \stackrel{iid}{\sim} \text{mvnorm}(0, \Sigma)$$

Model characteristics: Nodes nearby one another are more likely to have a tie, and will likely have similar ties to others:

$$a_i \approx a_j \Leftrightarrow \begin{cases} \Pr(Y_{i,j} = 1 | a_i, a_j) \approx \theta \\ \Pr(Y_{i,k} = 1 | a_i, a_k) \approx \Pr(Y_{j,k} = 1 | a_j, a_k) \end{cases}$$

Latent distance model: an exchangeable latent variable model

(Hoff, Raftery and Handcock 2002, Handcock, Raftery and Tantrum 2007)

- Each node i has an (unknown) latent position

$$a_i \in \mathbb{R}^K$$

- The probability of a tie from i to j depends on the distance between them

$$\log \text{odds} \Pr(Y_{i,j} = 1 | a_i, a_j) = \theta - |a_i - a_j|$$

- The positions are unknown but exchangeable a priori:

$$a_1, \ldots, a_n \overset{iid}{\sim} \text{mvnorm}(0, \Sigma)$$

Model characteristics: Nodes nearby one another are more likely to have a tie, and will likely have similar ties to others:

$$a_i \approx a_j \Leftrightarrow \begin{cases} \Pr(Y_{i,j} = 1 | a_i, a_j) \approx \theta \\ \Pr(Y_{i,k} = 1 | a_i, a_k) \approx \Pr(Y_{j,k} = 1 | a_j, a_k) \end{cases}$$

Latent distance model: an exchangeable latent variable model

(Hoff, Raftery and Handcock 2002, Handcock, Raftery and Tantrum 2007)

- Each node i has an (unknown) latent position

$$a_i \in \mathbb{R}^K$$

- The probability of a tie from i to j depends on the distance between them

$$\text{log odds} \Pr(Y_{i,j} = 1 | a_i, a_j) = \theta - |a_i - a_j|$$

- The positions are unknown but exchangeable a priori:

$$a_1, \ldots, a_n \overset{iid}{\sim} \text{mvnorm}(0, \Sigma)$$

Model characteristics: Nodes nearby one another are more likely to have a tie, and will likely have similar ties to others:

$$a_i \approx a_j \Leftrightarrow \begin{cases} \Pr(Y_{i,j} = 1 | a_i, a_j) \approx \theta \\ \Pr(Y_{i,k} = 1 | a_i, a_k) \approx \Pr(Y_{j,k} = 1 | a_j, a_k) \end{cases}$$

Latent distance model: an exchangeable latent variable model

(Hoff, Raftery and Handcock 2002, Handcock, Raftery and Tantrum 2007)

- Each node i has an (unknown) latent position

$$a_i \in \mathbb{R}^K$$

- The probability of a tie from i to j depends on the distance between them

$$\text{log odds} \Pr(Y_{i,j} = 1 | a_i, a_j) = \theta - |a_i - a_j|$$

- The positions are unknown but exchangeable a priori:

$$a_1, \ldots, a_n \stackrel{iid}{\sim} \text{mvnorm}(0, \Sigma)$$

Model characteristics: Nodes nearby one another are more likely to have a tie, and will likely have similar ties to others:

$$a_i \approx a_j \Leftrightarrow \begin{cases} \Pr(Y_{i,j} = 1 | a_i, a_j) \approx \theta \\ \Pr(Y_{i,k} = 1 | a_i, a_k) \approx \Pr(Y_{j,k} = 1 | a_j, a_k) \end{cases}$$

Latent factor model: an exchangeable latent variable model

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

- Each node i has an (unknown) latent factor

$$a_i \in \mathbb{R}^K$$

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

$$\log \text{odds} \Pr(Y_{i,j} = 1 | a_i, a_j) = \theta + a_i^T B a_j \ , B = \begin{pmatrix} b_1 & 0 & 0 \\ 0 & b_2 & 0 \\ 0 & 0 & b_3 \end{pmatrix}$$

- The positions are unknown but exchangeable a priori:

$$a_1, \ldots, a_n \stackrel{iid}{\sim} \text{mvnorm}(\mu, \Sigma)$$

Model characteristics:

nodes with similar factors may have a **large or small** probability of a tie

nodes with similar factors are approximately stochastically equivalent

Latent factor model: an exchangeable latent variable model

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

- Each node i has an (unknown) latent factor

$$a_i \in \mathbb{R}^K$$

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

$$\log \text{odds} \Pr(Y_{i,j} = 1 | a_i, a_j) = \theta + a_i^T B a_j \;, B = \begin{pmatrix} b_1 & 0 & 0 \\ 0 & b_2 & 0 \\ 0 & 0 & b_3 \end{pmatrix}$$

- The positions are unknown but exchangeable a priori:

$$a_1, \ldots, a_n \overset{iid}{\sim} \text{mvnorm}(\mu, \Sigma)$$

Model characteristics:

nodes with similar factors may have a **large or small** probability of a tie

nodes with similar factors are approximately stochastically equivalent

Latent factor model: an exchangeable latent variable model

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

- Each node i has an (unknown) latent factor

$$a_i \in \mathbb{R}^K$$

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

$$\log \text{odds} \Pr(Y_{i,j} = 1 | a_i, a_j) = \theta + a_i^T B a_j \text{ , } B = \begin{pmatrix} b_1 & 0 & 0 \\ 0 & b_2 & 0 \\ 0 & 0 & b_3 \end{pmatrix}$$

- The positions are unknown but exchangeable a priori:

$$a_1, \ldots, a_n \stackrel{iid}{\sim} \text{mvnorm}(\mu, \Sigma)$$

Model characteristics:

nodes with similar factors may have a **large or small** probability of a tie

nodes with similar factors are approximately stochastically equivalent

Latent factor model: an exchangeable latent variable model

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

- Each node i has an (unknown) latent factor

$$a_i \in \mathbb{R}^K$$

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

$$\log \text{odds} \Pr(Y_{i,j} = 1 | a_i, a_j) = \theta + a_i^T B a_j \ , B = \begin{pmatrix} b_1 & 0 & 0 \\ 0 & b_2 & 0 \\ 0 & 0 & b_3 \end{pmatrix}$$

- The positions are unknown but exchangeable a priori:

$$a_1, \ldots, a_n \overset{iid}{\sim} \text{mvnorm}(\mu, \Sigma)$$

Model characteristics:

nodes with similar factors may have a **large or small** probability of a tie

nodes with similar factors are approximately stochastically equivalent

Incorporation into regression modeling

Consider expanding upon the simple LM or GLM:

$$Y_{i,j} \sim \boldsymbol{\beta}^T \mathbf{x}_{i,j} + \gamma_{i,j}$$

- The $\{\gamma_{i,j}\}$'s represent deviations from the simple regression model
- The matrix of deviations is itself a relational (unobserved) data matrix
- The latent variable structure can describe these deviations

$$\begin{aligned}
Y_{i,j} &\sim \boldsymbol{\beta}^T \mathbf{x}_{i,j} + \gamma_{i,j} \\
\gamma_{i,j} &= \theta_{a_i,a_j} && \text{(stochastic blockmodel)} \\
\gamma_{i,j} &= -|a_i - a_j| && \text{(distance model)} \\
\gamma_{i,j} &= a_i^T \mathbf{B} a_j && \text{(factor model)}
\end{aligned}$$

Incorporation into regression modeling

Consider expanding upon the simple LM or GLM:

$$Y_{i,j} \sim \boldsymbol{\beta}^T \mathbf{x}_{i,j} + \gamma_{i,j}$$

- The $\{\gamma_{i,j}\}$'s represent deviations from the simple regression model
- The matrix of deviations is itself a relational (unobserved) data matrix
- The latent variable structure can describe these deviations

$$\begin{aligned}
Y_{i,j} &\sim \boldsymbol{\beta}^T \mathbf{x}_{i,j} + \gamma_{i,j} \\
\gamma_{i,j} &= \theta_{a_i,a_j} && \text{(stochastic blockmodel)} \\
\gamma_{i,j} &= -|a_i - a_j| && \text{(distance model)} \\
\gamma_{i,j} &= a_i^T \mathbf{B} a_j && \text{(factor model)}
\end{aligned}$$

Incorporation into regression modeling

Consider expanding upon the simple LM or GLM:

$$Y_{i,j} \sim \boldsymbol{\beta}^T \mathbf{x}_{i,j} + \gamma_{i,j}$$

- The $\{\gamma_{i,j}\}$'s represent deviations from the simple regression model
- The matrix of deviations is itself a relational (unobserved) data matrix
- The latent variable structure can describe these deviations

$$
\begin{aligned}
Y_{i,j} &\sim \boldsymbol{\beta}^T \mathbf{x}_{i,j} + \gamma_{i,j} & \\
\gamma_{i,j} &= \theta_{a_i,a_j} & \text{(stochastic blockmodel)} \\
\gamma_{i,j} &= -|a_i - a_j| & \text{(distance model)} \\
\gamma_{i,j} &= a_i^T \mathbf{B} a_j & \text{(factor model)}
\end{aligned}
$$

Incorporation into regression modeling

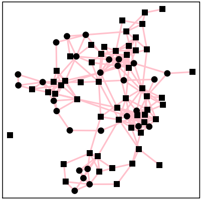
Consider expanding upon the simple LM or GLM:

$$Y_{i,j} \sim \boldsymbol{\beta}^T \mathbf{x}_{i,j} + \gamma_{i,j}$$

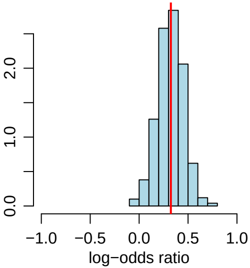
- The $\{\gamma_{i,j}\}$'s represent deviations from the simple regression model
- The matrix of deviations is itself a relational (unobserved) data matrix
- The latent variable structure can describe these deviations

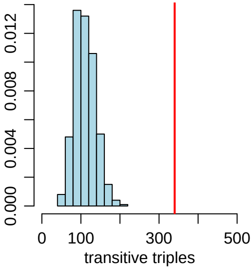
$$\begin{aligned} Y_{i,j} &\sim \boldsymbol{\beta}^T \mathbf{x}_{i,j} + \gamma_{i,j} & \\ \gamma_{i,j} &= \theta_{a_i,a_j} & \text{(stochastic blockmodel)} \\ \gamma_{i,j} &= -|a_i - a_j| & \text{(distance model)} \\ \gamma_{i,j} &= a_i^T \mathbf{B} a_j & \text{(factor model)} \end{aligned}$$

High school social network: additive effects fit



$$Y_{i,j} \sim \boldsymbol{\beta}^T \mathbf{x}_{i,j} + a_i + a_j$$



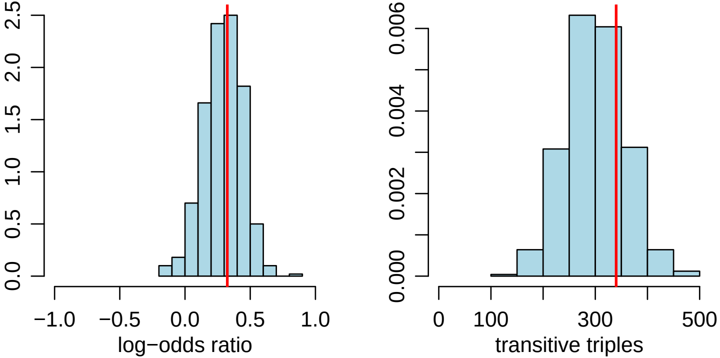


High school social network: Latent factor fit

$$Y_{i,j} \sim \boldsymbol{\beta}^T \mathbf{x}_{i,j} + \mathbf{a}_i^T \mathbf{B} \mathbf{a}_j$$

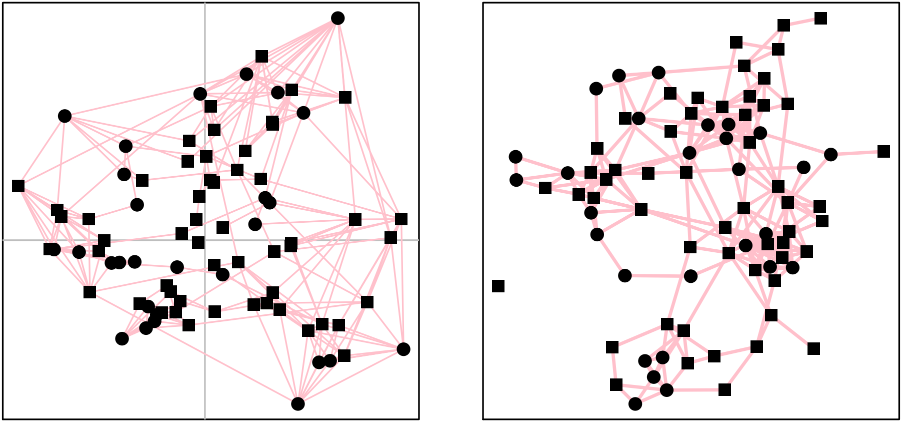
Parameters in this model can be fit with the `eigenmodel` package in R:

```
eigenmodel_mcmc(Y,X,R=3)
```

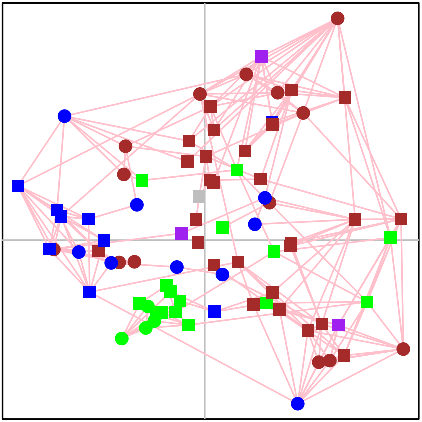


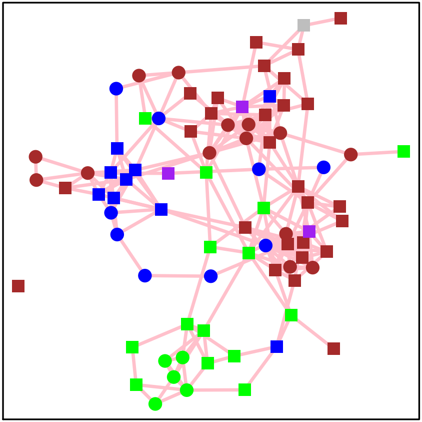
The latent factors are able to represent the network transitivity.

Underlying structure



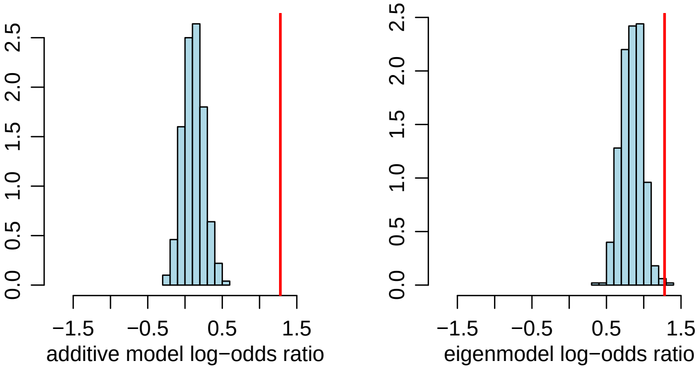
Missing variables





Missing variables

The eigenmodel, without having explicit race information, captures a large degree of the racial homophily in friendship:



Model comparisons

How do the different latent variable models compare?

What structures do they 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): $\rho_{i,j} = \mathrm{cor}(\mathbf{y}_{[i,]}, \mathbf{y}_{[j,]})$

Model comparisons

How do the different latent variable models compare?

What structures do they 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): $\rho_{i,j} = \text{cor}(\mathbf{y}_{[i,]}, \mathbf{y}_{[j,]})$

Model comparisons

How do the different latent variable models compare?

