Introduction to Social Network Analysis in R

Lorien Jasny

1Q-Step Centre, Exeter University
l.jasny@exeter.ac.uk

Sunbelt XXXVIII, Utrecht University
26 June 2018
Think Formally

A network is not just a metaphor: it is a precise, mathematical construct.
Think Formally

A network is not just a metaphor: it is a precise, mathematical construct of nodes (vertices, actors) $N$. 
Think Formally

A network is not just a metaphor: it is a precise, mathematical construct of nodes (vertices, actors) $N$ and edges (ties, relations) $E$
A network is not just a metaphor: it is a precise, mathematical construct of nodes (vertices, actors) $N$ and edges (ties, relations) $E$ that can be directed.
A network is not just a metaphor: it is a precise, mathematical construct of nodes (vertices, actors) $N$ and edges (ties, relations) $E$ that can be directed or undirected.
Think Formally

A network is not just a metaphor: it is a precise, mathematical construct of nodes (vertices, actors) $N$ and edges (ties, relations) $E$ that can be directed or undirected. We can include information (attributes) on the nodes.
A network is not just a metaphor: it is a precise, mathematical construct of nodes (vertices, actors) $N$ and edges (ties, relations) $E$ that can be directed or undirected. We can include information (attributes) on the nodes as well as the edges.
Network Intuition
Network Intuition
Why network methods

- We need a new language to describe what's going on
- Cannot simply use existing statistical methods
- The whole point is that observations are interdependent
- Want to explicitly model these interdependencies
Why network methods

- We need a new language to describe what’s going on
Why network methods

- We need a new language to describe what’s going on
- Cannot simply use existing statistical methods
Why network methods

- We need a new language to describe what’s going on
- Cannot simply use existing statistical methods
- The whole point is that observations are interdependent
Why network methods

- We need a new language to describe what’s going on
- Cannot simply use existing statistical methods
- The whole point is that observations are interdependent
- Want to explicitly model these interdependencies
Data Structures
Data Structures

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>n1</td>
<td>n2</td>
<td>n3</td>
<td>n4</td>
<td>n5</td>
<td>n6</td>
</tr>
<tr>
<td>2</td>
<td>n1</td>
<td>n2</td>
<td>n3</td>
<td>n4</td>
<td>n5</td>
<td>n6</td>
</tr>
<tr>
<td>3</td>
<td>n1</td>
<td>n2</td>
<td>n3</td>
<td>n4</td>
<td>n5</td>
<td>n6</td>
</tr>
<tr>
<td>4</td>
<td>n1</td>
<td>n2</td>
<td>n3</td>
<td>n4</td>
<td>n5</td>
<td>n6</td>
</tr>
<tr>
<td>5</td>
<td>n1</td>
<td>n2</td>
<td>n3</td>
<td>n4</td>
<td>n5</td>
<td>n6</td>
</tr>
<tr>
<td>6</td>
<td>n1</td>
<td>n2</td>
<td>n3</td>
<td>n4</td>
<td>n5</td>
<td>n6</td>
</tr>
</tbody>
</table>

Diagram:

- n1 (Allison) connected to n2 (Drew)
- n3 (Eliot) connected to n4 (Keith)
- n5 (Ross) connected to n6 (Sarah)

The diagram shows a network structure with nodes and connections indicating relationships.
Data Structures

```
<table>
<thead>
<tr>
<th></th>
<th>n1</th>
<th>n2</th>
<th>n3</th>
<th>n4</th>
<th>n5</th>
<th>n6</th>
</tr>
</thead>
<tbody>
<tr>
<td>n1</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>n2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>n3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>n4</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>n5</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>n6</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
```
SNA
LJasny
Intro
Data Structures
Descriptives
Hypothesis Testing

Data Structures

<table>
<thead>
<tr>
<th></th>
<th>$n_1$</th>
<th>$n_2$</th>
<th>$n_3$</th>
<th>$n_4$</th>
<th>$n_5$</th>
<th>$n_6$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$n_1$</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$n_2$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$n_3$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$n_4$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$n_5$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$n_6$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Image of a network with nodes labeled Allison, Drew, Eliot, Keith, Ross, and Sarah. The network is directed with arrows pointing from one node to another, indicating connections between the individuals.
Data Structures

<table>
<thead>
<tr>
<th></th>
<th>$n_1$</th>
<th>$n_2$</th>
<th>$n_3$</th>
<th>$n_4$</th>
<th>$n_5$</th>
<th>$n_6$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$n_1$</td>
<td>1</td>
<td></td>
<td>1</td>
<td>1</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>$n_2$</td>
<td></td>
<td></td>
<td>1</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$n_3$</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$n_4$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1</td>
</tr>
<tr>
<td>$n_5$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>$n_6$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1</td>
</tr>
</tbody>
</table>
**Data Structures**

