Introduction

567 Statistical analysis of social networks

Peter Hoff

Statistics, University of Washington
Networks are relational data

**Relationship**: An irreducible property of two or more entities.
- Contrast to properties of entities alone (*attributes*).
- Relations are possibly affected, but not determined, by entity attributes.

**Focus of RDA/SNA**: The study of relational data arising from social entities.
- **Entities**: people, animals, groups, locations, organizations, regions, etc..
- **Relationships**: communication, acquaintanceship, sexual contact, trade, migration rate, alliance/conflict, etc..
Networks are relational data

**Relationship**: An irreducible property of two or more entities.
- Contrast to properties of entities alone (attributes).
- Relations are possibly affected, but not determined, by entity attributes.

**Focus of RDA/SNA**: The study of relational data arising from social entities.
- **Entities**: people, animals, groups, locations, organizations, regions, etc..
- **Relationships**: communication, acquaintanceship, sexual contact, trade, migration rate, alliance/conflict, etc..
Networks are relational data

**Relationship**: An irreducible property of two or more entities.
- Contrast to properties of entities alone (**attributes**).
- Relations are possibly affected, but not determined, by entity attributes.

**Focus of RDA/SNA**: The study of relational data arising from social entities.
- **Entities**: people, animals, groups, locations, organizations, regions, etc..
- **Relationships**: communication, acquaintanceship, sexual contact, trade, migration rate, alliance/conflict, etc..
Networks are relational data

**Relationship:** An irreducible property of two or more entities.
- Contrast to properties of entities alone (attributes).
- Relations are possibly affected, but not determined, by entity attributes.

**Focus of RDA/SNA:** The study of relational data arising from social entities.
- **Entities:** people, animals, groups, locations, organizations, regions, etc..
- **Relationships:** communication, acquaintanceship, sexual contact, trade, migration rate, alliance/conflict, etc..
Networks are relational data

**Relationship**: An irreducible property of two or more entities.
- Contrast to properties of entities alone (**attributes**).
- Relations are possibly affected, but not determined, by entity attributes.

**Focus of RDA/SNA**: The study of relational data arising from social entities.
- **Entities**: people, animals, groups, locations, organizations, regions, etc..
- **Relationships**: communication, acquaintanceship, sexual contact, trade, migration rate, alliance/conflict, etc..
Relationship: An irreducible property of two or more entities.

- Contrast to properties of entities alone (attributes).
- Relations are possibly affected, but not determined, by entity attributes.

Focus of RDA/SNA: The study of relational data arising from social entities.

- **Entities**: people, animals, groups, locations, organizations, regions, etc..
- **Relationships**: communication, acquaintanceship, sexual contact, trade, migration rate, alliance/conflict, etc..
Relational data:
A collection of entities and a set of measured relations between them.

- Entities: nodes, actors, egos, units.
- Relations: ties, links, edges.

Relations can be
- directed or undirected;
- valued or dichotomous (binary).
Relational data:
A collection of entities and a set of measured relations between them.

- Entities: nodes, actors, egos, units.
- Relations: ties, links, edges.

Relations can be
- directed or undirected;
- valued or dichotomous (binary).
Some vocabulary

**Relational data:** A collection of entities and a set of measured relations between them.

- **Entities:** nodes, actors, egos, units.
- **Relations:** ties, links, edges.

Relations can be

- directed or undirected;
- valued or dichotomous (binary).
Some vocabulary

Relational data:
A collection of entities and a set of measured relations between them.

- Entities: nodes, actors, egos, units.
- Relations: ties, links, edges.

Relations can be
- directed or undirected;
- valued or dichotomous (binary).
Some vocabulary

**Relational data:**
A collection of entities and a set of measured relations between them.

- Entities: nodes, actors, egos, units.
- Relations: ties, links, edges.

Relations can be
- directed or undirected;
- valued or dichotomous (binary).
Relational data:
A collection of entities and a set of measured relations between them.

- Entities: nodes, actors, egos, units.
- Relations: ties, links, edges.