What structures do they 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)

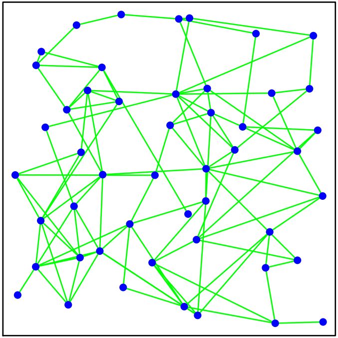
Stochastic equivalence: Similar nodes have similar relational patterns

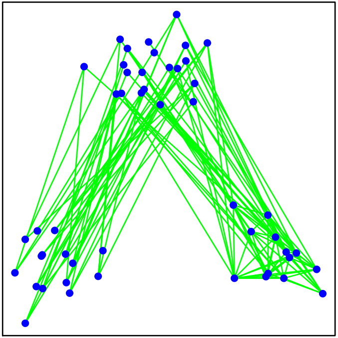
- similar nodes may or may not link to each other
- equivalent nodes can be thought of as having the same "role"

Descriptive measures:

- Transitivity (global measure): $\sum_{i,j,k} y_{i,j} y_{j,k} y_{k,i}$
- Stochastic equivalence (local measure): $\rho_{i,j} = \mathrm{cor}(\mathbf{y}_{[i,]}, \mathbf{y}_{[j,]})$

Homophily and stochastic equivalence

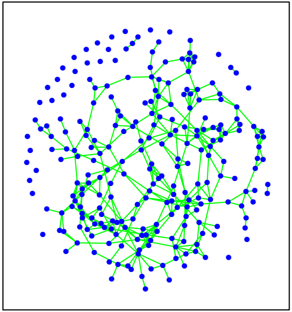


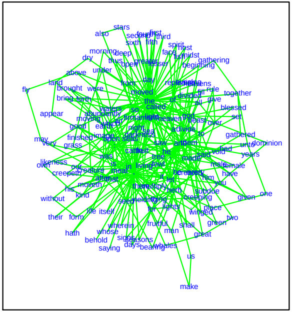


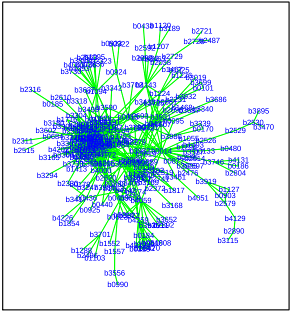
How well can the distance model represent these networks?

How well can the latent class 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

Model comparison via cross validation

1. Randomly divide the $\binom{n}{2}$ data values into 5 sets letting $s_{i,j}$ be the set to which pair $\{i,j\}$ is assigned.
2. For each $s \in \{1,\ldots,5\}$:
   2.1 Estimate model parameters with $\{y_{i,j} : s_{i,j} \neq s\}$, the data not in set s.
   2.2 Predict $\{y_{i,j} : s_{i,j} \neq s\}$ from these estimated parameters

This generates a sociomatrix $\hat{\mathbf{Y}}$, in which each entry $\hat{y}_{i,j}$ is a predicted value obtained from using a subset of the data that does not include $y_{i,j}$.

(Hoff 2008)

Model comparison via cross validation

1. Randomly divide the $\binom{n}{2}$ data values into 5 sets letting $s_{i,j}$ be the set to which pair $\{i, j\}$ is assigned.
2. For each $s \in \{1, \ldots, 5\}$:
   2.1 Estimate model parameters with $\{y_{i,j} : s_{i,j} \neq s\}$, the data not in set s.
   2.2 Predict $\{y_{i,j} : s_{i,j} \neq s\}$ from these estimated parameters

This generates a sociomatrix $\hat{\mathbf{Y}}$, in which each entry $\hat{y}_{i,j}$ is a predicted value obtained from using a subset of the data that does not include $y_{i,j}$.

(Hoff 2008)

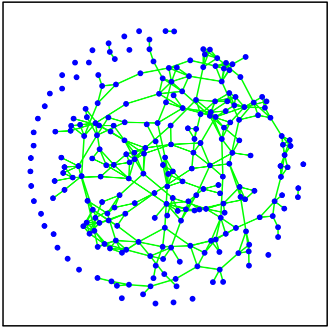
Model comparison via cross validation

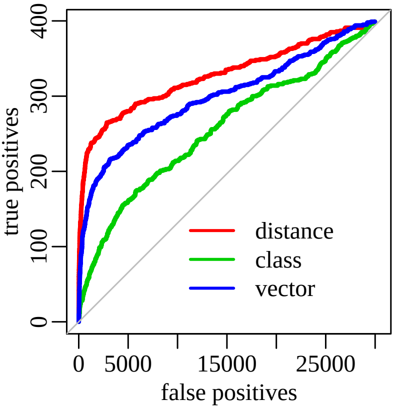
1. Randomly divide the $\binom{n}{2}$ data values into 5 sets letting $s_{i,j}$ be the set to which pair $\{i, j\}$ is assigned.
2. For each $s \in \{1, \ldots, 5\}$:
   2.1 Estimate model parameters with $\{y_{i,j} : s_{i,j} \neq s\}$, the data not in set s.
   2.2 Predict $\{y_{i,j} : s_{i,j} \neq s\}$ from these estimated parameters

This generates a sociomatrix $\hat{\mathbf{Y}}$, in which each entry $\hat{y}_{i,j}$ is a predicted value obtained from using a subset of the data that does not include $y_{i,j}$.

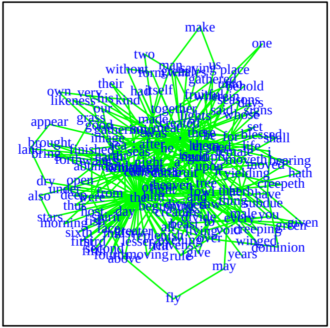
(Hoff 2008)

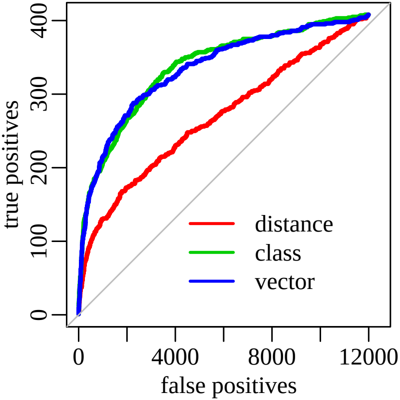
AddHealth friendships



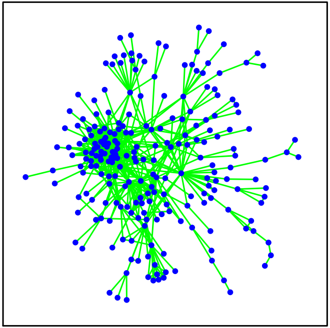


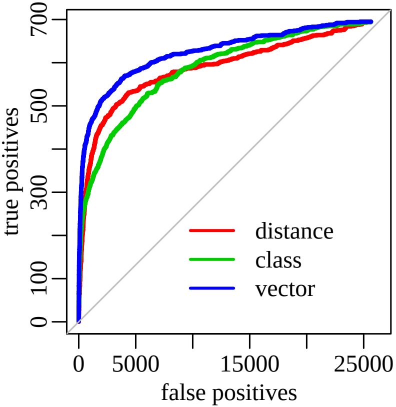
Genesis word neighbors





Protein bindings





More cross validation results

K	Add health			Genesis			Protein interaction		
	dist	class	eigen	dist	class	eigen	dist	class	eigen
3	0.82	0.64	0.75	0.62	0.82	0.82	0.83	0.79	0.88
5	0.81	0.70	0.78	0.66	0.82	0.82	0.84	0.84	0.90
10	0.76	0.69	0.80	0.74	0.82	0.82	0.85	0.86	0.90

The eigenmodel is generally as good or better than the others in each case (it can be made more comparable to the distance model if a diffuse prior is used).

More cross validation results

K	**Add health**			**Genesis**			**Protein interaction**		
	dist	class	eigen	dist	class	eigen	dist	class	eigen
3	0.82	0.64	0.75	0.62	0.82	0.82	0.83	0.79	0.88
5	0.81	0.70	0.78	0.66	0.82	0.82	0.84	0.84	0.90
10	0.76	0.69	0.80	0.74	0.82	0.82	0.85	0.86	0.90

The eigenmodel is generally as good or better than the others in each case (it can be made more comparable to the distance model if a diffuse prior is used).

Model flexibility

Probit versions of the three latent variable models all have the following form:

$$\begin{aligned} y_{i,j} &= \begin{cases} 1 & \text{if } z_{i,j} > 0 \\ 0 & \text{if } z_{i,j} \leq 0 \end{cases} \\ z_{i,j} &= \mu + \alpha(a_i, a_j) + \epsilon_{i,j} \\ \{\epsilon_{i,j} : 1 \leq i < j \leq n\} &\sim \text{i.i.d. normal}(0,1) \\ \{a_1, \ldots, a_n\} &\sim \text{i.i.d. } f(a|\psi) \end{aligned}$$

where

Latent class model:

$$\alpha(a_i, a_j) = \theta_{a_i, a_j}$$
$$a_i \in \{1, \ldots, K\}, \; i \in \{1, \ldots, n\}$$
Θ a $K \times K$ symmetric matrix

Latent distance model:

$$\alpha(a_i, a_j) = -|a_i - a_j|$$
$$a_i \in \mathbb{R}^K, \; i \in \{1, \ldots, n\}$$

Latent factor model:

$$\alpha(a_i, a_j) = a_i^T \Lambda a_j$$
$$a_i \in \mathbb{R}^K, \; i \in \{1, \ldots, n\}$$
Λ a $K \times K$ diagonal matrix.

Model flexibility

Probit versions of the three latent variable models all have the following form:

$$
\begin{aligned}
y_{i,j} &= \begin{cases} 1 & \text{if } z_{i,j} > 0 \\ 0 & \text{if } z_{i,j} \leq 0 \end{cases} \\
z_{i,j} &= \mu + \alpha(a_i, a_j) + \epsilon_{i,j} \\
\{\epsilon_{i,j} : 1 \leq i < j \leq n\} &\sim \text{i.i.d. normal}(0,1) \\
\{a_1, \ldots, a_n\} &\sim \text{i.i.d. } f(a|\psi)
\end{aligned}
$$

where

Latent class model:

$\alpha(a_i, a_j) = \theta_{a_i, a_j}$
$a_i \in \{1, \ldots, K\}, \; i \in \{1, \ldots, n\}$
Θ a $K \times K$ symmetric matrix

Latent distance model:

$\alpha(a_i, a_j) = -|a_i - a_j|$
$a_i \in \mathbb{R}^K, \; i \in \{1, \ldots, n\}$

Latent factor model:

$\alpha(a_i, a_j) = a_i^T \Lambda a_j$
$a_i \in \mathbb{R}^K, \; i \in \{1, \ldots, n\}$
Λ a $K \times K$ diagonal matrix.

Model flexibility

Probit versions of the three latent variable models all have the following form:

$$
\begin{aligned}
y_{i,j} &= \begin{cases} 1 & \text{if } z_{i,j} > 0 \\ 0 & \text{if } z_{i,j} \leq 0 \end{cases} \\
z_{i,j} &= \mu + \alpha(a_i, a_j) + \epsilon_{i,j} \\
\{\epsilon_{i,j} : 1 \leq i < j \leq n\} &\sim \text{i.i.d. normal}(0,1) \\
\{a_1, \ldots, a_n\} &\sim \text{i.i.d. } f(a|\psi)
\end{aligned}
$$

where

Latent class model:

$$
\begin{aligned}
&\alpha(a_i, a_j) = \theta_{a_i, a_j} \\
&a_i \in \{1, \ldots, K\},\ i \in \{1, \ldots, n\} \\
&\Theta \text{ a } K \times K \text{ symmetric matrix}
\end{aligned}
$$

Latent distance model:

$$
\begin{aligned}
&\alpha(a_i, a_j) = -|a_i - a_j| \\
&a_i \in \mathbb{R}^K,\ i \in \{1, \ldots, n\}
\end{aligned}
$$

Latent factor model:

$$
\begin{aligned}
&\alpha(a_i, a_j) = a_i^T \Lambda a_j \\
&a_i \in \mathbb{R}^K,\ i \in \{1, \ldots, n\} \\
&\Lambda \text{ a } K \times K \text{ diagonal matrix.}
\end{aligned}
$$

Model flexibility

Probit versions of the three latent variable models all have the following form:

$$
\begin{aligned}
y_{i,j} &= \begin{cases} 1 & \text{if } z_{i,j} > 0 \\ 0 & \text{if } z_{i,j} \leq 0 \end{cases} \\
z_{i,j} &= \mu + \alpha(a_i, a_j) + \epsilon_{i,j} \\
\{\epsilon_{i,j} : 1 \leq i < j \leq n\} &\sim \text{i.i.d. normal}(0,1) \\
\{a_1, \ldots, a_n\} &\sim \text{i.i.d. } f(a|\psi)
\end{aligned}
$$

where

Latent class model:

$\alpha(a_i, a_j) = \theta_{a_i, a_j}$
$a_i \in \{1, \ldots, K\},\ i \in \{1, \ldots, n\}$
Θ a $K \times K$ symmetric matrix

Latent distance model:

$\alpha(a_i, a_j) = -|a_i - a_j|$
$a_i \in \mathbb{R}^K,\ i \in \{1, \ldots, n\}$

Latent factor model:

$\alpha(a_i, a_j) = a_i^T \Lambda a_j$
$a_i \in \mathbb{R}^K,\ i \in \{1, \ldots, n\}$
Λ a $K \times K$ diagonal matrix.

Model flexibility

Let $\mathcal{S}_n$ be the set of symmetric $n \times n$ matrices, and let

$$
\begin{aligned}
\mathcal{C}_K &= \{C \in \mathcal{S}_n : c_{i,j} = \theta_{a_i,a_j},\ a_i \in \{1,\ldots,K\},\ \Theta \text{ a } K \times K \text{ symmetric matrix}\};\\
\mathcal{D}_K &= \{D \in \mathcal{S}_n : d_{i,j} = -|a_i - a_j|,\ a_i \in \mathbb{R}^K\};\\
\mathcal{E}_K &= \{E \in \mathcal{S}_n : e_{i,j} = a_i^T \Lambda a_j,\ a_i \in \mathbb{R}^K,\ \ \Lambda \text{ a } K \times K \text{ diagonal matrix}\}.
\end{aligned}
$$

$\mathcal{C}_K$, $\mathcal{D}_K$ and $\mathcal{E}_K$ describe the patterns representable by the class, distance and factor models respectively.

Theoretical results:

- $\mathcal{E}_K$ generalizes $\mathcal{C}_K$
- $\mathcal{E}_{K+1}$ weakly generalizes $\mathcal{D}_K$
- $\mathcal{D}_K$ does not weakly generalize $\mathcal{E}_1$

Model flexibility

Let $\mathcal{S}_n$ be the set of symmetric $n \times n$ matrices, and let

$$\begin{aligned}
\mathcal{C}_K &= \{C \in \mathcal{S}_n : c_{i,j} = \theta_{a_i,a_j},\ a_i \in \{1,\ldots,K\},\ \Theta \text{ a } K \times K \text{ symmetric matrix}\};\\
\mathcal{D}_K &= \{D \in \mathcal{S}_n : d_{i,j} = -|a_i - a_j|,\ a_i \in \mathbb{R}^K\};\\
\mathcal{E}_K &= \{E \in \mathcal{S}_n : e_{i,j} = a_i^T \Lambda a_j,\ a_i \in \mathbb{R}^K,\ \ \Lambda \text{ a } K \times K \text{ diagonal matrix}\}.
\end{aligned}$$

$\mathcal{C}_K$, $\mathcal{D}_K$ and $\mathcal{E}_K$ describe the patterns representable by the class, distance and factor models respectively.