<table>
<thead>
<tr>
<th></th>
<th>$n_1$</th>
<th>$n_2$</th>
<th>$n_3$</th>
<th>$n_4$</th>
<th>$n_5$</th>
<th>$n_6$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$n_1$</td>
<td>-</td>
<td>1</td>
<td>1</td>
<td>-</td>
<td>1</td>
<td>-</td>
</tr>
<tr>
<td>$n_2$</td>
<td>-</td>
<td>-</td>
<td>1</td>
<td>1</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>$n_3$</td>
<td>1</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>1</td>
</tr>
<tr>
<td>$n_4$</td>
<td>1</td>
<td>1</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>$n_5$</td>
<td>-</td>
<td>1</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>$n_6$</td>
<td>1</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>
Data Structures

<table>
<thead>
<tr>
<th></th>
<th>( n_1 )</th>
<th>( n_2 )</th>
<th>( n_3 )</th>
<th>( n_4 )</th>
<th>( n_5 )</th>
<th>( n_6 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( n_1 )</td>
<td>-</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>( n_2 )</td>
<td>0</td>
<td>-</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>( n_3 )</td>
<td>0</td>
<td>1</td>
<td>-</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>( n_4 )</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>-</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>( n_5 )</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>-</td>
<td>1</td>
</tr>
<tr>
<td>( n_6 )</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>-</td>
</tr>
</tbody>
</table>
Data Structures
Data Structures

<table>
<thead>
<tr>
<th>Sender</th>
<th>Receiver</th>
<th>Weight</th>
</tr>
</thead>
</table>

Diagram:

- **n₁** Allison
- **n₂** Drew
- **n₃** Eliot
- **n₄** Keith
- **n₅** Ross
- **n₆** Sarah
Data Structures

<table>
<thead>
<tr>
<th>Sender</th>
<th>Receiver</th>
<th>Weight</th>
</tr>
</thead>
<tbody>
<tr>
<td>$n_1$</td>
<td>$n_2$</td>
<td>1</td>
</tr>
</tbody>
</table>

Diagram of social network:
- $n_1$ (Allison)
- $n_2$ (Drew)
- $n_3$ (Eliot)
- $n_4$ (Keith)
- $n_5$ (Ross)
- $n_6$ (Sarah)

Connections and weights:
- $n_1$ to $n_2$: weight 1
- $n_2$ to $n_3$: weight 1
- $n_3$ to $n_4$: weight 1
- $n_4$ to $n_5$: weight 1
- $n_5$ to $n_6$: weight 1
- $n_6$ to $n_1$: weight 1
Data Structures

<table>
<thead>
<tr>
<th>Sender</th>
<th>Receiver</th>
<th>Weight</th>
</tr>
</thead>
<tbody>
<tr>
<td>n₁</td>
<td>n₂</td>
<td>1</td>
</tr>
<tr>
<td>n₁</td>
<td>n₅</td>
<td>1</td>
</tr>
<tr>
<td>n₂</td>
<td>n₃</td>
<td>1</td>
</tr>
<tr>
<td>n₂</td>
<td>n₆</td>
<td>1</td>
</tr>
<tr>
<td>n₃</td>
<td>n₂</td>
<td>1</td>
</tr>
<tr>
<td>n₄</td>
<td>n₅</td>
<td>1</td>
</tr>
<tr>
<td>n₅</td>
<td>n₆</td>
<td>1</td>
</tr>
<tr>
<td>n₆</td>
<td>n₂</td>
<td>1</td>
</tr>
</tbody>
</table>
The R Environment

• R is both a language and a software platform
• R software is open-source, cross-platform, and free
• Its home on the web: http://www.r-project.org
The **R** Environment

- **R** is both a language and a software platform
The R Environment

- R is both a language and a software platform
- R software is open-source, cross-platform, and free
The R Environment

- **R** is both a language and a software platform
- R software is open-source, cross-platform, and free
- Its home on the web: http://www.r-project.org
Basics

- R uses a command-like environment, like stata or sas
- R is highly extendable
- You can write your own custom functions
- There are 6000 free add-on packages
- Generally good at reading in/writing out other file formats
- Everything in R is an object – data, functions, everything
Basics

- R uses a command-like environment, like stata or sas
• **R** uses a command-like environment, like stata or sas
• **R** is highly extendable
R uses a command-like environment, like stata or sas
R is highly extendable
  * You can write your own custom functions
**Basics**