Relations can be
- directed or undirected;
- valued or dichotomous (binary).
Different perspectives on network analysis

- **Social sciences** *(social theory, description, survey design)*
- **Machine learning** *(clustering, prediction, computation)*
- **Physics and applied math** *(agent-based models, emergent features)*
- **Statistics** *(statistical modeling, estimation and testing, design based inference)*
Different perspectives on network analysis

- **Social sciences** (social theory, description, survey design)
- **Machine learning** (clustering, prediction, computation)
- **Physics and applied math** (agent-based models, emergent features)
- **Statistics** (statistical modeling, estimation and testing, design based inference)
Different perspectives on network analysis

- Social sciences (social theory, description, survey design)
- Machine learning (clustering, prediction, computation)
- Physics and applied math (agent-based models, emergent features)
- Statistics (statistical modeling, estimation and testing, design based inference)
Different perspectives on network analysis

- Social sciences (social theory, description, survey design)
- Machine learning (clustering, prediction, computation)
- Physics and applied math (agent-based models, emergent features)
- Statistics (statistical modeling, estimation and testing, design based inference)
Core areas of statistical network analysis

1. **Statistical modeling**: evaluation and fitting of network models.
   - Testing: evaluation of competing theories of network formation.
   - Estimation: evaluation of parameters in a presumed network model.
   - Description: summaries of main network patterns.
   - Prediction: prediction of missing or future network relations.

2. **Design-based inference**: Network inference from sampled data.
   - Design: survey and data-gathering procedures.
   - Inference: generalization of sample data to the full network.
Core areas of statistical network analysis

1. **Statistical modeling**: evaluation and fitting of network models.
   - Testing: evaluation of competing theories of network formation.
   - Estimation: evaluation of parameters in a presumed network model.
   - Description: summaries of main network patterns.
   - Prediction: prediction of missing or future network relations.

2. **Design-based inference**: Network inference from sampled data.
   - Design: survey and data-gathering procedures.
   - Inference: generalization of sample data to the full network.
1. **Statistical modeling**: evaluation and fitting of network models.
   - Testing: evaluation of competing theories of network formation.
   - Estimation: evaluation of parameters in a presumed network model.
   - Description: summaries of main network patterns.
   - Prediction: prediction of missing or future network relations.

2. **Design-based inference**: Network inference from sampled data.
   - Design: survey and data-gathering procedures.
   - Inference: generalization of sample data to the full network.
1. **Statistical modeling**: evaluation and fitting of network models.
   - Testing: evaluation of competing theories of network formation.
   - Estimation: evaluation of parameters in a presumed network model.
   - Description: summaries of main network patterns.
   - Prediction: prediction of missing or future network relations.

2. **Design-based inference**: Network inference from sampled data.
   - Design: survey and data-gathering procedures.
   - Inference: generalization of sample data to the full network.
Core areas of statistical network analysis

1. **Statistical modeling**: evaluation and fitting of network models.
   - Testing: evaluation of competing theories of network formation.
   - Estimation: evaluation of parameters in a presumed network model.
   - Description: summaries of main network patterns.
   - Prediction: prediction of missing or future network relations.

2. **Design-based inference**: Network inference from sampled data.
   - Design: survey and data-gathering procedures.
   - Inference: generalization of sample data to the full network.
Core areas of statistical network analysis

1. **Statistical modeling**: evaluation and fitting of network models.
   - Testing: evaluation of competing theories of network formation.
   - Estimation: evaluation of parameters in a presumed network model.
   - Description: summaries of main network patterns.
   - Prediction: prediction of missing or future network relations.

2. **Design-based inference**: Network inference from sampled data.
   - Design: survey and data-gathering procedures.
   - Inference: generalization of sample data to the full network.
Core areas of statistical network analysis

1. **Statistical modeling**: evaluation and fitting of network models.
   - Testing: evaluation of competing theories of network formation.
   - Estimation: evaluation of parameters in a presumed network model.
   - Description: summaries of main network patterns.
   - Prediction: prediction of missing or future network relations.

2. **Design-based inference**: Network inference from sampled data.
   - Design: survey and data-gathering procedures.
   - Inference: generalization of sample data to the full network.
Core areas of statistical network analysis

1. **Statistical modeling**: evaluation and fitting of network models.
   - Testing: evaluation of competing theories of network formation.
   - Estimation: evaluation of parameters in a presumed network model.
   - Description: summaries of main network patterns.
   - Prediction: prediction of missing or future network relations.