Theoretical results:

- $\mathcal{E}_K$ generalizes $\mathcal{C}_K$
- $\mathcal{E}_{K+1}$ weakly generalizes $\mathcal{D}_K$
- $\mathcal{D}_K$ does not weakly generalize $\mathcal{E}_1$

Model flexibility

Let $\mathcal{S}_n$ be the set of symmetric $n \times n$ matrices, and let

$$\begin{aligned}
\mathcal{C}_K &= \{C \in \mathcal{S}_n : c_{i,j} = \theta_{a_i,a_j},\ a_i \in \{1,\ldots,K\},\ \Theta \text{ a } K \times K \text{ symmetric matrix}\}; \\
\mathcal{D}_K &= \{D \in \mathcal{S}_n : d_{i,j} = -|a_i - a_j|,\ a_i \in \mathbb{R}^K\}; \\
\mathcal{E}_K &= \{E \in \mathcal{S}_n : e_{i,j} = a_i^T \Lambda a_j,\ a_i \in \mathbb{R}^K,\ \ \Lambda \text{ a } K \times K \text{ diagonal matrix}\}.
\end{aligned}$$

$\mathcal{C}_K$, $\mathcal{D}_K$ and $\mathcal{E}_K$ describe the patterns representable by the class, distance and factor models respectively.

Theoretical results:

- $\mathcal{E}_K$ generalizes $\mathcal{C}_K$
- $\mathcal{E}_{K+1}$ weakly generalizes $\mathcal{D}_K$
- $\mathcal{D}_K$ does not weakly generalize $\mathcal{E}_1$

Model flexibility

Let $\mathcal{S}_n$ be the set of symmetric $n \times n$ matrices, and let

$$
\begin{aligned}
\mathcal{C}_K &= \{C \in \mathcal{S}_n : c_{i,j} = \theta_{a_i,a_j},\ a_i \in \{1,\ldots,K\},\ \Theta \text{ a } K \times K \text{ symmetric matrix}\}; \\
\mathcal{D}_K &= \{D \in \mathcal{S}_n : d_{i,j} = -|a_i - a_j|,\ a_i \in \mathbb{R}^K\}; \\
\mathcal{E}_K &= \{E \in \mathcal{S}_n : e_{i,j} = a_i^T \Lambda a_j,\ a_i \in \mathbb{R}^K,\ \ \Lambda \text{ a } K \times K \text{ diagonal matrix}\}.
\end{aligned}
$$

$\mathcal{C}_K$, $\mathcal{D}_K$ and $\mathcal{E}_K$ describe the patterns representable by the class, distance and factor models respectively.

Theoretical results:

- $\mathcal{E}_K$ generalizes $\mathcal{C}_K$
- $\mathcal{E}_{K+1}$ weakly generalizes $\mathcal{D}_K$
- $\mathcal{D}_K$ does not weakly generalize $\mathcal{E}_1$

Model flexibility

Let $\mathcal{S}_n$ be the set of symmetric $n \times n$ matrices, and let

$$
\begin{aligned}
\mathcal{C}_K &= \{C \in \mathcal{S}_n : c_{i,j} = \theta_{a_i,a_j},\ a_i \in \{1,\ldots,K\},\ \Theta \text{ a } K \times K \text{ symmetric matrix}\}; \\
\mathcal{D}_K &= \{D \in \mathcal{S}_n : d_{i,j} = -|a_i - a_j|,\ a_i \in \mathbb{R}^K\}; \\
\mathcal{E}_K &= \{E \in \mathcal{S}_n : e_{i,j} = a_i^T \Lambda a_j,\ a_i \in \mathbb{R}^K,\ \ \Lambda \text{ a } K \times K \text{ diagonal matrix}\}.
\end{aligned}
$$

$\mathcal{C}_K$, $\mathcal{D}_K$ and $\mathcal{E}_K$ describe the patterns representable by the class, distance and factor models respectively.

Theoretical results:

- $\mathcal{E}_K$ generalizes $\mathcal{C}_K$
- $\mathcal{E}_{K+1}$ weakly generalizes $\mathcal{D}_K$
- $\mathcal{D}_K$ does not weakly generalize $\mathcal{E}_1$

Matrix decompositions

Probit version of the latent factor model:

$$
\begin{aligned}
y_{i,j} &= g(z_{i,j}) \ , && \text{where } g \text{ is a nondecreasing function} \\
z_{i,j} &= \mathbf{u}_i^T \Lambda \mathbf{u}_j + \epsilon_{i,j} \ , && \text{where } \mathbf{u}_i \in \mathbb{R}^K \ , \ \Lambda = \text{diag}(\lambda_1, \ldots, \lambda_K) \\
\{\epsilon_{i,j}\} &\stackrel{iid}{\sim} \text{normal}(0,1)
\end{aligned}
$$

Writing $\{z_{i,j}\}$ as a matrix ,

$$\mathbf{Z} = \mathbf{U}\Lambda\mathbf{U}^T + \mathbf{E}$$

Recall from linear algebra:

- Every $n \times n$ symmetric matrix $\mathbf{Z}$ can be written

  $$\mathbf{Z} = \mathbf{U}\Lambda\mathbf{U}^T$$

  where $\Lambda = \text{diag}(\lambda_1, \ldots, \lambda_n)$ and $\mathbf{U}$ is orthonormal.

- If $\mathbf{U}\Lambda\mathbf{U}^T$ is the eigendecomposition of $\mathbf{Z}$, then

  $$\hat{\mathbf{Z}}_k \equiv \mathbf{U}_{[,1:k]}\Lambda_{[1:k,1:k]}\mathbf{U}_{[,1:k]}^T$$

  is the least-squares rank-k approximation to $\mathbf{Z}$.

Matrix decompositions

Probit version of the latent factor model:

$$
\begin{aligned}
y_{i,j} &= g(z_{i,j}) \ , && \text{where } g \text{ is a nondecreasing function} \\
z_{i,j} &= \mathbf{u}_i^T \Lambda \mathbf{u}_j + \epsilon_{i,j} \ , && \text{where } \mathbf{u}_i \in \mathbb{R}^K \ , \ \Lambda = \text{diag}(\lambda_1, \ldots, \lambda_K) \\
\{\epsilon_{i,j}\} &\overset{iid}{\sim} \text{normal}(0,1)
\end{aligned}
$$

Writing $\{z_{i,j}\}$ as a matrix ,

$$\mathbf{Z} = \mathbf{U}\Lambda\mathbf{U}^T + \mathbf{E}$$

Recall from linear algebra:

- Every $n \times n$ symmetric matrix $\mathbf{Z}$ can be written

$$\mathbf{Z} = \mathbf{U}\Lambda\mathbf{U}^T$$

  where $\Lambda = \text{diag}(\lambda_1, \ldots, \lambda_n)$ and $\mathbf{U}$ is orthonormal.

- If $\mathbf{U}\Lambda\mathbf{U}^T$ is the eigendecomposition of $\mathbf{Z}$, then

$$\hat{\mathbf{Z}}_k \equiv \mathbf{U}_{[,1:k]}\Lambda_{[1:k,1:k]}\mathbf{U}_{[,1:k]}^T$$

  is the least-squares rank-k approximation to $\mathbf{Z}$.

Matrix decompositions

Probit version of the latent factor model:

$$\begin{aligned} y_{i,j} &= g(z_{i,j}) \;, && \text{where } g \text{ is a nondecreasing function} \\ z_{i,j} &= \mathbf{u}_i^T \Lambda \mathbf{u}_j + \epsilon_{i,j} \;, && \text{where } \mathbf{u}_i \in \mathbb{R}^K \;,\; \Lambda = \text{diag}(\lambda_1, \ldots, \lambda_K) \\ \{\epsilon_{i,j}\} &\overset{iid}{\sim} \text{normal}(0,1) \end{aligned}$$

Writing $\{z_{i,j}\}$ as a matrix ,

$$\mathbf{Z} = \mathbf{U}\Lambda\mathbf{U}^T + \mathbf{E}$$

Recall from linear algebra:

- Every $n \times n$ symmetric matrix $\mathbf{Z}$ can be written

$$\mathbf{Z} = \mathbf{U}\Lambda\mathbf{U}^T$$

  where $\Lambda = \text{diag}(\lambda_1, \ldots, \lambda_n)$ and $\mathbf{U}$ is orthonormal.
- If $\mathbf{U}\Lambda\mathbf{U}^T$ is the eigendecomposition of $\mathbf{Z}$, then

$$\hat{\mathbf{Z}}_k \equiv \mathbf{U}_{[,1:k]}\Lambda_{[1:k,1:k]}\mathbf{U}_{[,1:k]}^T$$

  is the least-squares rank-k approximation to $\mathbf{Z}$.

Matrix decompositions

Probit version of the latent factor model:

$$
\begin{aligned}
y_{i,j} &= g(z_{i,j}) , && \text{where } g \text{ is a nondecreasing function} \\
z_{i,j} &= \mathbf{u}_i^T \Lambda \mathbf{u}_j + \epsilon_{i,j} , && \text{where } \mathbf{u}_i \in \mathbb{R}^K , \ \Lambda = \text{diag}(\lambda_1, \ldots, \lambda_K) \\
\{\epsilon_{i,j}\} &\stackrel{iid}{\sim} \text{normal}(0,1)
\end{aligned}
$$

Writing $\{z_{i,j}\}$ as a matrix ,

$$\mathbf{Z} = \mathbf{U}\Lambda\mathbf{U}^T + \mathbf{E}$$

Recall from linear algebra:

- Every $n \times n$ symmetric matrix $\mathbf{Z}$ can be written

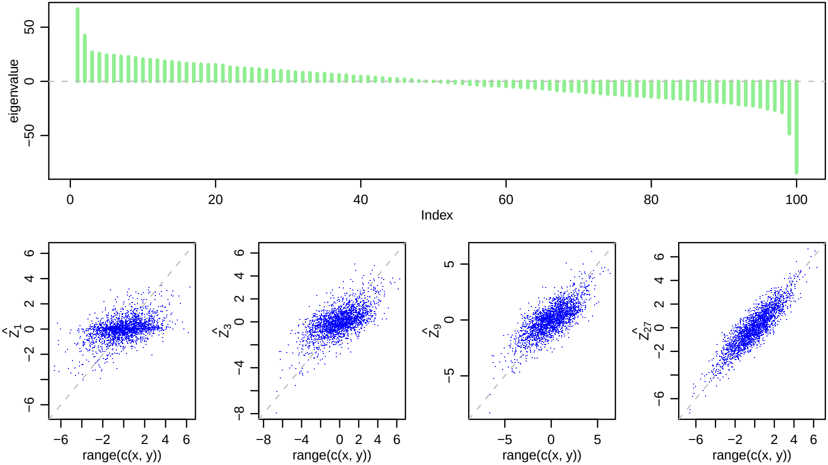
  $$\mathbf{Z} = \mathbf{U}\Lambda\mathbf{U}^T$$

  where $\Lambda = \text{diag}(\lambda_1, \ldots, \lambda_n)$ and $\mathbf{U}$ is orthonormal.
- If $\mathbf{U}\Lambda\mathbf{U}^T$ is the eigendecomposition of $\mathbf{Z}$, then

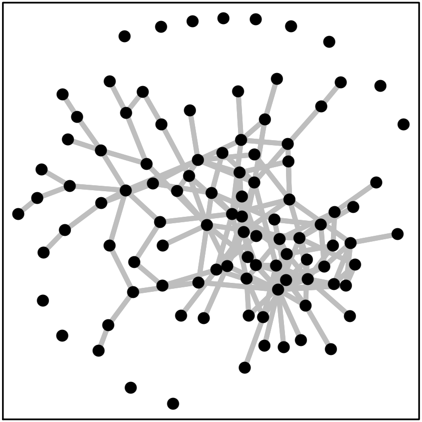
  $$\hat{\mathbf{Z}}_k \equiv \mathbf{U}_{[,1:k]}\Lambda_{[1:k,1:k]}\mathbf{U}^T_{[,1:k]}$$

  is the least-squares rank-k approximation to $\mathbf{Z}$.

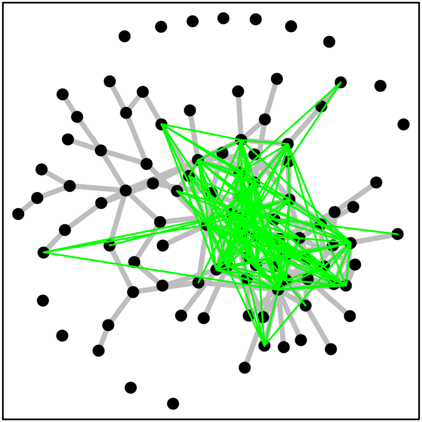
Least squares approximations of increasing rank



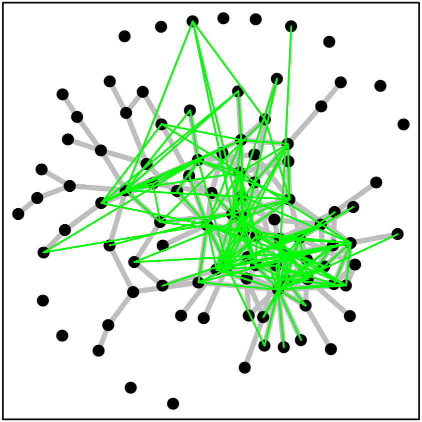
Least squares approximations of increasing rank



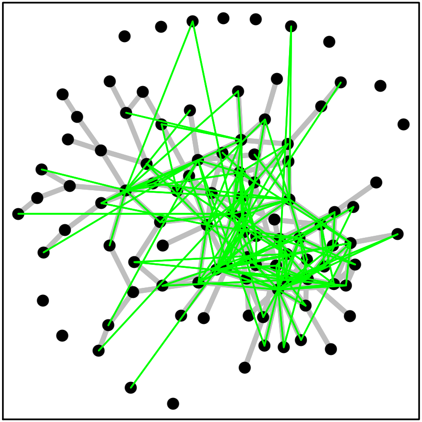
Least squares approximations of increasing rank



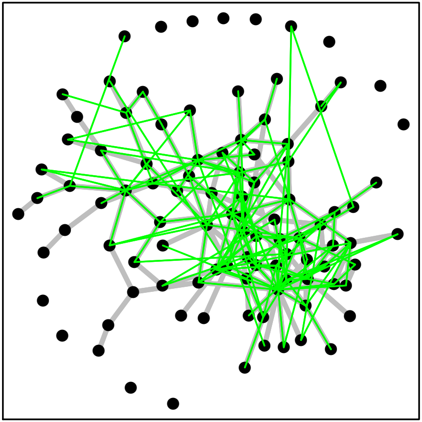
Least squares approximations of increasing rank



Least squares approximations of increasing rank



Least squares approximations of increasing rank



Estimation and Inference

Data: $\mathbf{Y} = \{y_{i,j}, 1 \leq i < j \leq n\}$

Model:

$$
\begin{aligned}
y_{i,j} &= 1 \text{ if } z_{i,j} > 0 \ , \ 0 \text{ else} \\
z_{i,j} &= \mu + \mathbf{u}_i^T \Lambda \mathbf{u}_j + \epsilon_{i,j} \ , \\
\{\epsilon_{i,j}\} &\stackrel{iid}{\sim} \text{normal}(0,1)
\end{aligned}
$$

Posterior inference:

$$
\begin{aligned}
p(\mathbf{Z}, \mu, \mathbf{U}, \Lambda | \mathbf{Y}) &\propto p(\mathbf{Y}|\mathbf{Z}, \mu, \mathbf{U}, \Lambda) p(\mathbf{Z}, \mu, \mathbf{U}, \Lambda) \\
&= p(\mathbf{Y}|\mathbf{Z}) p(\mathbf{Z}|\mu, \mathbf{U}, \Lambda) p(\mu) p(\mathbf{U}) p(\Lambda)
\end{aligned}
$$