- **R** uses a command-like environment, like stata or sas
- **R** is highly extendable
  - You can write your own custom functions
  - There are 6000 free add-on packages
Basics

• **R** uses a command-like environment, like stata or sas
• **R** is highly extendable
  • You can write your own custom functions
  • There are 6000 free add-on packages
• Generally good at reading in/writing out other file formats
• R uses a command-like environment, like stata or sas
• R is highly extendable
  • You can write your own custom functions
  • There are 6000 free add-on packages
• Generally good at reading in/writing out other file formats
• Everything in R is an object – data, functions, everything
Fundamentals
Fundamentals

- When you type commands at the prompt ‘>’ and hit ENTER
Fundamentals

- When you type commands at the prompt ‘>’ and hit **ENTER**
  - **R** tries to interpret what you’ve asked it to do (evaluation)
Fundamentals

- When you type commands at the prompt ‘>’ and hit ENTER
  - R tries to interpret what you’ve asked it to do (evaluation)
  - If it understands what you’ve written, it does it (execution)
Fundamentals

• When you type commands at the prompt ‘>’ and hit \textit{ENTER}:
  • \textbf{R} tries to interpret what you’ve asked it to do (evaluation)
  • If it understands what you’ve written, it does it (execution)
  • If it doesn’t, it will likely give you an error or a warning
Fundamentals

- When you type commands at the prompt ‘>’ and hit `ENTER`
  - \textbf{R} tries to interpret what you’ve asked it to do (evaluation)
  - If it understands what you’ve written, it does it (execution)
  - If it doesn’t, it will likely give you an error or a warning
- Some commands trigger \textbf{R} to print to the screen, others don’t
Fundamentals

- When you type commands at the prompt ‘>’ and hit ENTER
  - \textbf{R} tries to interpret what you’ve asked it to do (evaluation)
  - If it understands what you’ve written, it does it (execution)
  - If it doesn’t, it will likely give you an error or a warning
- Some commands trigger \textbf{R} to print to the screen, others don’t
- If you type an incomplete command, \textbf{R} will usually respond by changing the command prompt to the ‘+’ character
Fundamentals

• When you type commands at the prompt ‘>’ and hit \textit{ENTER}
  • \textbf{R} tries to interpret what you’ve asked it to do (evaluation)
  • If it understands what you’ve written, it does it (execution)
  • If it doesn’t, it will likely give you an error or a warning
• Some commands trigger \textbf{R} to print to the screen, others don’t
• If you type an incomplete command, \textbf{R} will usually respond by changing the command prompt to the ‘+’ character
  • Hit \textit{ESC} on a Mac to cancel
  • Type in \textit{Ctrl + C} on Windows and Linux to cancel
Data Structures in R

- R has several built-in data types and structures
- Common data types:
  - Numeric (integers, numbers, etc.)
  - 3.14
  - Strings (alphanumeric characters in quotation marks)
    - "hello"
    - "3.14"
- Common data structures
  - Vectors
  - Matrices
  - Data Frames
  - Network objects
Data Structures in R

- R has several built-in data types and structures
Data Structures in \textbf{R}

- \textbf{R} has several built-in data types and structures
- Common data types:
Data Structures in \textbf{R}

- \textbf{R} has several built-in data types and structures
- Common data types:
  - Numeric (integers, numbers, etc.)
Data Structures in **R**

- **R** has several built-in data types and structures
- Common data types:
  - Numeric (integers, numbers, etc.)
    - 12
    - 3.14
  - Strings (alphanumeric characters in quotation marks)
    - "hello"
    - "3.14"
Data Structures in R

- **R** has several built-in data types and structures
- **Common data types:**
  - Numeric (integers, numbers, etc.)
    - 12
    - 3.14
  - Strings (alphanumeric characters in quotation marks)
Data Structures in R

- **R** has several built-in data types and structures
- **Common data types:**
  - Numeric (integers, numbers, etc.)
    - 12
    - 3.14
  - Strings (alphanumeric characters in quotation marks)
    - “hello”
    - “3.14”
Data Structures in \textbf{R}

- \textbf{R} has several built-in data types and structures
- Common data types:
  - Numeric (integers, numbers, etc.)
    - 12
    - 3.14
  - Strings (alphanumeric characters in quotation marks)
    - “hello”
    - “3.14”
Data Structures in **R**

- **R** has several built-in data types and structures
- **Common data types:**
  - Numeric (integers, numbers, etc.)
    - 12
    - 3.14
  - Strings (alphanumeric characters in quotation marks)
    - “hello”
    - “3.14”
- **Common data structures**
Data Structures in \textbf{R}