2. **Design-based inference**: Network inference from sampled data.
   - Design: survey and data-gathering procedures.
   - Inference: generalization of sample data to the full network.
1. **Statistical modeling**: evaluation and fitting of network models.
   - Testing: evaluation of competing theories of network formation.
   - Estimation: evaluation of parameters in a presumed network model.
   - Description: summaries of main network patterns.
   - Prediction: prediction of missing or future network relations.

2. **Design-based inference**: Network inference from sampled data.
   - Design: survey and data-gathering procedures.
   - Inference: generalization of sample data to the full network.
Core areas of statistical network analysis

1. **Statistical modeling**: evaluation and fitting of network models.
   - Testing: evaluation of competing theories of network formation.
   - Estimation: evaluation of parameters in a presumed network model.
   - Description: summaries of main network patterns.
   - Prediction: prediction of missing or future network relations.

2. **Design-based inference**: Network inference from sampled data.
   - Design: survey and data-gathering procedures.
   - Inference: generalization of sample data to the full network.
Example: AddHealth friendships

- “Add Health” - The National Longitudinal Study of Adolescent Health:
  - A school-based study of adolescent health and social behaviors;
  - www.cpc.unc.edu/projects/addhealth.

- Data from 160 schools across the US:
  - The smallest had 69 adolescents in grades 7–12;
  - The largest had thousands of participants.

- Relational data:
  - Participants nominated and ranked up to 5 boys and 5 girls as friends.
Example: AddHealth friendships

- “Add Health” - The National Longitudinal Study of Adolescent Health:
  - A school-based study of adolescent health and social behaviors;
  - www.cpc.unc.edu/projects/addhealth.

- Data from 160 schools across the US:
  - The smallest had 69 adolescents in grades 7–12;
  - The largest had thousands of participants.

- Relational data:
  - Participants nominated and ranked up to 5 boys and 5 girls as friends.
Example: AddHealth friendships

- “Add Health” - The National Longitudinal Study of Adolescent Health:
  - A school-based study of adolescent health and social behaviors;
  - www.cpc.unc.edu/projects/addhealth.

- Data from 160 schools across the US:
  - The smallest had 69 adolescents in grades 7–12;
  - The largest had thousands of participants.

- Relational data:
  - Participants nominated and ranked up to 5 boys and 5 girls as friends.
Example: AddHealth friendships

- “Add Health” - The National Longitudinal Study of Adolescent Health:
  - A school-based study of adolescent health and social behaviors;

- Data from 160 schools across the US:
  - The smallest had 69 adolescents in grades 7–12;
  - The largest had thousands of participants.

- Relational data:
  - Participants nominated and ranked up to 5 boys and 5 girls as friends.
Example: AddHealth friendships

- “Add Health” - The National Longitudinal Study of Adolescent Health:
  - A school-based study of adolescent health and social behaviors;
  - www.cpc.unc.edu/projects/addhealth.

- Data from 160 schools across the US:
  - The smallest had 69 adolescents in grades 7–12;
  - The largest had thousands of participants.

- Relational data:
  - Participants nominated and ranked up to 5 boys and 5 girls as friends.
Example: AddHealth friendships

- “Add Health” - The National Longitudinal Study of Adolescent Health:
  - A school-based study of adolescent health and social behaviors;
  - www.cpc.unc.edu/projects/addhealth.

- Data from 160 schools across the US:
  - The smallest had 69 adolescents in grades 7–12;
  - The largest had thousands of participants.

- Relational data:
  - Participants nominated and ranked up to 5 boys and 5 girls as friends.
Example: AddHealth friendships

- “Add Health” - The National Longitudinal Study of Adolescent Health:
  - A school-based study of adolescent health and social behaviors;
  - www.cpc.unc.edu/projects/addhealth.

- Data from 160 schools across the US:
  - The smallest had 69 adolescents in grades 7–12;
  - The largest had thousands of participants.

- Relational data:
  - Participants nominated and ranked up to 5 boys and 5 girls as friends.
Example: AddHealth friendships

- “Add Health” - The National Longitudinal Study of Adolescent Health:
  - A school-based study of adolescent health and social behaviors;
  - www.cpc.unc.edu/projects/addhealth.