Gibbs sampling: MCMC approximation to $p(\mathbf{Z}, \mu, \mathbf{U}, \Lambda|\mathbf{Y})$:

1. sample $z_{i,j} \sim p(z_{i,j}|y_{i,j}, \mu)$ for each pair (i,j)
2. sample $\mathbf{u}_i \sim p(\mathbf{u}_i|\mathbf{U}_{[-i],}, \mathbf{Z}_{[i,-i]}, \mu, \Lambda)$
3. sample $\Lambda \sim p(\Lambda|\mathbf{U}, \mathbf{Z}, \mu)$
4. sample $\mu \sim p(\mu|\mathbf{Y}, \mathbf{Z}, \mathbf{U}, \Lambda)$

Estimation and Inference

Data: $\mathbf{Y} = \{y_{i,j}, 1 \leq i < j \leq n\}$

Model:

$$\begin{aligned} y_{i,j} &= 1 \text{ if } z_{i,j} > 0 \quad , \quad 0 \text{ else} \\ z_{i,j} &= \mu + \mathbf{u}_i^T \Lambda \mathbf{u}_j + \epsilon_{i,j} \quad , \\ \{\epsilon_{i,j}\} &\stackrel{iid}{\sim} \text{normal}(0,1) \end{aligned}$$

Posterior inference:

$$\begin{aligned} p(\mathbf{Z}, \mu, \mathbf{U}, \Lambda | \mathbf{Y}) &\propto p(\mathbf{Y} | \mathbf{Z}, \mu, \mathbf{U}, \Lambda) p(\mathbf{Z}, \mu, \mathbf{U}, \Lambda) \\ &= p(\mathbf{Y}|\mathbf{Z}) p(\mathbf{Z}|\mu, \mathbf{U}, \Lambda) p(\mu) p(\mathbf{U}) p(\Lambda) \end{aligned}$$

Gibbs sampling: MCMC approximation to $p(\mathbf{Z}, \mu, \mathbf{U}, \Lambda | \mathbf{Y})$:

1. sample $z_{i,j} \sim p(z_{i,j} | y_{i,j}, \mu)$ for each pair (i,j)
2. sample $\mathbf{u}_i \sim p(\mathbf{u}_i | \mathbf{U}_{[-i],}, \mathbf{Z}_{[i,-i]}, \mu, \Lambda)$
3. sample $\Lambda \sim p(\Lambda | \mathbf{U}, \mathbf{Z}, \mu)$
4. sample $\mu \sim p(\mu | \mathbf{Y}, \mathbf{Z}, \mathbf{U}, \Lambda)$

Estimation and Inference

Data: $\mathbf{Y} = \{y_{i,j}, 1 \leq i < j \leq n\}$

Model:

$$\begin{aligned} y_{i,j} &= 1 \text{ if } z_{i,j} > 0 \quad , \quad 0 \text{ else} \\ z_{i,j} &= \mu + \mathbf{u}_i^T \Lambda \mathbf{u}_j + \epsilon_{i,j} \quad , \\ \{\epsilon_{i,j}\} &\stackrel{iid}{\sim} \text{normal}(0, 1) \end{aligned}$$

Posterior inference:

$$\begin{aligned} p(\mathbf{Z}, \mu, \mathbf{U}, \Lambda | \mathbf{Y}) &\propto p(\mathbf{Y} | \mathbf{Z}, \mu, \mathbf{U}, \Lambda) p(\mathbf{Z}, \mu, \mathbf{U}, \Lambda) \\ &= p(\mathbf{Y}|\mathbf{Z}) p(\mathbf{Z}|\mu, \mathbf{U}, \Lambda) p(\mu) p(\mathbf{U}) p(\Lambda) \end{aligned}$$

Gibbs sampling: MCMC approximation to $p(\mathbf{Z}, \mu, \mathbf{U}, \Lambda | \mathbf{Y})$:

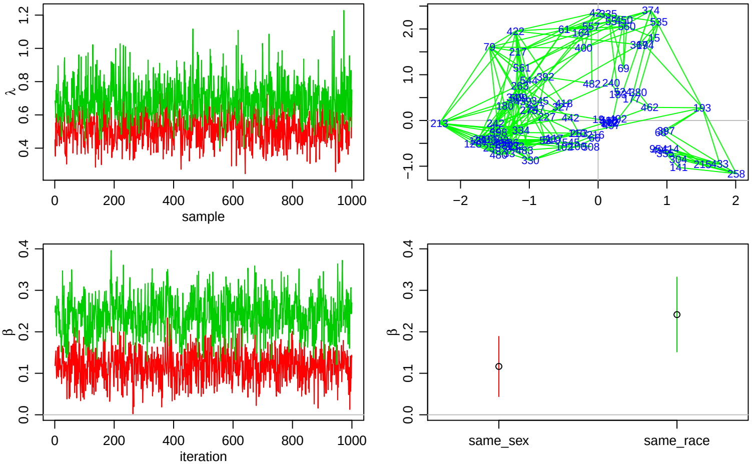
1. sample $z_{i,j} \sim p(z_{i,j} | y_{i,j}, \mu)$ for each pair (i, j)
2. sample $\mathbf{u}_i \sim p(\mathbf{u}_i | \mathbf{U}_{[-i],}, \mathbf{Z}_{[i,-i]}, \mu, \Lambda)$
3. sample $\Lambda \sim p(\Lambda | \mathbf{U}, \mathbf{Z}, \mu)$
4. sample $\mu \sim p(\mu | \mathbf{Y}, \mathbf{Z}, \mathbf{U}, \Lambda)$

R-Package eigenmodel

```
Description:

     Construct approximate samples from the posterior distribution of
     the parameters and latent variables in an eigenmodel for symmetric
     relational data.

Usage:

     eigenmodel_mcmc(Y, X = NULL, R = 2, S = 1000, seed = 1, Nss = min(S-burn, 1000), burn = 0)

Arguments:

       Y: an n x n symmetric matrix with missing diagonal entries.
          Off-diagonal missing values are allowed.

       X: an n x n x p array of regressors

       R: the rank of the approximating factor matrix

       S: number of samples from the Markov chain

    seed: a random seed

     Nss: number of samples to be saved

    burn: number of initial scans of the Markov chain to be dropped

Value: a list with the following components:

Z_postmean: posterior mean of the latent variable in the probit specification

ULU_postmean: posterior mean of the reduced-rank approximating matrix

Y_postmean: the original data matrix with missing values replaced by posterior means

L_postsamp: samples of the eigenvalues

b_postsamp: samples of the regression coefficients
```

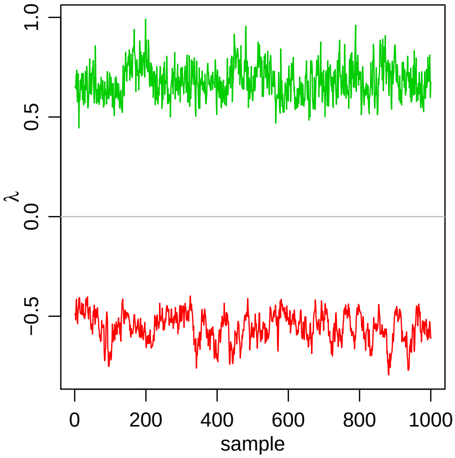
Friendship example

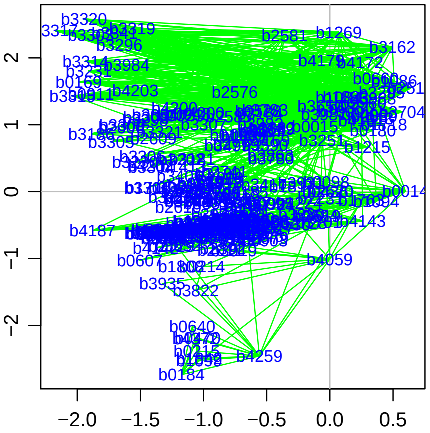
```
> library(eigenmodel)
> data(YX_Friend)
> fit<-eigenmodel_mcmc(Y=YX_Friend$Y,X=YX_Friend$X,R=2,S=100000,burn=5000)
```



Protein interaction example

```
> library(eigenmodel)
> data(Y_Pro)
> fit<-eigenmodel_mcmc(Y=Y_Pro,R=2,S=100000,burn=5000)
```





R-Software `svdmodel`

The same idea, except for asymmetric data...

Recall from linear algebra:

- Every $m \times n$ symmetric matrix $\mathbf{Z}$ can be written

$$\mathbf{Z} = \mathbf{U}\mathbf{D}\mathbf{V}^T$$

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

- If $\mathbf{U}\mathbf{D}\mathbf{V}^T$ is the svd of $\mathbf{Z}$, then

$$\hat{\mathbf{Z}}_k \equiv \mathbf{U}_{[,1:k]}\mathbf{D}_{[1:k,1:k]}\mathbf{V}^T_{[,1:k]}$$

  is the least-squares rank-k approximation to $\mathbf{Z}$.

Model:

$$\begin{aligned} y_{i,j} &= 1 \text{ if } z_{i,j} > 0 \;,\; 0 \text{ else} \\ z_{i,j} &= \mu + \mathbf{u}_i^T \mathbf{D} \mathbf{v}_j + \epsilon_{i,j} \;, \\ \{\epsilon_{i,j}\} &\overset{iid}{\sim} \text{normal}(0,1) \end{aligned}$$

R-Software svdmodel

The same idea, except for asymmetric data...

Recall from linear algebra:

- Every $m \times n$ symmetric matrix $\mathbf{Z}$ can be written

$$\mathbf{Z} = \mathbf{U}\mathbf{D}\mathbf{V}^T$$

  where $\mathbf{D} = \text{diag}(d_1, \ldots, d_n)$, $\mathbf{U}$ and $\mathbf{V}$ are orthonormal.
- If $\mathbf{U}\mathbf{D}\mathbf{V}^T$ is the svd of $\mathbf{Z}$, then

$$\hat{\mathbf{Z}}_k \equiv \mathbf{U}_{[,1:k]}\mathbf{D}_{[1:k,1:k]}\mathbf{V}^T_{[,1:k]}$$

  is the least-squares rank-k approximation to $\mathbf{Z}$.

Model:

$$\begin{aligned} y_{i,j} &= 1 \text{ if } z_{i,j} > 0 \ , \ 0 \text{ else} \\ z_{i,j} &= \mu + \mathbf{u}_i^T \mathbf{D} \mathbf{v}_j + \epsilon_{i,j} \ , \\ \{\epsilon_{i,j}\} &\stackrel{iid}{\sim} \text{normal}(0,1) \end{aligned}$$

R-Software `svdmodel`

The same idea, except for asymmetric data...

Recall from linear algebra:

- Every $m \times n$ symmetric matrix $\mathbf{Z}$ can be written

$$\mathbf{Z} = \mathbf{U}\mathbf{D}\mathbf{V}^T$$

  where $\mathbf{D} = \text{diag}(d_1, \ldots, d_n)$, $\mathbf{U}$ and $\mathbf{V}$ are orthonormal.
- If $\mathbf{U}\mathbf{D}\mathbf{V}^T$ is the svd of $\mathbf{Z}$, then

$$\hat{\mathbf{Z}}_k \equiv \mathbf{U}_{[,1:k]}\mathbf{D}_{[1:k,1:k]}\mathbf{V}^T_{[,1:k]}$$

  is the least-squares rank-k approximation to $\mathbf{Z}$.

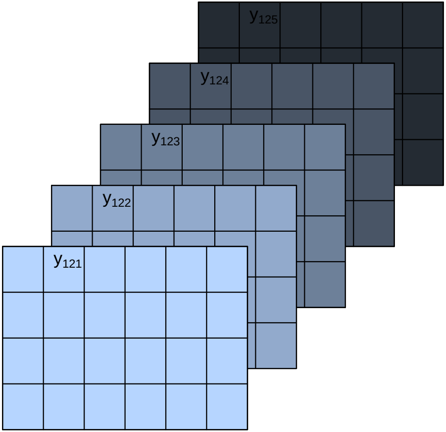
Model:

$$\begin{aligned} y_{i,j} &= 1 \text{ if } z_{i,j} > 0 \;\; , \;\; 0 \text{ else} \\ z_{i,j} &= \mu + \mathbf{u}_i^T \mathbf{D} \mathbf{v}_j + \epsilon_{i,j} \;\; , \\ \{\epsilon_{i,j}\} &\stackrel{iid}{\sim} \text{normal}(0,1) \end{aligned}$$

Multiway relational arrays

$y_{i,j,k} =$

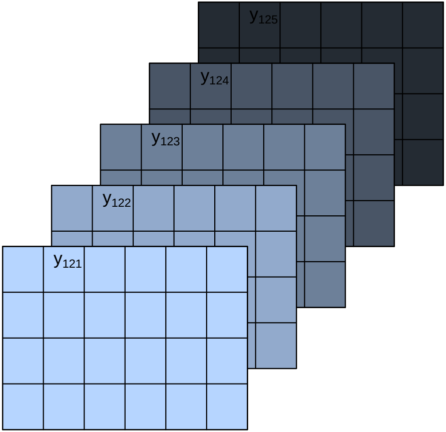
- jth measurement on ith subject under condition k (psychometrics)
- type-k relationship between i and j (relational data/network)
- relationship between i and j at time t (longitudinal relational data)



Multiway relational arrays

$y_{i,j,k} =$

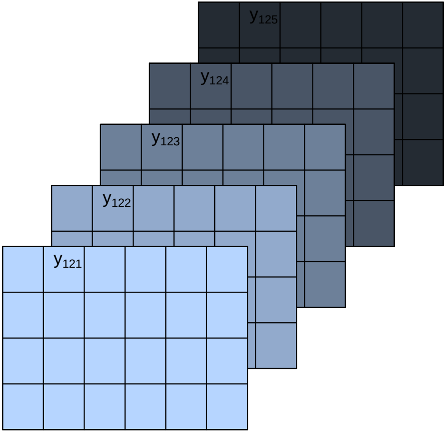
- jth measurement on ith subject under condition k (psychometrics)
- type-k relationship between i and j (relational data/network)
- relationship between i and j at time t (longitudinal relational data)



Multiway relational arrays

$y_{i,j,k} =$

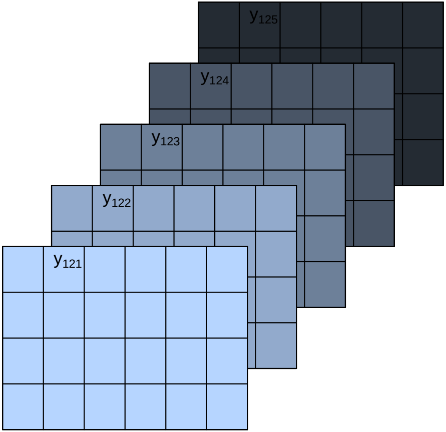
- jth measurement on ith subject under condition k (psychometrics)
- type-k relationship between i and j (relational data/network)
- relationship between i and j at time t (longitudinal relational data)



Multiway relational arrays

$y_{i,j,k} =$

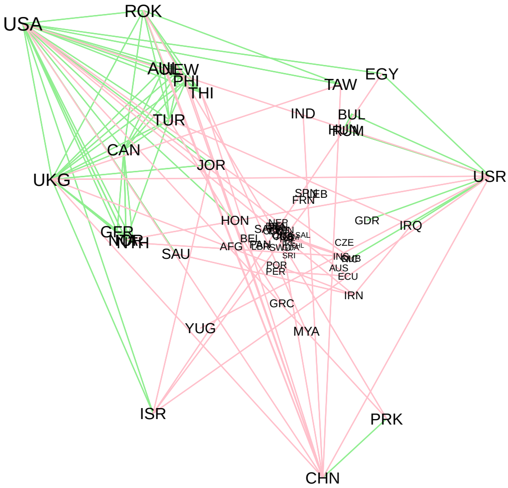
- jth measurement on ith subject under condition k (psychometrics)
- type-k relationship between i and j (relational data/network)
- relationship between i and j at time t (longitudinal relational data)