- \textbf{R} has several built-in data types and structures
- Common data types:
  - Numeric (integers, numbers, etc.)
    - 12
    - 3.14
  - Strings (alphanumeric characters in quotation marks)
    - “hello”
    - “3.14”
- Common data structures
  - Vectors
Data Structures in R

- R has several built-in data types and structures
- Common data types:
  - Numeric (integers, numbers, etc.)
    - 12
    - 3.14
  - Strings (alphanumeric characters in quotation marks)
    - “hello”
    - “3.14”
- Common data structures
  - Vectors
  - Matrices
Data Structures in \textbf{R}

- \textbf{R} has several built-in data types and structures
- Common data types:
  - Numeric (integers, numbers, etc.)
    - 12
    - 3.14
  - Strings (alphanumeric characters in quotation marks)
    - “hello”
    - “3.14”
- Common data structures
  - Vectors
  - Matrices
  - Data Frames
**Data Structures in R**

- **R** has several built-in data types and structures
- **Common data types:**
  - Numeric (integers, numbers, etc.)
    - 12
    - 3.14
  - Strings (alphanumeric characters in quotation marks)
    - “hello”
    - “3.14”
- **Common data structures**
  - Vectors
  - Matrices
  - Data Frames
  - Network objects
A vector is a one-dimensional data structure
Vectors

A vector is a one-dimensional data structure

- Vectors are indexed starting at 1
A vector is a one-dimensional data structure

- Vectors are indexed starting at 1
- A vector of length $n$ has $n$ cells
A vector is a one-dimensional data structure

- Vectors are indexed starting at 1
- A vector of length $n$ has $n$ cells
- Each cell can hold a single value
Vectors

A vector is a one-dimensional data structure

- Vectors are indexed starting at 1
- A vector of length $n$ has $n$ cells
- Each cell can hold a single value
- Vectors can only store data of the same type – either all strings or all numerical but not both
Working with Vectors in \textbf{R}
• We index vectors in \( \mathbf{R} \) using “square bracket notation”
Working with Vectors in \( \mathbb{R} \)

- We index vectors in \( \mathbb{R} \) using “square bracket notation”
- Example:
  - you have a vector of numeric values called \texttt{testScores}
Working with Vectors in $\mathbf{R}$

- We index vectors in $\mathbf{R}$ using “square bracket notation”
- Example:
  - you have a vector of numeric values called `testScores`
  - To retrieve the value in the third cell, type `testScores[3]`
Working with Vectors in \textbf{R}

- We index vectors in \textbf{R} using “square bracket notation”
- Example:
  - you have a vector of numeric values called \texttt{testScores}
  - To retrieve the value in the third cell, type \texttt{testScores[3]}
  - To retrieve all BUT the third value, type \texttt{testScores[-3]}
Two-dimensional data in $\mathbf{R}$

- Most (all?) of us are familiar with two-dimensional data like that in spreadsheets
Two-dimensional data in $\mathbf{R}$

- Most (all?) of us are familiar with two-dimensional data like that in spreadsheets
- $\mathbf{R}$ has two built-in data structures for storing two-dimensional data
Two-dimensional data in R

- Most (all?) of us are familiar with two-dimensional data like that in spreadsheets
- \textbf{R} has two built-in data structures for storing two-dimensional data
  - Matrices
Two-dimensional data in R

- Most (all?) of us are familiar with two-dimensional data like that in spreadsheets
- R has two built-in data structures for storing two-dimensional data
  - Matrices
  - Data Frames
Two-dimensional data in R

- Most (all?) of us are familiar with two-dimensional data like that in spreadsheets
- R has two built-in data structures for storing two-dimensional data
  - Matrices
  - Data Frames
- In most instances, they behave the same
Two-dimensional data in \( \mathbf{R} \)

- Most (all?) of us are familiar with two-dimensional data like that in spreadsheets
- \( \mathbf{R} \) has two built-in data structures for storing two-dimensional data
  - Matrices
  - Data Frames
- In most instances, they behave the same
- Most functions will accept either a matrix or a data frame
Matrices versus data frames in \texttt{R}

- Matrices can only store data of one type
  - Either all strings or all numbers, but not both
  - If you try to give it multiple types, \texttt{R} converts everything to string format
- Data frames can store data of multiple types
  - Ideal for classical data analysis where you might have a mix of numerical and string data
Matrices versus data frames in R

- Matrices can only store data of one type
Matrices versus data frames in R

- Matrices can only store data of one type
  - Either all strings or all numbers, but not both
Matrices versus data frames in $\mathbf{R}$

- Matrices can only store data of one type
  - Either all strings or all numbers, but not both
  - If you try to give it multiple types, $\mathbf{R}$ converts everything to string format
Matrices versus data frames in R