- Data from 160 schools across the US:
  - The smallest had 69 adolescents in grades 7–12;
  - The largest had thousands of participants.

- Relational data:
  - Participants nominated and ranked up to 5 boys and 5 girls as friends.
Example: AddHealth friendships

Notice: Homophily by nodal attributes.
Example: AddHealth friendships

Notice: Homophily by nodal attributes.
Example: AddHealth friendships

Notice: Homophily by nodal attributes.
Example: AddHealth friendships

**Question**: Why might this plot be misleading?
Example: Protein interaction data

Notice: Network structure as compared to the friendship data.
In addition to associations to nodal and dyadic attributes, many networks exhibit the following features:

- **Reciprocity** of ties
- **Degree heterogeneity** in the propensity to form or receive ties
  - sociability
  - popularity
- **Homophily** by actor attributes
  - higher propensity to form ties between actors with similar attributes
  - attributes may be observed or unobserved
- **Transitivity** of relationships
  - friends of friends have a higher propensity to be friends
- **Balance** of relationships
  - liking those who dislike whom you dislike
- **Equivalence** of nodes
  - some nodes may have identical or similar patterns of relationships
Features of many social networks

In addition to associations to nodal and dyadic attributes, many networks exhibit the following features:

- **Reciprocity** of ties
- **Degree heterogeneity** in the propensity to form or receive ties
  - sociability
  - popularity
- **Homophily** by actor attributes
  - higher propensity to form ties between actors with similar attributes
  - attributes may be observed or unobserved
- **Transitivity** of relationships
  - friends of friends have a higher propensity to be friends
- **Balance** of relationships
  - liking those who dislike whom you dislike
- **Equivalence** of nodes
  - some nodes may have identical or similar patterns of relationships
In addition to associations to nodal and dyadic attributes, many networks exhibit the following features:

- **Reciprocity** of ties
- **Degree heterogeneity** in the propensity to form or receive ties
  - sociability
  - popularity
- **Homophily** by actor attributes
  - higher propensity to form ties between actors with similar attributes
  - attributes may be observed or unobserved
- **Transitivity** of relationships
  - friends of friends have a higher propensity to be friends
- **Balance** of relationships
  - liking those who dislike whom you dislike
- **Equivalence** of nodes
  - some nodes may have identical or similar patterns of relationships
Features of many social networks

In addition to associations to nodal and dyadic attributes, many networks exhibit the following features:

- **Reciprocity** of ties
- **Degree heterogeneity** in the propensity to form or receive ties
  - sociability
  - popularity
- **Homophily** by actor attributes
  - higher propensity to form ties between actors with similar attributes
  - attributes may be observed or unobserved
- **Transitivity** of relationships
  - friends of friends have a higher propensity to be friends
- **Balance** of relationships
  - liking those who dislike whom you dislike
- **Equivalence** of nodes
  - some nodes may have identical or similar patterns of relationships
Features of many social networks

In addition to associations to nodal and dyadic attributes, many networks exhibit the following features:

- **Reciprocity** of ties
- **Degree heterogeneity** in the propensity to form or receive ties
  - sociability
  - popularity
- **Homophily** by actor attributes
  - higher propensity to form ties between actors with similar attributes
  - attributes may be observed or unobserved
- **Transitivity** of relationships
  - friends of friends have a higher propensity to be friends
- **Balance** of relationships
  - liking those who dislike whom you dislike
- **Equivalence** of nodes
  - some nodes may have identical or similar patterns of relationships
Features of many social networks

In addition to associations to nodal and dyadic attributes, many networks exhibit the following features:

- **Reciprocity** of ties
- **Degree heterogeneity** in the propensity to form or receive ties
  - sociability
  - popularity
- **Homophily** by actor attributes
  - higher propensity to form ties between actors with similar attributes
  - attributes may be observed or unobserved
- **Transitivity** of relationships
  - friends of friends have a higher propensity to be friends
- **Balance** of relationships
  - liking those who dislike whom you dislike
- **Equivalence** of nodes
  - some nodes may have identical or similar patterns of relationships
Features of many social networks

In addition to associations to nodal and dyadic attributes, many networks exhibit the following features:

- **Reciprocity** of ties
- **Degree heterogeneity** in the propensity to form or receive ties
  - sociability
  - popularity
- **Homophily** by actor attributes
  - higher propensity to form ties between actors with similar attributes
  - attributes may be observed or unobserved
- **Transitivity** of relationships
  - friends of friends have a higher propensity to be friends
- **Balance** of relationships
  - liking those who dislike whom you dislike
- **Equivalence** of nodes
  - some nodes may have identical or similar patterns of relationships
Features of many social networks

In addition to associations to nodal and dyadic attributes, many networks exhibit the following features:

- **Reciprocity** of ties
- **Degree heterogeneity** in the propensity to form or receive ties
  - sociability
  - popularity
- **Homophily** by actor attributes
  - higher propensity to form ties between actors with similar attributes
  - attributes may be observed or unobserved
- **Transitivity** of relationships
  - friends of friends have a higher propensity to be friends
- **Balance** of relationships
  - liking those who dislike whom you dislike
- **Equivalence** of nodes
  - some nodes may have identical or similar patterns of relationships
Features of many social networks

In addition to associations to nodal and dyadic attributes, many networks exhibit the following features:

- **Reciprocity** of ties
- **Degree heterogeneity** in the propensity to form or receive ties
  - sociability
  - popularity
- **Homophily** by actor attributes
  - higher propensity to form ties between actors with similar attributes
  - attributes may be observed or unobserved
- **Transitivity** of relationships
  - friends of friends have a higher propensity to be friends
- **Balance** of relationships
  - liking those who dislike whom you dislike
- **Equivalence** of nodes
  - some nodes may have identical or similar patterns of relationships
Dependent relational data

On notion of statistical dependence is as follows:

**Dependence:**
Two outcomes are dependent if knowing one gives you information about the other.

**Exercise:**
How might network features give rise to statistical dependence?

**Ubiquitous feature of network data:**
Statistical dependence among relational measurements.
On notion of statistical dependence is as follows:

**Dependence:**
Two outcomes are dependent if knowing one gives you information about the other.

**Exercise:**
How might network features give rise to statistical dependence?

**Ubiquitous feature of network data:**
Statistical dependence among relational measurements.
On notion of statistical dependence is as follows:

**Dependence:**
Two outcomes are dependent if knowing one gives you information about the other.

**Exercise:**
How might network features give rise to statistical dependence?

**Ubiquitous feature of network data:**
Statistical dependence among relational measurements.
Dependent relational data

On notion of statistical dependence is as follows:

**Dependence:**
Two outcomes are dependent if knowing one gives you information about the other.

**Exercise:**
How might network features give rise to statistical dependence?

**Ubiquitous feature of network data:**
Statistical dependence among relational measurements.
How do network features drive our data analysis?

1. How can we describe features of social relations?
   (reciprocity/sociability/popularity/transitivity: descriptive statistics)
2. How can we identify nodes with similar network roles?
   (stochastic equivalence: node partitioning)
3. How do we relate the network to covariate information?
   (homophily: regression modeling)
How do network features drive our data analysis?

1. How can we describe features of social relations?
   (reciprocity/sociability/popularity/transitivity : descriptive statistics)

2. How can we identify nodes with similar network roles?
   (stochastic equivalence : node partitioning)

3. How do we relate the network to covariate information?
   (homophily : regression modeling)
Data analysis goals

How do network features drive our data analysis?

1. How can we describe features of social relations?
   (reciprocity/sociability/popularity/transitivity : descriptive statistics)
2. How can we identify nodes with similar network roles?
   (stochastic equivalence : node partitioning)
3. How do we relate the network to covariate information?
   (homophily : regression modeling)
How do network features drive our data analysis?

1. How can we describe features of social relations?
   (reciprocity/sociability/popularity/transitivity : descriptive statistics)
2. How can we identify nodes with similar network roles?
   (stochastic equivalence : node partitioning)
3. How do we relate the network to covariate information?
   (homophily : regression modeling)
How do network features drive our data analysis?

1. How can we describe features of social relations?
   (reciprocity/sociability/popularity/transitivity : descriptive statistics)

2. How can we identify nodes with similar network roles?
   (stochastic equivalence : node partitioning)

3. How do we relate the network to covariate information?
   (homophily : regression modeling)
How do network features drive our data analysis?