Longitudinal network example

Cold war cooperation and conflict

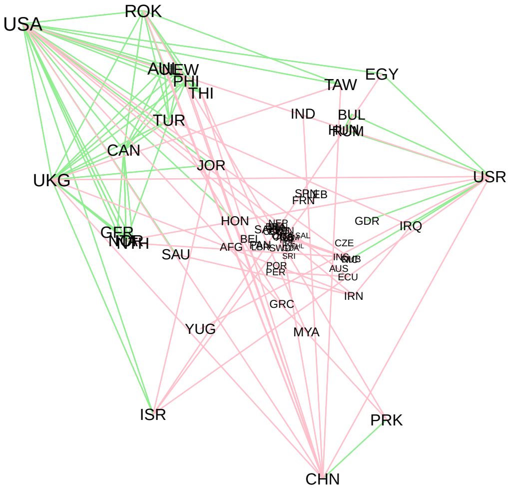
- 66 countries
- 8 years (1950,1955,...,1980,1985)
- $y_{i,j,t}$ =relation between i, j in year t
- also have data on gdp and polity



Longitudinal network example

Cold war cooperation and conflict

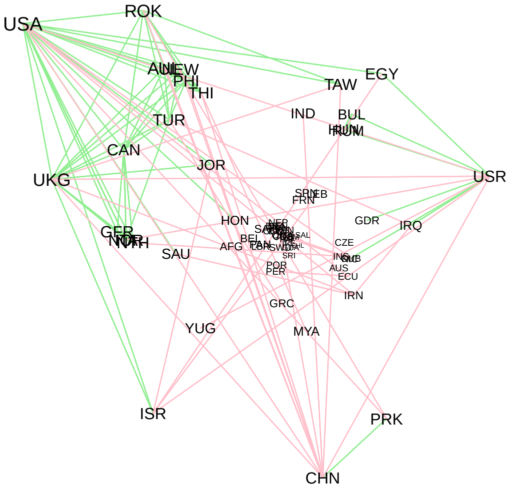
- 66 countries
- 8 years (1950,1955,...,1980,1985)
- $y_{i,j,t}$ =relation between i, j in year t
- also have data on gdp and polity



Longitudinal network example

Cold war cooperation and conflict

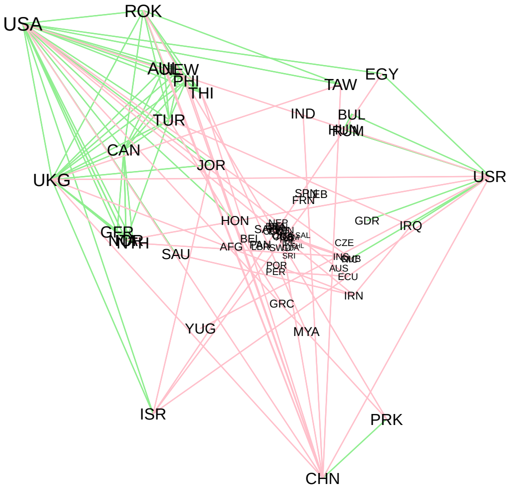
- 66 countries
- 8 years (1950,1955,...,1980,1985)
- $y_{i,j,t}$ =relation between i, j in year t
- also have data on gdp and polity



Longitudinal network example

Cold war cooperation and conflict

- 66 countries
- 8 years (1950,1955,...,1980,1985)
- $y_{i,j,t}$ =relation between i, j in year t
- also have data on gdp and polity



Reduced rank models

$$\mathbf{Y} = \mathbf{\Theta} + \mathbf{E}$$

- $\mathbf{\Theta}$ contains the "main features" we hope to recover,
- $\mathbf{E}$ is "patternless."

Matrix decomposition: If Θ is a rank-R matrix, then

$$\theta_{i,j} = \langle \mathbf{u}_i, \mathbf{v}_j \rangle = \sum_{r=1}^{R} u_{i,r} v_{j,r} \qquad \mathbf{\Theta} = \sum_{r=1}^{R} \mathbf{u}_r \mathbf{v}_r^T = \sum_{r=1}^{R} \mathbf{u}_r \circ \mathbf{v}_r$$

Array decomposition: If Θ is a rank-R array, then

$$\theta_{i,j,k} = \langle \mathbf{u}_i, \mathbf{v}_j, \mathbf{w}_k \rangle = \sum_{r=1}^{R} u_{i,r} v_{j,r} w_{k,r} \qquad \mathbf{\Theta} = \sum_{r=1}^{R} \mathbf{u}_r \circ \mathbf{v}_r \circ \mathbf{w}_r$$

(Harshman[1970], Kruskal[1976,1977], Harshman and Lundy[1984], Kruskal[1989])

Reduced rank models

$$\mathbf{Y} = \mathbf{\Theta} + \mathbf{E}$$

- $\mathbf{\Theta}$ contains the "main features" we hope to recover,
- $\mathbf{E}$ is "patternless."

Matrix decomposition: If Θ is a rank-R matrix, then

$$\theta_{i,j} = \langle \mathbf{u}_i, \mathbf{v}_j \rangle = \sum_{r=1}^{R} u_{i,r} v_{j,r} \qquad \mathbf{\Theta} = \sum_{r=1}^{R} \mathbf{u}_r \mathbf{v}_r^T = \sum_{r=1}^{R} \mathbf{u}_r \circ \mathbf{v}_r$$

Array decomposition: If Θ is a rank-R array, then

$$\theta_{i,j,k} = \langle \mathbf{u}_i, \mathbf{v}_j, \mathbf{w}_k \rangle = \sum_{r=1}^{R} u_{i,r} v_{j,r} w_{k,r} \qquad \mathbf{\Theta} = \sum_{r=1}^{R} \mathbf{u}_r \circ \mathbf{v}_r \circ \mathbf{w}_r$$

(Harshman[1970], Kruskal[1976,1977], Harshman and Lundy[1984], Kruskal[1989])

Reduced rank models

$$\mathbf{Y} = \mathbf{\Theta} + \mathbf{E}$$

- $\mathbf{\Theta}$ contains the "main features" we hope to recover,
- $\mathbf{E}$ is "patternless."

Matrix decomposition: If Θ is a rank-R matrix, then

$$\theta_{i,j} = \langle \mathbf{u}_i, \mathbf{v}_j \rangle = \sum_{r=1}^{R} u_{i,r} v_{j,r} \qquad \mathbf{\Theta} = \sum_{r=1}^{R} \mathbf{u}_r \mathbf{v}_r^T = \sum_{r=1}^{R} \mathbf{u}_r \circ \mathbf{v}_r$$

Array decomposition: If Θ is a rank-R array, then

$$\theta_{i,j,k} = \langle \mathbf{u}_i, \mathbf{v}_j, \mathbf{w}_k \rangle = \sum_{r=1}^{R} u_{i,r} v_{j,r} w_{k,r} \qquad \mathbf{\Theta} = \sum_{r=1}^{R} \mathbf{u}_r \circ \mathbf{v}_r \circ \mathbf{w}_r$$

(Harshman[1970], Kruskal[1976,1977], Harshman and Lundy[1984], Kruskal[1989])

Reduced rank models

$$\mathbf{Y} = \mathbf{\Theta} + \mathbf{E}$$

- $\mathbf{\Theta}$ contains the "main features" we hope to recover,
- $\mathbf{E}$ is "patternless."

Matrix decomposition: If Θ is a rank-R matrix, then

$$\theta_{i,j} = \langle \mathbf{u}_i, \mathbf{v}_j \rangle = \sum_{r=1}^{R} u_{i,r} v_{j,r} \qquad \mathbf{\Theta} = \sum_{r=1}^{R} \mathbf{u}_r \mathbf{v}_r^T = \sum_{r=1}^{R} \mathbf{u}_r \circ \mathbf{v}_r$$

Array decomposition: If Θ is a rank-R array, then

$$\theta_{i,j,k} = \langle \mathbf{u}_i, \mathbf{v}_j, \mathbf{w}_k \rangle = \sum_{r=1}^{R} u_{i,r} v_{j,r} w_{k,r} \qquad \mathbf{\Theta} = \sum_{r=1}^{R} \mathbf{u}_r \circ \mathbf{v}_r \circ \mathbf{w}_r$$

(Harshman[1970], Kruskal[1976,1977], Harshman and Lundy[1984], Kruskal[1989])

Reduced rank models

$$\mathbf{Y} = \mathbf{\Theta} + \mathbf{E}$$

- $\mathbf{\Theta}$ contains the "main features" we hope to recover,
- $\mathbf{E}$ is "patternless."

Matrix decomposition: If Θ is a rank-R matrix, then

$$\theta_{i,j} = \langle \mathbf{u}_i, \mathbf{v}_j \rangle = \sum_{r=1}^{R} u_{i,r} v_{j,r} \qquad \mathbf{\Theta} = \sum_{r=1}^{R} \mathbf{u}_r \mathbf{v}_r^T = \sum_{r=1}^{R} \mathbf{u}_r \circ \mathbf{v}_r$$

Array decomposition: If Θ is a rank-R array, then

$$\theta_{i,j,k} = \langle \mathbf{u}_i, \mathbf{v}_j, \mathbf{w}_k \rangle = \sum_{r=1}^{R} u_{i,r} v_{j,r} w_{k,r} \qquad \mathbf{\Theta} = \sum_{r=1}^{R} \mathbf{u}_r \circ \mathbf{v}_r \circ \mathbf{w}_r$$

(Harshman[1970], Kruskal[1976,1977], Harshman and Lundy[1984], Kruskal[1989])

Some things you should know

1. Computing the rank
   - matrix: easy to do
   - array: no known algorithm
2. Possible rank
   - matrix: $R_{\max} = \min(m_1, m_2)$
   - array: $\max(m_1, m_2, m_3) \leq R_{\max} \leq \min(m_1 m_2, m_1 m_3, m_2 m_3)$
3. Probable rank
   - matrix: "almost all" matrices have full rank.
   - array: a nonzero fraction (w.r.t. Lebesgue measure) have less than full rank.
4. Least squares approximation
   - matrix: SVD of $\mathbf{Y}$ provides the rank R least-squares approximation to $\mathbf{\Theta}$.
   - array: iterative "least squares" methods, but solution may not exist (de Silva and Lim[2008])
5. Uniqueness
   - matrix: The representation $\mathbf{\Theta} = \langle \mathbf{U}, \mathbf{V} \rangle = \mathbf{U}\mathbf{V}^T$ is not unique.
   - array: The representation $\mathbf{\Theta} = \langle \mathbf{U}, \mathbf{V}, \mathbf{W} \rangle$ is essentially unique.

A model-based approach

For a K-way array $\mathbf{Y}$,

$$\begin{aligned}
\mathbf{Y} &= \mathbf{\Theta} + \mathbf{E} \\
\mathbf{\Theta} &= \sum_{r=1}^{R} \mathbf{u}_r^{(1)} \circ \cdots \circ \mathbf{u}_r^{(K)} \equiv \langle \mathbf{U}^{(1)}, \ldots, \mathbf{U}^{(K)} \rangle \\
\mathbf{u}_1^{(k)}, \ldots, \mathbf{u}_{m_k}^{(k)} &\overset{\text{iid}}{\sim} \text{multivariate normal}(\boldsymbol{\mu}_k, \boldsymbol{\Psi}_k),
\end{aligned}$$

with $\{\boldsymbol{\mu}_k, \boldsymbol{\Psi}_k, k = 1, \ldots, K\}$ to be estimated.

Some motivation:

- shrinkage: $\mathbf{\Theta}$ contains lots of parameters.
- hierarchical: covariance among columns of $\mathbf{U}^{(k)}$ is identifiable.
- estimation: $p(\mathbf{Y}|\mathbf{U}^{(1)}, \ldots, vU^{(K)})$ multimodal, MCMC “stochastic search”
- adaptability: incorporate reduced rank arrays as a model component
  - multilinear predictor in a GLM
  - multilinear effects for regression parameters

A model-based approach

For a K-way array $\mathbf{Y}$,

$$\begin{aligned} \mathbf{Y} &= \mathbf{\Theta} + \mathbf{E} \\ \mathbf{\Theta} &= \sum_{r=1}^{R} \mathbf{u}_r^{(1)} \circ \cdots \circ \mathbf{u}_r^{(K)} \equiv \langle \mathbf{U}^{(1)}, \ldots, \mathbf{U}^{(K)} \rangle \\ \mathbf{u}_1^{(k)}, \ldots, \mathbf{u}_{m_k}^{(k)} &\stackrel{\text{iid}}{\sim} \text{multivariate normal}(\boldsymbol{\mu}_k, \boldsymbol{\Psi}_k), \end{aligned}$$

with $\{\boldsymbol{\mu}_k, \boldsymbol{\Psi}_k, k = 1, \ldots, K\}$ to be estimated.

Some motivation:

- shrinkage: $\mathbf{\Theta}$ contains lots of parameters.
- hierarchical: covariance among columns of $\mathbf{U}^{(k)}$ is identifiable.
- estimation: $p(\mathbf{Y}|\mathbf{U}^{(1)}, \ldots, vU^{(K)})$ multimodal, MCMC "stochastic search"
- adaptability: incorporate reduced rank arrays as a model component
  - multilinear predictor in a GLM
  - multilinear effects for regression parameters

A model-based approach

For a K-way array $\mathbf{Y}$,

$$\begin{aligned} \mathbf{Y} &= \mathbf{\Theta} + \mathbf{E} \\ \mathbf{\Theta} &= \sum_{r=1}^{R} \mathbf{u}_r^{(1)} \circ \cdots \circ \mathbf{u}_r^{(K)} \equiv \langle \mathbf{U}^{(1)}, \ldots, \mathbf{U}^{(K)} \rangle \\ \mathbf{u}_1^{(k)}, \ldots, \mathbf{u}_{m_k}^{(k)} &\overset{\text{iid}}{\sim} \text{multivariate normal}(\boldsymbol{\mu}_k, \mathbf{\Psi}_k), \end{aligned}$$

with $\{\boldsymbol{\mu}_k, \mathbf{\Psi}_k, k = 1, \ldots, K\}$ to be estimated.

Some motivation:

- shrinkage: $\mathbf{\Theta}$ contains lots of parameters.
- hierarchical: covariance among columns of $\mathbf{U}^{(k)}$ is identifiable.
- estimation: $p(\mathbf{Y}|\mathbf{U}^{(1)}, \ldots, vU^{(K)})$ multimodal, MCMC "stochastic search"
- adaptability: incorporate reduced rank arrays as a model component
  - multilinear predictor in a GLM
  - multilinear effects for regression parameters

A model-based approach

For a K-way array $\mathbf{Y}$,

$$\begin{aligned}\mathbf{Y} &= \boldsymbol{\Theta} + \mathbf{E} \\ \boldsymbol{\Theta} &= \sum_{r=1}^{R} \mathbf{u}_r^{(1)} \circ \cdots \circ \mathbf{u}_r^{(K)} \equiv \langle \mathbf{U}^{(1)}, \ldots, \mathbf{U}^{(K)} \rangle \\ \mathbf{u}_1^{(k)}, \ldots, \mathbf{u}_{m_k}^{(k)} &\overset{\text{iid}}{\sim} \text{multivariate normal}(\boldsymbol{\mu}_k, \boldsymbol{\Psi}_k),\end{aligned}$$

with $\{\boldsymbol{\mu}_k, \boldsymbol{\Psi}_k, k = 1, \ldots, K\}$ to be estimated.