- Matrices can only store data of one type
  - Either all strings or all numbers, but not both
  - If you try to give it multiple types, R converts everything to string format
- Data frames can store data of multiple types
Matrices versus data frames in \texttt{R}

- Matrices can only store data of one type
  - Either all strings or all numbers, but not both
  - If you try to give it multiple types, \texttt{R} converts everything to string format
- Data frames can store data of multiple types
  - Ideal for classical data analysis where you might have a mix of numerical and string data
### Working with matrices

<table>
<thead>
<tr>
<th>id</th>
<th>name</th>
<th>age</th>
<th>sex</th>
<th>handed</th>
<th>lastDocVisit</th>
</tr>
</thead>
<tbody>
<tr>
<td>5012</td>
<td>Danielle</td>
<td>44</td>
<td>F</td>
<td>R</td>
<td>2012</td>
</tr>
<tr>
<td>2331</td>
<td>Josh</td>
<td>44</td>
<td>M</td>
<td>R</td>
<td>2008</td>
</tr>
<tr>
<td>1989</td>
<td>Mark</td>
<td>40</td>
<td>M</td>
<td>R</td>
<td>2010</td>
</tr>
<tr>
<td>2217</td>
<td>Emma</td>
<td>32</td>
<td>F</td>
<td>L</td>
<td>2012</td>
</tr>
<tr>
<td>2912</td>
<td>Sarah</td>
<td>33</td>
<td>F</td>
<td>R</td>
<td>2011</td>
</tr>
</tbody>
</table>

• Can also use "square bracket notation"
• Inside the square brackets, the first position refers to the row(s) and the second to the column(s)
• If this matrix is called `friendSurvey`, the command to retrieve Josh's age is `friendSurvey[2,3]`
## Working with matrices

<table>
<thead>
<tr>
<th>id</th>
<th>name</th>
<th>age</th>
<th>sex</th>
<th>handed</th>
<th>lastDocVisit</th>
</tr>
</thead>
<tbody>
<tr>
<td>5012</td>
<td>Danielle</td>
<td>44</td>
<td>F</td>
<td>R</td>
<td>2012</td>
</tr>
<tr>
<td>2331</td>
<td>Josh</td>
<td>44</td>
<td>M</td>
<td>R</td>
<td>2008</td>
</tr>
<tr>
<td>1989</td>
<td>Mark</td>
<td>40</td>
<td>M</td>
<td>R</td>
<td>2010</td>
</tr>
<tr>
<td>2217</td>
<td>Emma</td>
<td>32</td>
<td>F</td>
<td>L</td>
<td>2012</td>
</tr>
<tr>
<td>2912</td>
<td>Sarah</td>
<td>33</td>
<td>F</td>
<td>R</td>
<td>2011</td>
</tr>
</tbody>
</table>

- Can also use “square bracket notation”
Working with matrices

<table>
<thead>
<tr>
<th>id</th>
<th>name</th>
<th>age</th>
<th>sex</th>
<th>handed</th>
<th>lastDocVisit</th>
</tr>
</thead>
<tbody>
<tr>
<td>5012</td>
<td>Danielle</td>
<td>44</td>
<td>F</td>
<td>R</td>
<td>2012</td>
</tr>
<tr>
<td>2331</td>
<td>Josh</td>
<td>44</td>
<td>M</td>
<td>R</td>
<td>2008</td>
</tr>
<tr>
<td>1989</td>
<td>Mark</td>
<td>40</td>
<td>M</td>
<td>R</td>
<td>2010</td>
</tr>
<tr>
<td>2217</td>
<td>Emma</td>
<td>32</td>
<td>F</td>
<td>L</td>
<td>2012</td>
</tr>
<tr>
<td>2912</td>
<td>Sarah</td>
<td>33</td>
<td>F</td>
<td>R</td>
<td>2011</td>
</tr>
</tbody>
</table>

- Can also use “square bracket notation”
- Inside the square brackets, the first position refers to the row(s) and the second to the column(s)
Working with matrices

<table>
<thead>
<tr>
<th>id</th>
<th>name</th>
<th>age</th>
<th>sex</th>
<th>handed</th>
<th>lastDocVisit</th>
</tr>
</thead>
<tbody>
<tr>
<td>5012</td>
<td>Danielle</td>
<td>44</td>
<td>F</td>
<td>R</td>
<td>2012</td>
</tr>
<tr>
<td>2331</td>
<td>Josh</td>
<td>44</td>
<td>M</td>
<td>R</td>
<td>2008</td>
</tr>
<tr>
<td>1989</td>
<td>Mark</td>
<td>40</td>
<td>M</td>
<td>R</td>
<td>2010</td>
</tr>
<tr>
<td>2217</td>
<td>Emma</td>
<td>32</td>
<td>F</td>
<td>L</td>
<td>2012</td>
</tr>
<tr>
<td>2912</td>
<td>Sarah</td>
<td>33</td>
<td>F</td>
<td>R</td>
<td>2011</td>
</tr>
</tbody>
</table>