1. How can we describe features of social relations?
   (reciprocity/sociability/popularity/transitivity : descriptive statistics)

2. How can we identify nodes with similar network roles?
   (stochastic equivalence : node partitioning)

3. How do we relate the network to covariate information?
   (homophily : regression modeling)
Inferential goals in the regression framework

\( y_{i,j} \) measures \( i \to j \), \( x_{i,j} \) is a vector of explanatory variables.

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

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

Consider a basic (generalized) linear model

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

A model can provide

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

A recurring challenge will be to sufficiently account for dependence in the data.
Inferential goals in the regression framework

\( y_{i,j} \) measures \( i \rightarrow j \), \( x_{i,j} \) is a vector of explanatory variables.

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

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

Consider a basic (generalized) linear model

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

A model can provide

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

A recurring challenge will be to sufficiently account for dependence in the data.
Inferential goals in the regression framework

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

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

Consider a basic (generalized) linear model

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

A model can provide

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

A recurring challenge will be to sufficiently account for dependence in the data.
Inferential goals in the regression framework

\( y_{i,j} \) measures \( i \rightarrow j \), \( x_{i,j} \) is a vector of explanatory variables.

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

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

Consider a basic (generalized) linear model

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

A model can provide

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

A recurring challenge will be to sufficiently account for dependence in the data.
Inferential goals in the regression framework

$y_{i,j}$ measures $i \rightarrow j$, $x_{i,j}$ is a vector of explanatory variables.

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

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

Consider a basic (generalized) linear model

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

A model can provide

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

A recurring challenge will be to sufficiently account for dependence in the data.
Course outline

1. Representations of relational data
   - matrix representations
   - graph representations

2. Descriptive statistics and summaries
   - matrix-based (row/column summaries, matrix decompositions)
   - graph-based (degrees, dyads, triads, paths)
   - covariates

3. Inference for complete relational data
   - model comparison via hypothesis testing
   - regression models
   - $p_1$ and ERGM models
   - social relations model
   - latent variable models (random effects, latent factors and blockmodels)

4. Inference for incomplete or sampled relational data
   - sampling designs (link tracing, egocentric)
   - sample-based inference
   - model-based inference

5. Longitudinal and multivariate relational data
Course outline

1. Representations of relational data
   - matrix representations
   - graph representations

2. Descriptive statistics and summaries
   - matrix-based (row/column summaries, matrix decompositions)
   - graph-based (degrees, dyads, triads, paths)
   - covariates

3. Inference for complete relational data
   - model comparison via hypothesis testing
   - regression models
   - $p_1$ and ERGM models
   - social relations model
   - latent variable models (random effects, latent factors and blockmodels)

4. Inference for incomplete or sampled relational data
   - sampling designs (link tracing, egocentric)
   - sample-based inference
   - model-based inference

5. Longitudinal and multivariate relational data
Course outline

1. Representations of relational data
   - matrix representations
   - graph representations

2. Descriptive statistics and summaries
   - matrix-based (row/column summaries, matrix decompositions)
   - graph-based (degrees, dyads, triads, paths)
   - covariates

3. Inference for complete relational data
   - model comparison via hypothesis testing
   - regression models
   - $p_1$ and ERGM models
   - social relations model
   - latent variable models (random effects, latent factors and blockmodels)

4. Inference for incomplete or sampled relational data
   - sampling designs (link tracing, egocentric)
   - sample-based inference
   - model-based inference

5. Longitudinal and multivariate relational data
Course outline

1. Representations of relational data
   - matrix representations
   - graph representations

2. Descriptive statistics and summaries
   - matrix-based (row/column summaries, matrix decompositions)
   - graph-based (degrees, dyads, triads, paths)
   - covariates

3. Inference for complete relational data
   - model comparison via hypothesis testing
   - regression models
   - \( p_1 \) and ERGM models
   - social relations model
   - latent variable models (random effects, latent factors and blockmodels)

4. Inference for incomplete or sampled relational data
   - sampling designs (link tracing, egocentric)
   - sample-based inference
   - model-based inference

5. Longitudinal and multivariate relational data
Course outline

1. Representations of relational data
   - matrix representations
   - graph representations

2. Descriptive statistics and summaries
   - matrix-based (row/column summaries, matrix decompositions)
   - graph-based (degrees, dyads, triads, paths)
   - covariates