Some motivation:

- shrinkage: $\boldsymbol{\Theta}$ contains lots of parameters.
- hierarchical: covariance among columns of $\mathbf{U}^{(k)}$ is identifiable.
- estimation: $p(\mathbf{Y}|\mathbf{U}^{(1)}, \ldots, vU^{(K)})$ multimodal, MCMC "stochastic search"
- adaptability: incorporate reduced rank arrays as a model component
  - multilinear predictor in a GLM
  - multilinear effects for regression parameters

A model-based approach

For a K-way array $\mathbf{Y}$,

$$\begin{aligned}
\mathbf{Y} &= \mathbf{\Theta} + \mathbf{E} \\
\mathbf{\Theta} &= \sum_{r=1}^{R} \mathbf{u}_r^{(1)} \circ \cdots \circ \mathbf{u}_r^{(K)} \equiv \langle \mathbf{U}^{(1)}, \ldots, \mathbf{U}^{(K)} \rangle \\
\mathbf{u}_1^{(k)}, \ldots, \mathbf{u}_{m_k}^{(k)} &\stackrel{\text{iid}}{\sim} \text{multivariate normal}(\boldsymbol{\mu}_k, \mathbf{\Psi}_k),
\end{aligned}$$

with $\{\boldsymbol{\mu}_k, \mathbf{\Psi}_k, k = 1, \ldots, K\}$ to be estimated.

Some motivation:

- shrinkage: $\mathbf{\Theta}$ contains lots of parameters.
- hierarchical: covariance among columns of $\mathbf{U}^{(k)}$ is identifiable.
- estimation: $p(\mathbf{Y}|\mathbf{U}^{(1)}, \ldots, vU^{(K)})$ multimodal, MCMC "stochastic search"
- adaptability: incorporate reduced rank arrays as a model component
  - multilinear predictor in a GLM
  - multilinear effects for regression parameters

A model-based approach

For a K-way array $\mathbf{Y}$,

$$
\begin{aligned}
\mathbf{Y} &= \mathbf{\Theta} + \mathbf{E} \\
\mathbf{\Theta} &= \sum_{r=1}^{R} \mathbf{u}_r^{(1)} \circ \cdots \circ \mathbf{u}_r^{(K)} \equiv \langle \mathbf{U}^{(1)}, \ldots, \mathbf{U}^{(K)} \rangle \\
\mathbf{u}_1^{(k)}, \ldots, \mathbf{u}_{m_k}^{(k)} &\overset{\text{iid}}{\sim} \text{multivariate normal}(\boldsymbol{\mu}_k, \boldsymbol{\Psi}_k),
\end{aligned}
$$

with $\{\boldsymbol{\mu}_k, \boldsymbol{\Psi}_k, k = 1, \ldots, K\}$ to be estimated.

Some motivation:

- shrinkage: $\mathbf{\Theta}$ contains lots of parameters.
- hierarchical: covariance among columns of $\mathbf{U}^{(k)}$ is identifiable.
- estimation: $p(\mathbf{Y}|\mathbf{U}^{(1)}, \ldots, vU^{(K)})$ multimodal, MCMC "stochastic search"
- adaptability: incorporate reduced rank arrays as a model component
  - multilinear predictor in a GLM
  - multilinear effects for regression parameters

Longitudinal network example

- $y_{i,j,t} \in \{-5, -4, \ldots, +1, +2\}$, the level of military conflict/cooperation
- $x_{i,j,t,1} = \log \mathrm{gdp}_i + \log \mathrm{gdp}_j$, the sum of the log gdps of the two countries;
- $x_{i,j,t,2} = (\log \mathrm{gdp}_i) \times (\log \mathrm{gdp}_j)$, the product of the log gdps;
- $x_{i,j,t,3} = \mathrm{polity}_i \times \mathrm{polity}_j$, where $\mathrm{polity}_i \in \{-1, 0, +1\}$;
- $x_{i,j,t,4} = (\mathrm{polity}_i > 0) \times (\mathrm{polity}_j > 0)$.

Model:

$$
\begin{aligned}
y_{i,j,t} &= \max\{y : z_{i,j,t} > c_y\} \\
z_{i,j,t} &= \beta^T \mathbf{x}_{i,j,t} + \langle \mathbf{u}_i, \mathbf{u}_j, \boldsymbol{\lambda}_t \rangle + \epsilon_{i,j,t} \\
\mathbf{u}_1, \ldots, \mathbf{u}_n &\sim \text{iid } p(\mathbf{u})
\end{aligned}
$$

“Interpretation”:

$$
\mathbf{Z}_t = \mathbf{U}\Lambda_t\mathbf{U}^T + \mathbf{E}_t
$$

Longitudinal network example

- $y_{i,j,t} \in \{-5, -4, \ldots, +1, +2\}$, the level of military conflict/cooperation
- $x_{i,j,t,1} = \log \mathrm{gdp}_i + \log \mathrm{gdp}_j$, the sum of the log gdps of the two countries;
- $x_{i,j,t,2} = (\log \mathrm{gdp}_i) \times (\log \mathrm{gdp}_j)$, the product of the log gdps;
- $x_{i,j,t,3} = \mathrm{polity}_i \times \mathrm{polity}_j$, where $\mathrm{polity}_i \in \{-1, 0, +1\}$;
- $x_{i,j,t,4} = (\mathrm{polity}_i > 0) \times (\mathrm{polity}_j > 0)$.

Model:

$$
\begin{aligned}
y_{i,j,t} &= \max\{y : z_{i,j,t} > c_y\} \\
z_{i,j,t} &= \boldsymbol{\beta}^T \mathbf{x}_{i,j,t} + \langle \mathbf{u}_i, \mathbf{u}_j, \boldsymbol{\lambda}_t \rangle + \epsilon_{i,j,t} \\
\mathbf{u}_1, \ldots, \mathbf{u}_n &\sim \text{iid } p(\mathbf{u})
\end{aligned}
$$

"Interpretation":

$$\mathbf{Z}_t = \mathbf{U} \Lambda_t \mathbf{U}^T + \mathbf{E}_t$$

Longitudinal network example

- $y_{i,j,t} \in \{-5, -4, \ldots, +1, +2\}$, the level of military conflict/cooperation
- $x_{i,j,t,1} = \log \mathrm{gdp}_i + \log \mathrm{gdp}_j$, the sum of the log gdps of the two countries;
- $x_{i,j,t,2} = (\log \mathrm{gdp}_i) \times (\log \mathrm{gdp}_j)$, the product of the log gdps;
- $x_{i,j,t,3} = \mathrm{polity}_i \times \mathrm{polity}_j$, where $\mathrm{polity}_i \in \{-1, 0, +1\}$;
- $x_{i,j,t,4} = (\mathrm{polity}_i > 0) \times (\mathrm{polity}_j > 0)$.

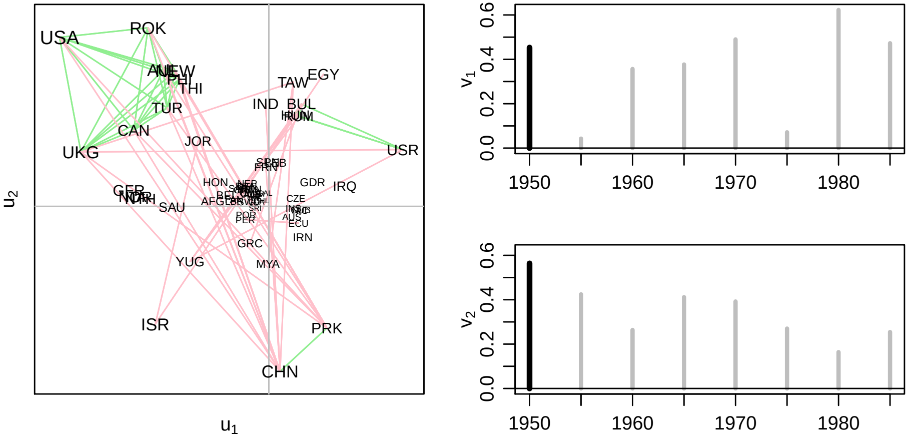
Model:

$$
\begin{aligned}
y_{i,j,t} &= \max\{y : z_{i,j,t} > c_y\} \\
z_{i,j,t} &= \boldsymbol{\beta}^T \mathbf{x}_{i,j,t} + \langle \mathbf{u}_i, \mathbf{u}_j, \boldsymbol{\lambda}_t \rangle + \epsilon_{i,j,t} \\
\mathbf{u}_1, \ldots, \mathbf{u}_n &\sim \text{iid } p(\mathbf{u})
\end{aligned}
$$

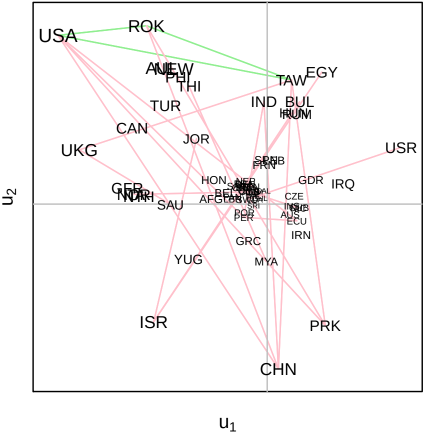
"Interpretation":

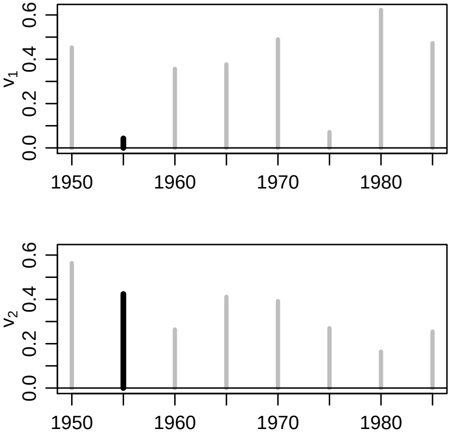
$$\mathbf{Z}_t = \mathbf{U} \Lambda_t \mathbf{U}^T + \mathbf{E}_t$$

Longitudinal network example

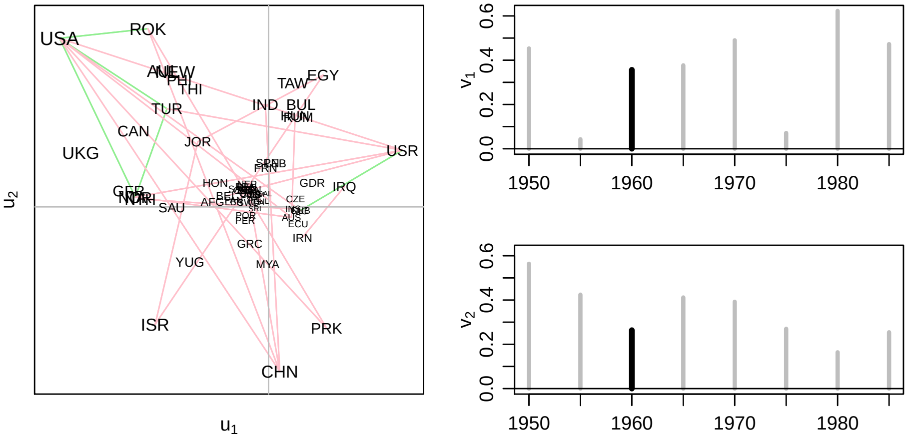


Longitudinal network example

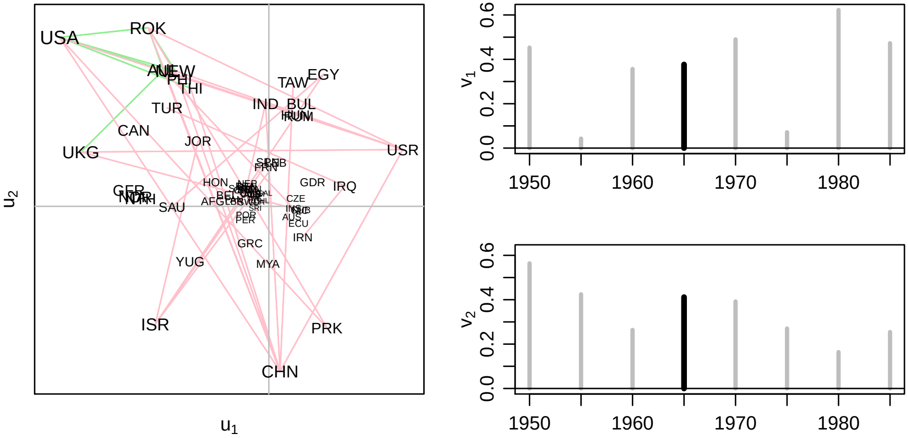




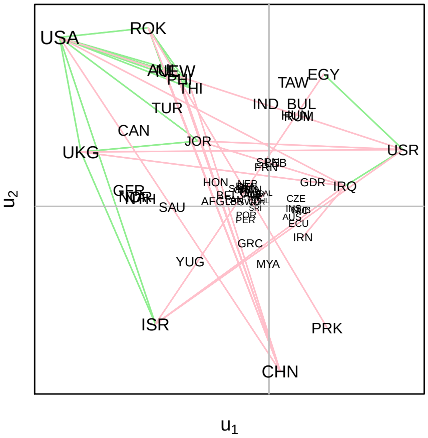
Longitudinal network example

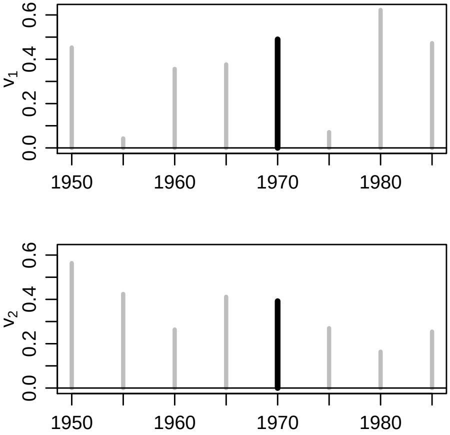


Longitudinal network example

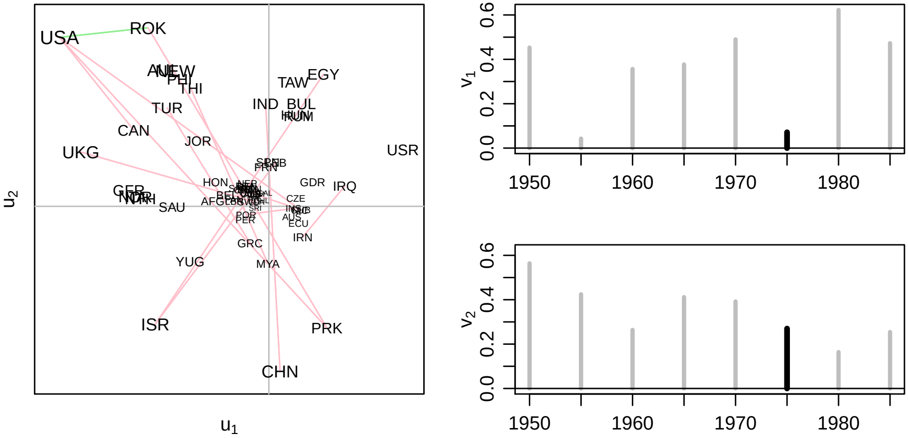


Longitudinal network example

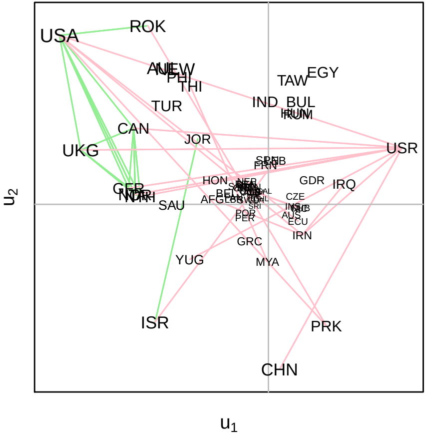


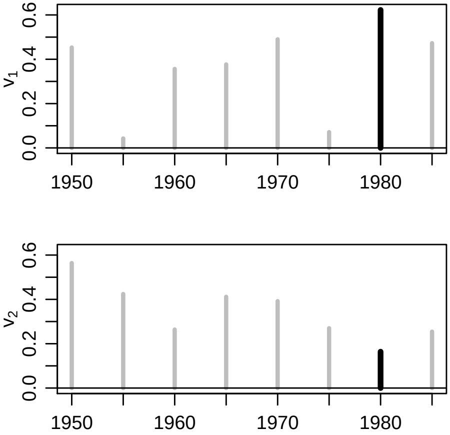


Longitudinal network example

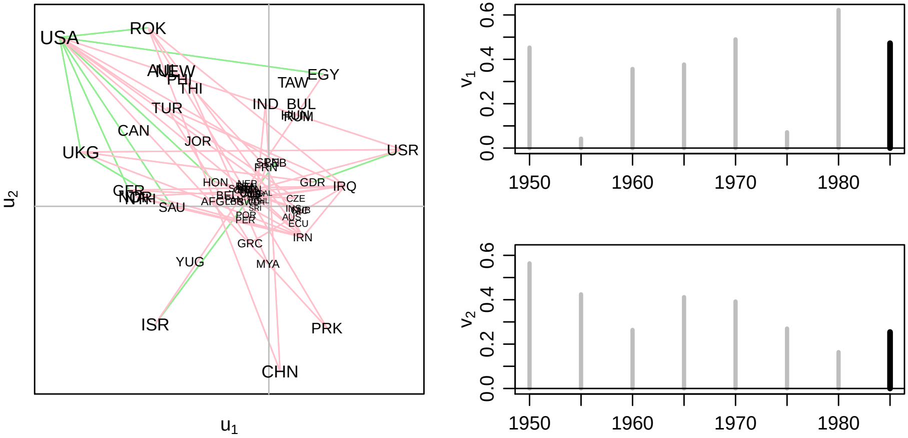


Longitudinal network example





Longitudinal network example



Covariance structure of multiple relational arrays

Yearly change in log exports (2000 dollars) : $\mathbf{Y} = \{y_{i,j,k,l}\} \in \mathbb{R}^{30\times30\times6\times10}$