- Can also use “square bracket notation”
- Inside the square brackets, the first position refers to the row(s) and the second to the column(s)
- If this matrix is called friendSurvey, the command to retrieve Josh’s age is friendSurvey[2,3]
Working with data frames
Working with data frames

• Square bracket notation works for data frames as well
Working with data frames

- Square bracket notation works for data frames as well
- Data frames provide another option: dollar sign notation
## Working with data frames

<table>
<thead>
<tr>
<th>id</th>
<th>name</th>
<th>age</th>
<th>sex</th>
<th>handed</th>
<th>lastDocVisit</th>
</tr>
</thead>
<tbody>
<tr>
<td>5012</td>
<td>Danielle</td>
<td>44</td>
<td>F</td>
<td>R</td>
<td>2012</td>
</tr>
<tr>
<td>2331</td>
<td>Josh</td>
<td>44</td>
<td>M</td>
<td>R</td>
<td>2008</td>
</tr>
<tr>
<td>1989</td>
<td>Mark</td>
<td>40</td>
<td>M</td>
<td>R</td>
<td>2010</td>
</tr>
<tr>
<td>2217</td>
<td>Emma</td>
<td>32</td>
<td>F</td>
<td>L</td>
<td>2012</td>
</tr>
<tr>
<td>2912</td>
<td>Sarah</td>
<td>33</td>
<td>F</td>
<td>R</td>
<td>2011</td>
</tr>
</tbody>
</table>

- To retrieve the 'sex' column as a vector, use `friendSurvey$sex`
- To retrieve Josh's age, use `friendSurvey$age[2]`
Working with data frames

<table>
<thead>
<tr>
<th>id</th>
<th>name</th>
<th>age</th>
<th>sex</th>
<th>handed</th>
<th>lastDocVisit</th>
</tr>
</thead>
<tbody>
<tr>
<td>5012</td>
<td>Danielle</td>
<td>44</td>
<td>F</td>
<td>R</td>
<td>2012</td>
</tr>
<tr>
<td>2331</td>
<td>Josh</td>
<td>44</td>
<td>M</td>
<td>R</td>
<td>2008</td>
</tr>
<tr>
<td>1989</td>
<td>Mark</td>
<td>40</td>
<td>M</td>
<td>R</td>
<td>2010</td>
</tr>
<tr>
<td>2217</td>
<td>Emma</td>
<td>32</td>
<td>F</td>
<td>L</td>
<td>2012</td>
</tr>
<tr>
<td>2912</td>
<td>Sarah</td>
<td>33</td>
<td>F</td>
<td>R</td>
<td>2011</td>
</tr>
</tbody>
</table>

- To retrieve the ‘sex’ column as a vector, use `friendSurvey$sex`
## Working with data frames

<table>
<thead>
<tr>
<th>id</th>
<th>name</th>
<th>age</th>
<th>sex</th>
<th>handed</th>
<th>lastDocVisit</th>
</tr>
</thead>
<tbody>
<tr>
<td>5012</td>
<td>Danielle</td>
<td>44</td>
<td>F</td>
<td>R</td>
<td>2012</td>
</tr>
<tr>
<td>2331</td>
<td>Josh</td>
<td>44</td>
<td>M</td>
<td>R</td>
<td>2008</td>
</tr>
<tr>
<td>1989</td>
<td>Mark</td>
<td>40</td>
<td>M</td>
<td>R</td>
<td>2010</td>
</tr>
<tr>
<td>2217</td>
<td>Emma</td>
<td>32</td>
<td>F</td>
<td>L</td>
<td>2012</td>
</tr>
<tr>
<td>2912</td>
<td>Sarah</td>
<td>33</td>
<td>F</td>
<td>R</td>
<td>2011</td>
</tr>
</tbody>
</table>

- To retrieve the ‘sex’ column as a vector, use `friendSurvey$sex`
- To retrieve Josh’s age, use `friendSurvey$age[2]`
Network Objects

• stores an adjacency matrix or an edgelist as well as metadata
• vertex, edge, and network attributes
• can use square-bracket notation just like a matrix
Network Objects