3. Inference for complete relational data
   - model comparison via hypothesis testing
   - regression models
   - $p_1$ and ERGM models
   - social relations model
   - latent variable models (random effects, latent factors and blockmodels)

4. Inference for incomplete or sampled relational data
   - sampling designs (link tracing, egocentric)
   - sample-based inference
   - model-based inference

5. Longitudinal and multivariate relational data
Course outline

1. Representations of relational data
   - matrix representations
   - graph representations

2. Descriptive statistics and summaries
   - matrix-based (row/column summaries, matrix decompositions)
   - graph-based (degrees, dyads, triads, paths)
   - covariates

3. Inference for complete relational data
   - model comparison via hypothesis testing
   - regression models
   - $p_1$ and ERGM models
   - social relations model
   - latent variable models (random effects, latent factors and blockmodels)

4. Inference for incomplete or sampled relational data
   - sampling designs (link tracing, egocentric)
   - sample-based inference
   - model-based inference

5. Longitudinal and multivariate relational data
Course outline

1. Representations of relational data
   - matrix representations
   - graph representations

2. Descriptive statistics and summaries
   - matrix-based (row/column summaries, matrix decompositions)
   - graph-based (degrees, dyads, triads, paths)
   - covariates

3. Inference for complete relational data
   - model comparison via hypothesis testing
   - regression models
   - \( p_1 \) and ERGM models
   - social relations model
   - latent variable models (random effects, latent factors and blockmodels)

4. Inference for incomplete or sampled relational data
   - sampling designs (link tracing, egocentric)
   - sample-based inference
   - model-based inference

5. Longitudinal and multivariate relational data
Course outline

1. Representations of relational data
   • matrix representations
   • graph representations

2. Descriptive statistics and summaries
   • matrix-based (row/column summaries, matrix decompositions)
   • graph-based (degrees, dyads, triads, paths)
   • covariates

3. Inference for complete relational data
   • model comparison via hypothesis testing
   • regression models
   • $p_1$ and ERGM models
   • social relations model
   • latent variable models (random effects, latent factors and blockmodels)

4. Inference for incomplete or sampled relational data
   • sampling designs (link tracing, egocentric)
   • sample-based inference
   • model-based inference

5. Longitudinal and multivariate relational data
1. Representations of relational data
   - matrix representations
   - graph representations

2. Descriptive statistics and summaries
   - matrix-based (row/column summaries, matrix decompositions)
   - graph-based (degrees, dyads, triads, paths)
   - covariates

3. Inference for complete relational data
   - model comparison via hypothesis testing
   - regression models
   - \( p_1 \) and ERGM models
   - social relations model
   - latent variable models (random effects, latent factors and blockmodels)

4. Inference for incomplete or sampled relational data
   - sampling designs (link tracing, egocentric)
   - sample-based inference
   - model-based inference

5. Longitudinal and multivariate relational data
Course outline

1. Representations of relational data
   - matrix representations
   - graph representations

2. Descriptive statistics and summaries
   - matrix-based (row/column summaries, matrix decompositions)
   - graph-based (degrees, dyads, triads, paths)
   - covariates

3. Inference for complete relational data
   - model comparison via hypothesis testing
   - regression models
   - $p_1$ and ERGM models
   - social relations model
   - latent variable models (random effects, latent factors and blockmodels)

4. Inference for incomplete or sampled relational data
   - sampling designs (link tracing, egocentric)
   - sample-based inference
   - model-based inference

5. Longitudinal and multivariate relational data
Course outline

1. Representations of relational data
   - matrix representations
   - graph representations

2. Descriptive statistics and summaries
   - matrix-based (row/column summaries, matrix decompositions)
   - graph-based (degrees, dyads, triads, paths)
   - covariates

3. Inference for complete relational data
   - model comparison via hypothesis testing
   - regression models
   - $p_1$ and ERGM models
   - social relations model
   - latent variable models (random effects, latent factors and blockmodels)

4. Inference for incomplete or sampled relational data
   - sampling designs (link tracing, egocentric)
   - sample-based inference
   - model-based inference

5. Longitudinal and multivariate relational data
1. Representations of relational data
   - matrix representations
   - graph representations