- $i \in \{1,\ldots,30\}$ indexes exporting nation
- $j \in \{1,\ldots,30\}$ indexes importing nation
- $k \in \{1,\ldots,6\}$ indexes commodity
- $l \in \{1,\ldots,10\}$ indexes year

"Replications" over time: $\mathbf{Y} = \{\mathbf{Y}_1,\ldots,\mathbf{Y}_{10}\}$

$$\mathbf{Y}_t = \mathbf{M} + \mathbf{E}_t$$

- $\mathbf{M} \in \mathbb{R}^{30\times30\times6}$, constant over time;
- $\mathbf{E}_t \in \mathbb{R}^{30\times30\times6}$, changing over time.

How should the covariance among $\{\mathbf{E}_1,\ldots,\mathbf{E}_{10}\}$ be described?

Covariance structure of multiple relational arrays

Yearly change in log exports (2000 dollars) : $\mathbf{Y} = \{y_{i,j,k,l}\} \in \mathbb{R}^{30\times30\times6\times10}$

- $i \in \{1, \ldots, 30\}$ indexes exporting nation
- $j \in \{1, \ldots, 30\}$ indexes importing nation
- $k \in \{1, \ldots, 6\}$ indexes commodity
- $l \in \{1, \ldots, 10\}$ indexes year

"Replications" over time: $\mathbf{Y} = \{\mathbf{Y}_1, \ldots, \mathbf{Y}_{10}\}$

$$\mathbf{Y}_t = \mathbf{M} + \mathbf{E}_t$$

- $\mathbf{M} \in \mathbb{R}^{30\times30\times6}$, constant over time;
- $\mathbf{E}_t \in \mathbb{R}^{30\times30\times6}$, changing over time.

How should the covariance among $\{\mathbf{E}_1, \ldots, \mathbf{E}_{10}\}$ be described?

Covariance structure of multiple relational arrays

Yearly change in log exports (2000 dollars) : $\mathbf{Y} = \{y_{i,j,k,l}\} \in \mathbb{R}^{30\times 30\times 6\times 10}$

- $i \in \{1,\ldots,30\}$ indexes exporting nation
- $j \in \{1,\ldots,30\}$ indexes importing nation
- $k \in \{1,\ldots,6\}$ indexes commodity
- $l \in \{1,\ldots,10\}$ indexes year

"Replications" over time: $\mathbf{Y} = \{\mathbf{Y}_1,\ldots,\mathbf{Y}_{10}\}$

$$\mathbf{Y}_t = \mathbf{M} + \mathbf{E}_t$$

- $\mathbf{M} \in \mathbb{R}^{30\times 30\times 6}$, constant over time;
- $\mathbf{E}_t \in \mathbb{R}^{30\times 30\times 6}$, changing over time.

How should the covariance among $\{\mathbf{E}_1,\ldots,\mathbf{E}_{10}\}$ be described?

Covariance structure of multiple relational arrays

Yearly change in log exports (2000 dollars) : $\mathbf{Y} = \{y_{i,j,k,l}\} \in \mathbb{R}^{30\times30\times6\times10}$

- $i \in \{1,\ldots,30\}$ indexes exporting nation
- $j \in \{1,\ldots,30\}$ indexes importing nation
- $k \in \{1,\ldots,6\}$ indexes commodity
- $l \in \{1,\ldots,10\}$ indexes year

"Replications" over time: $\mathbf{Y} = \{\mathbf{Y}_1,\ldots,\mathbf{Y}_{10}\}$

$$\mathbf{Y}_t = \mathbf{M} + \mathbf{E}_t$$

- $\mathbf{M} \in \mathbb{R}^{30\times30\times6}$, constant over time;
- $\mathbf{E}_t \in \mathbb{R}^{30\times30\times6}$, changing over time.

How should the covariance among $\{\mathbf{E}_1,\ldots,\mathbf{E}_{10}\}$ be described?

Covariance structure of multiple relational arrays

Yearly change in log exports (2000 dollars) : $\mathbf{Y} = \{y_{i,j,k,l}\} \in \mathbb{R}^{30\times 30\times 6\times 10}$

- $i \in \{1, \ldots, 30\}$ indexes exporting nation
- $j \in \{1, \ldots, 30\}$ indexes importing nation
- $k \in \{1, \ldots, 6\}$ indexes commodity
- $l \in \{1, \ldots, 10\}$ indexes year

"Replications" over time: $\mathbf{Y} = \{\mathbf{Y}_1, \ldots, \mathbf{Y}_{10}\}$

$$\mathbf{Y}_t = \mathbf{M} + \mathbf{E}_t$$

- $\mathbf{M} \in \mathbb{R}^{30\times 30\times 6}$, constant over time;
- $\mathbf{E}_t \in \mathbb{R}^{30\times 30\times 6}$, changing over time.

How should the covariance among $\{\mathbf{E}_1, \ldots, \mathbf{E}_{10}\}$ be described?

Covariance structure of multiple relational arrays

Yearly change in log exports (2000 dollars) : $\mathbf{Y} = \{y_{i,j,k,l}\} \in \mathbb{R}^{30\times 30\times 6\times 10}$

- $i \in \{1, \dots, 30\}$ indexes exporting nation
- $j \in \{1, \dots, 30\}$ indexes importing nation
- $k \in \{1, \dots, 6\}$ indexes commodity
- $l \in \{1, \dots, 10\}$ indexes year

"Replications" over time: $\mathbf{Y} = \{\mathbf{Y}_1, \dots, \mathbf{Y}_{10}\}$

$$\mathbf{Y}_t = \mathbf{M} + \mathbf{E}_t$$

- $\mathbf{M} \in \mathbb{R}^{30\times 30\times 6}$, constant over time;
- $\mathbf{E}_t \in \mathbb{R}^{30\times 30\times 6}$, changing over time.

How should the covariance among $\{\mathbf{E}_1, \dots, \mathbf{E}_{10}\}$ be described?

The Tucker product

$$\mathbf{Y} = \mathbf{\Theta} + \mathbf{E}$$

Decompose $\mathbf{\Theta}$ using the Tucker decomposition (Tucker 1964,1966):

$$\theta_{i,j,k} = \sum_{r=1}^{R}\sum_{s=1}^{S}\sum_{t=1}^{T} z_{r,s,t}\, a_{i,r}\, b_{j,r}\, c_{k,r}$$

$$\mathbf{\Theta} = \mathbf{Z} \times \{\mathbf{A}, \mathbf{B}, \mathbf{C}\}$$

- $\mathbf{Z}$ is the $R \times S \times T$ *core array*
- $\mathbf{A}$, $\mathbf{B}$, $\mathbf{C}$ are $R \times m_1$, $S \times m_2$, $T \times m_3$ matrices.
- R, S and T are the 1-rank, 2-rank and 3-rank of $\mathbf{\Theta}$
- "$\times$" is array-matrix multiplication (De Lathauwer et al., 2000)

The Tucker product

$$\mathbf{Y} = \mathbf{\Theta} + \mathbf{E}$$

Decompose $\mathbf{\Theta}$ using the Tucker decomposition (Tucker 1964,1966):

$$\theta_{i,j,k} = \sum_{r=1}^{R}\sum_{s=1}^{S}\sum_{t=1}^{T} z_{r,s,t}\, a_{i,r} b_{j,r} c_{k,r}$$

$$\mathbf{\Theta} = \mathbf{Z} \times \{\mathbf{A}, \mathbf{B}, \mathbf{C}\}$$

- $\mathbf{Z}$ is the $R \times S \times T$ *core array*
- $\mathbf{A}$, $\mathbf{B}$, $\mathbf{C}$ are $R \times m_1$, $S \times m_2$, $T \times m_3$ matrices.
- R, S and T are the 1-rank, 2-rank and 3-rank of $\mathbf{\Theta}$
- "$\times$" is array-matrix multiplication (De Lathauwer et al., 2000)

The Tucker product

$$\mathbf{Y} = \mathbf{\Theta} + \mathbf{E}$$

Decompose $\mathbf{\Theta}$ using the Tucker decomposition (Tucker 1964,1966):

$$\begin{aligned} \theta_{i,j,k} &= \sum_{r=1}^{R} \sum_{s=1}^{S} \sum_{t=1}^{T} z_{r,s,t} a_{i,r} b_{j,r} c_{k,r} \\ \mathbf{\Theta} &= \mathbf{Z} \times \{\mathbf{A}, \mathbf{B}, \mathbf{C}\} \end{aligned}$$

- $\mathbf{Z}$ is the $R \times S \times T$ *core array*
- $\mathbf{A}$, $\mathbf{B}$, $\mathbf{C}$ are $R \times m_1$, $S \times m_2$, $T \times m_3$ matrices.
- R, S and T are the 1-rank, 2-rank and 3-rank of $\mathbf{\Theta}$
- "$\times$" is array-matrix multiplication (De Lathauwer et al., 2000)

The Tucker product

$$\mathbf{Y} = \mathbf{\Theta} + \mathbf{E}$$

Decompose $\mathbf{\Theta}$ using the Tucker decomposition (Tucker 1964,1966):

$$\begin{aligned} \theta_{i,j,k} &= \sum_{r=1}^{R} \sum_{s=1}^{S} \sum_{t=1}^{T} z_{r,s,t} a_{i,r} b_{j,r} c_{k,r} \\ \mathbf{\Theta} &= \mathbf{Z} \times \{\mathbf{A}, \mathbf{B}, \mathbf{C}\} \end{aligned}$$

- $\mathbf{Z}$ is the $R \times S \times T$ *core array*
- $\mathbf{A}$, $\mathbf{B}$, $\mathbf{C}$ are $R \times m_1$, $S \times m_2$, $T \times m_3$ matrices.
- R, S and T are the 1-rank, 2-rank and 3-rank of $\mathbf{\Theta}$
- "$\times$" is array-matrix multiplication (De Lathauwer et al., 2000)

The Tucker product

$$\mathbf{Y} = \mathbf{\Theta} + \mathbf{E}$$

Decompose $\mathbf{\Theta}$ using the Tucker decomposition (Tucker 1964,1966):

$$\theta_{i,j,k} = \sum_{r=1}^{R} \sum_{s=1}^{S} \sum_{t=1}^{T} z_{r,s,t} a_{i,r} b_{j,r} c_{k,r}$$

$$\mathbf{\Theta} = \mathbf{Z} \times \{\mathbf{A}, \mathbf{B}, \mathbf{C}\}$$

- $\mathbf{Z}$ is the $R \times S \times T$ *core array*
- $\mathbf{A}$, $\mathbf{B}$, $\mathbf{C}$ are $R \times m_1$, $S \times m_2$, $T \times m_3$ matrices.
- R, S and T are the 1-rank, 2-rank and 3-rank of $\mathbf{\Theta}$
- "$\times$" is array-matrix multiplication (De Lathauwer et al., 2000)

The Tucker product

$$\mathbf{Y} = \boldsymbol{\Theta} + \mathbf{E}$$

Decompose $\boldsymbol{\Theta}$ using the Tucker decomposition (Tucker 1964,1966):

$$\theta_{i,j,k} = \sum_{r=1}^{R} \sum_{s=1}^{S} \sum_{t=1}^{T} z_{r,s,t} a_{i,r} b_{j,r} c_{k,r}$$

$$\boldsymbol{\Theta} = \mathbf{Z} \times \{\mathbf{A}, \mathbf{B}, \mathbf{C}\}$$

- $\mathbf{Z}$ is the $R \times S \times T$ *core array*
- $\mathbf{A}$, $\mathbf{B}$, $\mathbf{C}$ are $R \times m_1$, $S \times m_2$, $T \times m_3$ matrices.
- R, S and T are the 1-rank, 2-rank and 3-rank of $\boldsymbol{\Theta}$
- "$\times$" is array-matrix multiplication (De Lathauwer et al., 2000)

Separable covariance via Tucker products

Multivariate normal model:

$$\mathbf{z} = \{z_j : j = 1, \ldots, m\} \stackrel{iid}{\sim} \text{normal}(\mathbf{0}, 1)$$
$$\mathbf{y} = \boldsymbol{\mu} + \mathbf{A}\mathbf{z} \sim \text{multivariate normal}(\boldsymbol{\mu}, \boldsymbol{\Sigma} = \mathbf{A}\mathbf{A}^T)$$

Matrix normal model:

$$\mathbf{Z} = \{z_{i,j}\}_{i=1,j=1}^{m_1,m_2} \stackrel{iid}{\sim} \text{normal}(\mathbf{0}, 1)$$
$$\mathbf{Y} = \mathbf{M} + \mathbf{A}\mathbf{Z}\mathbf{B}^T \sim \text{matrix normal}(\mathbf{M}, \boldsymbol{\Sigma}_1 = \mathbf{A}\mathbf{A}^T, \Sigma_2 = \mathbf{B}\mathbf{B}^T)$$

NOTE: $\mathbf{A}\mathbf{Z}\mathbf{B}^T = \mathbf{Z} \times \{\mathbf{A}, \mathbf{B}\}$

Array normal model:

$$\mathbf{Z} = \{z_{i,j,k}\}_{i=1,j=1,k=1}^{m_1,m_2,m_3} \stackrel{iid}{\sim} \text{normal}(\mathbf{0}, 1)$$
$$\mathbf{Y} = \mathbf{M} + \mathbf{Z} \times \{\mathbf{A}, \mathbf{B}, \mathbf{C}\} \sim \text{array normal}(\mathbf{M}, \boldsymbol{\Sigma}_1 = \mathbf{A}\mathbf{A}^T, \boldsymbol{\Sigma}_2 = \mathbf{B}\mathbf{B}^T, \boldsymbol{\Sigma}_3 = \mathbf{C}\mathbf{C}^T)$$