- stores an adjacency matrix or an edgelist as well as metadata
Network Objects

- stores an adjacency matrix or an edgelist as well as metadata
  - vertex, edge, and network attributes
Network Objects

- stores an adjacency matrix or an edgelist as well as metadata
  - vertex, edge, and network attributes
- can use square-bracket notation just like a matrix
Network Objects

- stores an adjacency matrix or an edgelist as well as metadata
  - vertex, edge, and network attributes
- can use square-bracket notation just like a matrix

```r
Network attributes:
vertices = 18
directed = TRUE
hyper = FALSE
loops = FALSE
multiple = FALSE
bipartite = FALSE
total edges= 54
  missing edges= 0
  non-missing edges= 54

Vertex attribute names:
  Group vertex.names

Edge attribute names:
  Order
```
Code Time!

- Sections 1-3
Descriptives

- One isolate
- Two components
- Diameter is 5
- Medici is most popular
- Three triads
Descriptives

- One isolate
Descriptives

- One isolate
- Two components
Descriptives

- One isolate
- Two components
- Diameter is 5
Descriptives

- One isolate
- Two components
- Diameter is 5
- Medici is most popular
Descriptives

- One isolate
- Two components
- Diameter is 5
- Medici is most popular
- Three triads
For each node, its degree is: 

- Pucci has degree 0
- Lamberteschi has degree 1
- Guadagni has degree 4
- Medici has degree 6
Degree

For each node, its degree is

- the number of nodes adjacent to it
Degree

For each node, its degree is

• the number of nodes adjacent to it
• or, the number of lines incident with it
For each node, its degree is

- the number of nodes adjacent to it
- or, the number of lines incident with it
For each node, its degree is

- the number of nodes adjacent to it
- or, the number of lines incident with it

- Pucci has degree 0
For each node, its degree is

- the number of nodes adjacent to it
- or, the number of lines incident with it

- Pucci has degree 0
- Lamberteschi has degree 1
Degree

For each node, its degree is

- the number of nodes adjacent to it
- or, the number of lines incident with it

- Pucci has degree 0
- Lamberteschi has degree 1
- Guadagni has degree 4
For each node, its degree is

- the number of nodes adjacent to it
- or, the number of lines incident with it

- Pucci has degree 0
- Lamberteschi has degree 1
- Guadagni has degree 4
- Medici has degree 6
In directed graphs,

- A has 1 outdegree
- B has 1 outdegree and 3 indegree
- C has 2 outdegree and 1 indegree
- D has 1 outdegree and 1 indegree
Directed Degree

In directed graphs,

- *Indegree* indicates the number of received ties

![Directed Graph Example](image-url)
Directed Degree

In directed graphs,

- *Indegree* indicates the number of received ties
- *Outdegree* indicates the number of sent ties

![Diagram](attachment:diagram.png)
Directed Degree

In directed graphs,

- **Indegree** indicates the number of received ties
- **Outdegree** indicates the number of sent ties

- A has 1 outdegree
Directed Degree

In directed graphs,

- *Indegree* indicates the number of received ties
- *Outdegree* indicates the number of sent ties

- A has 1 outdegree
- B has 1 outdegree
Directed Degree

In directed graphs,

- *Indegree* indicates the number of received ties
- *Outdegree* indicates the number of sent ties

- A has 1 outdegree
- B has 1 outdegree and 3 indegree
In directed graphs,

- **Indegree** indicates the number of received ties
- **Outdegree** indicates the number of sent ties

- A has 1 outdegree
- B has 1 outdegree and 3 indegree
- C has 2 outdegree

![Diagram of directed graph with nodes A, B, C, and D connected by arrows indicating indegree and outdegree.]
In directed graphs,

- *Indegree* indicates the number of received ties
- *Outdegree* indicates the number of sent ties

- A has 1 outdegree
- B has 1 outdegree and 3 indegree
- C has 2 outdegree and 1 indegree
In directed graphs,

- **Indegree** indicates the number of received ties
- **Outdegree** indicates the number of sent ties

![Diagram](image-url)

- A has 1 outdegree
- B has 1 outdegree and 3 indegree
- C has 2 outdegree and 1 indegree
- D has 1 outdegree
Directed Degree

In directed graphs,

- *Indegree* indicates the number of received ties
- *Outdegree* indicates the number of sent ties

- A has 1 outdegree
- B has 1 outdegree and 3 indegree
- C has 2 outdegree and 1 indegree
- D has 1 outdegree and 1 indegree
Calculate degree centrality?
Calculate degree centrality?