2. Descriptive statistics and summaries
   - matrix-based (row/column summaries, matrix decompositions)
   - graph-based (degrees, dyads, triads, paths)
   - covariates

3. Inference for complete relational data
   - model comparison via hypothesis testing
   - regression models
   - $p_1$ and ERGM models
   - social relations model
   - latent variable models (random effects, latent factors and blockmodels)

4. Inference for incomplete or sampled relational data
   - sampling designs (link tracing, egocentric)
   - sample-based inference
   - model-based inference

5. Longitudinal and multivariate relational data
Course outline

1. Representations of relational data
   - matrix representations
   - graph representations

2. Descriptive statistics and summaries
   - matrix-based (row/column summaries, matrix decompositions)
   - graph-based (degrees, dyads, triads, paths)
   - covariates

3. Inference for complete relational data
   - model comparison via hypothesis testing
   - regression models
   - $p_1$ and ERGM models
   - social relations model
   - latent variable models (random effects, latent factors and blockmodels)

4. Inference for incomplete or sampled relational data
   - sampling designs (link tracing, egocentric)
   - sample-based inference
   - model-based inference

5. Longitudinal and multivariate relational data
Course outline

1. Representations of relational data
   - matrix representations
   - graph representations

2. Descriptive statistics and summaries
   - matrix-based (row/column summaries, matrix decompositions)
   - graph-based (degrees, dyads, triads, paths)
   - covariates

3. Inference for complete relational data
   - model comparison via hypothesis testing
   - regression models
   - $p_1$ and ERGM models
   - social relations model
   - latent variable models (random effects, latent factors and blockmodels)

4. Inference for incomplete or sampled relational data
   - sampling designs (link tracing, egocentric)
   - sample-based inference
   - model-based inference

5. Longitudinal and multivariate relational data
Course outline

1. Representations of relational data
   • matrix representations
   • graph representations

2. Descriptive statistics and summaries
   • matrix-based (row/column summaries, matrix decompositions)
   • graph-based (degrees, dyads, triads, paths)
   • covariates

3. Inference for complete relational data
   • model comparison via hypothesis testing
   • regression models
   • $p_1$ and ERGM models
   • social relations model
   • latent variable models (random effects, latent factors and blockmodels)

4. Inference for incomplete or sampled relational data
   • sampling designs (link tracing, egocentric)
     • sample-based inference
     • model-based inference

5. Longitudinal and multivariate relational data
Course outline

1. Representations of relational data
   - matrix representations
   - graph representations

2. Descriptive statistics and summaries
   - matrix-based (row/column summaries, matrix decompositions)
   - graph-based (degrees, dyads, triads, paths)
   - covariates

3. Inference for complete relational data
   - model comparison via hypothesis testing
   - regression models
   - $p_1$ and ERGM models
   - social relations model
   - latent variable models (random effects, latent factors and blockmodels)

4. Inference for incomplete or sampled relational data
   - sampling designs (link tracing, egocentric)
   - sample-based inference
   - model-based inference

5. Longitudinal and multivariate relational data
Course outline

1. Representations of relational data
   - matrix representations
   - graph representations

2. Descriptive statistics and summaries
   - matrix-based (row/column summaries, matrix decompositions)
   - graph-based (degrees, dyads, triads, paths)
   - covariates

3. Inference for complete relational data
   - model comparison via hypothesis testing
   - regression models
   - $p_1$ and ERGM models
   - social relations model
   - latent variable models (random effects, latent factors and blockmodels)

4. Inference for incomplete or sampled relational data
   - sampling designs (link tracing, egocentric)
   - sample-based inference
   - model-based inference

5. Longitudinal and multivariate relational data
Course outline

1. Representations of relational data
   - matrix representations
   - graph representations

2. Descriptive statistics and summaries
   - matrix-based (row/column summaries, matrix decompositions)
   - graph-based (degrees, dyads, triads, paths)
   - covariates

3. Inference for complete relational data
   - model comparison via hypothesis testing
   - regression models
   - $p_1$ and ERGM models
   - social relations model
   - latent variable models (random effects, latent factors and blockmodels)

4. Inference for incomplete or sampled relational data
   - sampling designs (link tracing, egocentric)
   - sample-based inference
   - model-based inference

5. Longitudinal and multivariate relational data