Separable covariance via Tucker products

Multivariate normal model:

$$\mathbf{z} = \{z_j : j = 1, \ldots, m\} \overset{iid}{\sim} \text{normal}(\mathbf{0}, 1)$$
$$\mathbf{y} = \boldsymbol{\mu} + \mathbf{Az} \sim \text{multivariate normal}(\boldsymbol{\mu}, \boldsymbol{\Sigma} = \mathbf{AA}^T)$$

Matrix normal model:

$$\mathbf{Z} = \{z_{i,j}\}_{i=1,j=1}^{m_1,m_2} \overset{iid}{\sim} \text{normal}(\mathbf{0}, 1)$$
$$\mathbf{Y} = \mathbf{M} + \mathbf{AZB}^T \sim \text{matrix normal}(\mathbf{M}, \boldsymbol{\Sigma}_1 = \mathbf{AA}^T, \Sigma_2 = \mathbf{BB}^T)$$

NOTE: $\mathbf{AZB}^T = \mathbf{Z} \times \{\mathbf{A}, \mathbf{B}\}$

Array normal model:

$$\mathbf{Z} = \{z_{i,j,k}\}_{i=1,j=1,k=1}^{m_1,m_2,m_3} \overset{iid}{\sim} \text{normal}(\mathbf{0}, 1)$$
$$\mathbf{Y} = \mathbf{M} + \mathbf{Z} \times \{\mathbf{A}, \mathbf{B}, \mathbf{C}\} \sim \text{array normal}(\mathbf{M}, \boldsymbol{\Sigma}_1 = \mathbf{AA}^T, \boldsymbol{\Sigma}_2 = \mathbf{BB}^T, \boldsymbol{\Sigma}_3 = \mathbf{CC}^T)$$

Separable covariance via Tucker products

Multivariate normal model:

$$\begin{aligned} \mathbf{z} = \{z_j : j = 1, \ldots, m\} &\overset{iid}{\sim} \text{normal}(\mathbf{0}, 1) \\ \mathbf{y} = \boldsymbol{\mu} + \mathbf{A}\mathbf{z} &\sim \text{multivariate normal}(\boldsymbol{\mu}, \boldsymbol{\Sigma} = \mathbf{A}\mathbf{A}^T) \end{aligned}$$

Matrix normal model:

$$\begin{aligned} \mathbf{Z} = \{z_{i,j}\}_{i=1,j=1}^{m_1,m_2} &\overset{iid}{\sim} \text{normal}(\mathbf{0}, 1) \\ \mathbf{Y} = \mathbf{M} + \mathbf{A}\mathbf{Z}\mathbf{B}^T &\sim \text{matrix normal}(\mathbf{M}, \boldsymbol{\Sigma}_1 = \mathbf{A}\mathbf{A}^T, \Sigma_2 = \mathbf{B}\mathbf{B}^T) \end{aligned}$$

NOTE: $\mathbf{A}\mathbf{Z}\mathbf{B}^T = \mathbf{Z} \times \{\mathbf{A}, \mathbf{B}\}$

Array normal model:

$$\begin{aligned} \mathbf{Z} = \{z_{i,j,k}\}_{i=1,j=1,k=1}^{m_1,m_2,m_3} &\overset{iid}{\sim} \text{normal}(\mathbf{0}, 1) \\ \mathbf{Y} = \mathbf{M} + \mathbf{Z} \times \{\mathbf{A}, \mathbf{B}, \mathbf{C}\} &\sim \text{array normal}(\mathbf{M}, \boldsymbol{\Sigma}_1 = \mathbf{A}\mathbf{A}^T, \boldsymbol{\Sigma}_2 = \mathbf{B}\mathbf{B}^T, \boldsymbol{\Sigma}_3 = \mathbf{C}\mathbf{C}^T) \end{aligned}$$

Separable covariance structure

For the matrix normal model:

$$\begin{aligned}
\mathrm{Cov}[\mathbf{Y}] &= \mathbf{\Sigma}_1 \circ \mathbf{\Sigma}_2 \\
\mathrm{Cov}[\mathrm{vec}(\mathbf{Y})] &= \mathbf{\Sigma}_2 \otimes \mathbf{\Sigma}_1 \\
\mathrm{E}[\mathbf{Y}\mathbf{Y}^T] &= \mathbf{\Sigma}_1 \times \mathrm{tr}(\mathbf{\Sigma}_2) \\
\mathrm{E}[\mathbf{Y}^T\mathbf{Y}] &= \mathbf{\Sigma}_2 \times \mathrm{tr}(\mathbf{\Sigma}_1)
\end{aligned}$$

For the array normal model:

$$\begin{aligned}
\mathrm{Cov}[\mathbf{Y}] &= \mathbf{\Sigma}_1 \circ \mathbf{\Sigma}_2 \circ \mathbf{\Sigma}_3 \\
\mathrm{Cov}[\mathrm{vec}(\mathbf{Y})] &= \mathbf{\Sigma}_K \otimes \cdots \otimes \mathbf{\Sigma}_1 \\
\mathrm{E}[\mathbf{Y}_{(k)}\mathbf{Y}_{(k)}^T] &= \mathbf{\Sigma}_k \times \prod_{j \neq k} \mathrm{tr}(\mathbf{\Sigma}_j)
\end{aligned}$$

International trade example

Yearly change in log exports (2000 dollars) : $\mathbf{Y} = \{y_{i,j,k,l}\} \in \mathbb{R}^{30\times30\times6\times7}$

- $i \in \{1,\ldots,30\}$ indexes exporting nation
- $j \in \{1,\ldots,30\}$ indexes importing nation
- $k \in \{1,\ldots,6\}$ indexes commodity
- $l \in \{1,\ldots,10\}$ indexes year

Full "cell means" model:

$$y_{i,j,k,l} = \mu_{i,j,k} + e_{i,j,k,l}$$

Let $\mathbf{E} = \{e_{i,j,k,l}\}$

- iid error model: $\mathbf{E} \sim$ array normal$(0,\mathbf{I},\mathbf{I},\mathbf{I},\sigma^2\mathbf{I})$
- vector normal error model: $\mathbf{E} \sim$ array normal$(0,\mathbf{I},\mathbf{I},\boldsymbol{\Sigma}_3,\mathbf{I})$
- matrix normal error model: $\mathbf{E} \sim$ array normal$(0,\mathbf{I},\mathbf{I},\boldsymbol{\Sigma}_3,\boldsymbol{\Sigma}_4)$
- array normal model: $\mathbf{E} \sim$ array normal$(\mathbf{0},\boldsymbol{\Sigma}_1,\boldsymbol{\Sigma}_2,\boldsymbol{\Sigma}_3,\boldsymbol{\Sigma}_4\}$

International trade example

Yearly change in log exports (2000 dollars) : $\mathbf{Y} = \{y_{i,j,k,l}\} \in \mathbb{R}^{30\times 30\times 6\times 7}$

- $i \in \{1, \ldots, 30\}$ indexes exporting nation
- $j \in \{1, \ldots, 30\}$ indexes importing nation
- $k \in \{1, \ldots, 6\}$ indexes commodity
- $l \in \{1, \ldots, 10\}$ indexes year

Full "cell means" model:

$$y_{i,j,k,l} = \mu_{i,j,k} + e_{i,j,k,l}$$

Let $\mathbf{E} = \{e_{i,j,k,l}\}$

- iid error model: $\mathbf{E} \sim \text{array normal}(0, \mathbf{I}, \mathbf{I}, \mathbf{I}, \sigma^2\mathbf{I})$
- vector normal error model: $\mathbf{E} \sim \text{array normal}(0, \mathbf{I}, \mathbf{I}, \mathbf{\Sigma}_3, \mathbf{I})$
- matrix normal error model: $\mathbf{E} \sim \text{array normal}(0, \mathbf{I}, \mathbf{I}, \mathbf{\Sigma}_3, \mathbf{\Sigma}_4)$
- array normal model: $\mathbf{E} \sim \text{array normal}(\mathbf{0}, \mathbf{\Sigma}_1, \mathbf{\Sigma}_2, \mathbf{\Sigma}_3, \mathbf{\Sigma}_4\}$

International trade example

Yearly change in log exports (2000 dollars) : $\mathbf{Y} = \{y_{i,j,k,l}\} \in \mathbb{R}^{30\times30\times6\times7}$

- $i \in \{1,\dots,30\}$ indexes exporting nation
- $j \in \{1,\dots,30\}$ indexes importing nation
- $k \in \{1,\dots,6\}$ indexes commodity
- $l \in \{1,\dots,10\}$ indexes year

Full "cell means" model:

$$y_{i,j,k,l} = \mu_{i,j,k} + e_{i,j,k,l}$$

Let $\mathbf{E} = \{e_{i,j,k,l}\}$

- iid error model: $\mathbf{E} \sim$ array normal$(0, \mathbf{I}, \mathbf{I}, \mathbf{I}, \sigma^2\mathbf{I})$
- vector normal error model: $\mathbf{E} \sim$ array normal$(0, \mathbf{I}, \mathbf{I}, \boldsymbol{\Sigma}_3, \mathbf{I})$
- matrix normal error model: $\mathbf{E} \sim$ array normal$(0, \mathbf{I}, \mathbf{I}, \boldsymbol{\Sigma}_3, \boldsymbol{\Sigma}_4)$
- array normal model: $\mathbf{E} \sim$ array normal$(\mathbf{0}, \boldsymbol{\Sigma}_1, \boldsymbol{\Sigma}_2, \boldsymbol{\Sigma}_3, \boldsymbol{\Sigma}_4\}$

International trade example

Yearly change in log exports (2000 dollars) : $\mathbf{Y} = \{y_{i,j,k,l}\} \in \mathbb{R}^{30\times 30\times 6\times 7}$

- $i \in \{1, \ldots, 30\}$ indexes exporting nation
- $j \in \{1, \ldots, 30\}$ indexes importing nation
- $k \in \{1, \ldots, 6\}$ indexes commodity
- $l \in \{1, \ldots, 10\}$ indexes year

Full "cell means" model:

$$y_{i,j,k,l} = \mu_{i,j,k} + e_{i,j,k,l}$$

Let $\mathbf{E} = \{e_{i,j,k,l}\}$

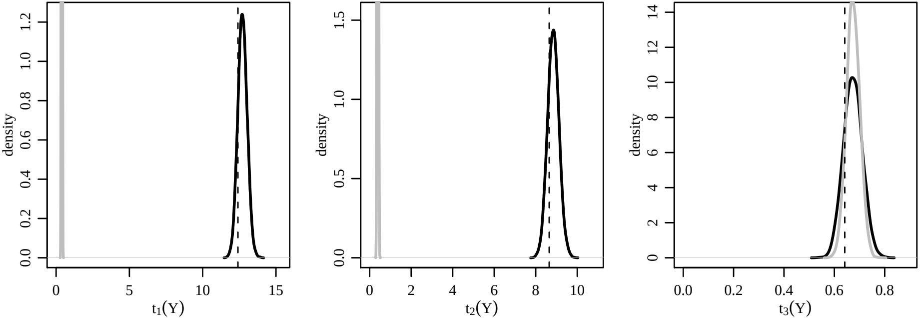
- iid error model: $\mathbf{E} \sim \text{array normal}(0, \mathbf{I}, \mathbf{I}, \mathbf{I}, \sigma^2\mathbf{I})$
- vector normal error model: $\mathbf{E} \sim \text{array normal}(0, \mathbf{I}, \mathbf{I}, \mathbf{\Sigma}_3, \mathbf{I})$
- matrix normal error model: $\mathbf{E} \sim \text{array normal}(0, \mathbf{I}, \mathbf{I}, \mathbf{\Sigma}_3, \mathbf{\Sigma}_4)$
- array normal model: $\mathbf{E} \sim \text{array normal}(\mathbf{0}, \mathbf{\Sigma}_1, \mathbf{\Sigma}_2, \mathbf{\Sigma}_3, \mathbf{\Sigma}_4\}$

International trade example

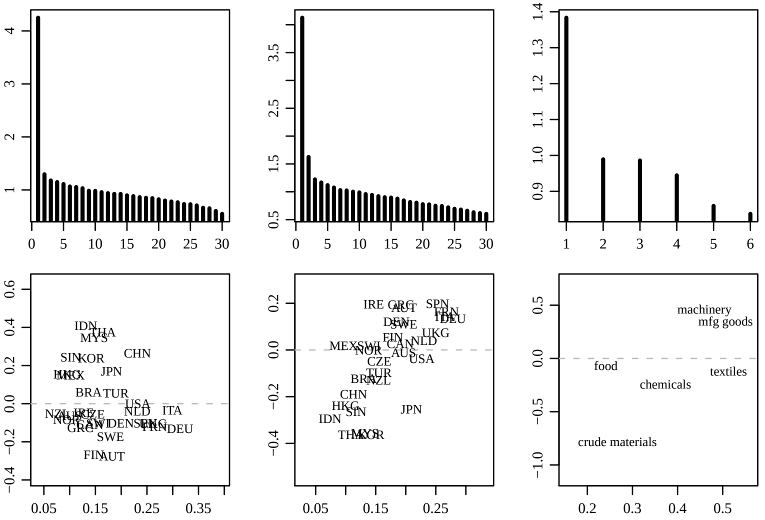
Model comparison:

reduced: array normal$(0, \mathbf{I}, \mathbf{I}, \mathbf{\Sigma}_3, \mathbf{\Sigma}_4)$

full: array normal$(0, \mathbf{\Sigma}_1, \mathbf{\Sigma}_2, \mathbf{\Sigma}_3, \mathbf{\Sigma}_4)$



International trade example



Summary

- Exchangeability implies a latent variable representation
- Matrix and array decompositions provide latent variable representations
- Lots of work to be done
    1. Theoretical: asymptotics, sampling frame, MDL
    2. Methodological: Rank selection, regularization
    3. Computational: MCMC, VB, other approximate solutions.

Summary

- Exchangeability implies a latent variable representation
- Matrix and array decompositions provide latent variable representations
- Lots of work to be done
  1. Theoretical: asymptotics, sampling frame, MDL
  2. Methodological: Rank selection, regularization
  3. Computational: MCMC, VB, other approximate solutions.

Summary

- Exchangeability implies a latent variable representation
- Matrix and array decompositions provide latent variable representations
- Lots of work to be done
  1. Theoretical: asymptotics, sampling frame, MDL
  2. Methodological: Rank selection, regularization
  3. Computational: MCMC, VB, other approximate solutions.

Summary

- Exchangeability implies a latent variable representation
- Matrix and array decompositions provide latent variable representations
- Lots of work to be done
  1. Theoretical: asymptotics, sampling frame, MDL
  2. Methodological: Rank selection, regularization
  3. Computational: MCMC, VB, other approximate solutions.

Summary

- Exchangeability implies a latent variable representation
- Matrix and array decompositions provide latent variable representations
- Lots of work to be done
  1. Theoretical: asymptotics, sampling frame, MDL
  2. Methodological: Rank selection, regularization
  3. Computational: MCMC, VB, other approximate solutions.

Summary

- Exchangeability implies a latent variable representation
- Matrix and array decompositions provide latent variable representations
- Lots of work to be done
  1. Theoretical: asymptotics, sampling frame, MDL
  2. Methodological: Rank selection, regularization
  3. Computational: MCMC, VB, other approximate solutions.

Summary

- Exchangeability implies a latent variable representation
- Matrix and array decompositions provide latent variable representations
- Lots of work to be done
  1. Theoretical: asymptotics, sampling frame, MDL
  2. Methodological: Rank selection, regularization
  3. Computational: MCMC, VB, other approximate solutions.