<table>
<thead>
<tr>
<th></th>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
<th>E</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>-</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>B</td>
<td>0</td>
<td>-</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>C</td>
<td>0</td>
<td>0</td>
<td>-</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>D</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>-</td>
<td>1</td>
</tr>
<tr>
<td>E</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>-</td>
</tr>
</tbody>
</table>
Pseudo Code
Pseudo Code

for each vertex{
   Count the number of 1’s in the row
}

R Code

```r
rs ← vector(length=nrow(network))
for(i in 1:nrow(network)) {
  network[rs] ← sum(network[i,])
}
Or, rowSums(network)
```
R Code

```r
network_rs ← vector(length=nrow(network))
for(i in 1:nrow(network)){
  network_rs ← sum(network[i,])
}
```

network_rs ← vector(length=nrow(network))
for(i in 1:nrow(network)){
    network_rs ← sum(network[i,])
}

Or, rowSums(network)
Betweenness Centrality

Betweenness Centrality is a measure of a node's importance in a network. It is defined as the proportion of shortest paths between all other pairs of nodes that the given node lies on.

\[ \text{Betweenness Centrality} = \frac{\text{Number of shortest paths that pass through the node}}{\text{Total number of shortest paths}} \]

For example, consider a network with nodes A, B, C, and D. The betweenness centrality of each node can be calculated as follows:

- **A** sits on no paths between others.
- **B** sits on some paths:
  - \( A \rightarrow C \) = 1
  - \( A \rightarrow D \) = 1
- **C** sits on some paths:
  - \( C \rightarrow D \) = 0
  - \( D \rightarrow C \) = 1
  - \( C \rightarrow A \) = 0
  - \( D \rightarrow A \) = 0
Betweenness Centrality

Proportion of shortest paths between all other pairs of nodes that the given node lies on
Betweenness Centrality

Proportion of shortest paths between all other pairs of nodes that the given node lies on

```
A  B  C  D
```

• A sits on no paths between others
• B sits on some paths:
  • $A \rightarrow C = 1$
  • $A \rightarrow D = 1$
  • $C \rightarrow D = 0$
  • $D \rightarrow C = 1$
  • $C \rightarrow A = 0$
  • $D \rightarrow A = 0$
Betweenness Centrality

Proportion of shortest paths between all other pairs of nodes that the given node lies on

- A sits on no paths between others
Betweenness Centrality

Proportion of shortest paths between all other pairs of nodes that the given node lies on

- A sits on no paths between others
- B sits on some paths:
Betweenness Centrality

Proportion of shortest paths between all other pairs of nodes that the given node lies on

- A sits on no paths between others
- B sits on some paths:
  - A→ C
  - A→ D
  - C→ D
  - D→ C
  - C→ A
  - D→ A
Betweenness Centrality

Proportion of shortest paths between all other pairs of nodes that the given node lies on

- A sits on no paths between others
- B sits on some paths:
  - $A \rightarrow C$
  - $A \rightarrow D$
  - $C \rightarrow D$
  - $D \rightarrow C$
  - $C \rightarrow A$
  - $D \rightarrow A$
Betweenness Centrality

Proportion of shortest paths between all other pairs of nodes that the given node lies on

- A sits on no paths between others
- B sits on some paths:
  - $A \rightarrow C = 1$
  - $A \rightarrow D$
  - $C \rightarrow D$
  - $D \rightarrow C$
  - $C \rightarrow A$
  - $D \rightarrow A$
Betweenness Centrality

Proportion of shortest paths between all other pairs of nodes that the given node lies on

- A sits on no paths between others
- B sits on some paths:
  - $A \rightarrow C = 1$
  - $A \rightarrow D$
  - $C \rightarrow D$
  - $D \rightarrow C$
  - $C \rightarrow A$
  - $D \rightarrow A$
Betweenness Centrality

Proportion of shortest paths between all other pairs of nodes that the given node lies on

- A sits on no paths between others
- B sits on some paths:
  - $A \rightarrow C = 1$
  - $A \rightarrow D = 1$
  - $C \rightarrow D$
  - $D \rightarrow C$
  - $C \rightarrow A$
  - $D \rightarrow A$
Betweenness Centrality

Proportion of shortest paths between all other pairs of nodes that the given node lies on

- A sits on no paths between others
- B sits on some paths:
  - A → C = 1
  - A → D = 1
  - C → D
  - D → C
  - C → A
  - D → A
Betweenness Centrality

Proportion of shortest paths between all other pairs of nodes that the given node lies on

- A sits on no paths between others
- B sits on some paths:
  - $A \rightarrow C = 1$
  - $A \rightarrow D = 1$
  - $C \rightarrow D = 0$
  - $D \rightarrow C$
  - $C \rightarrow A$
  - $D \rightarrow A$
Betweenness Centrality

Proportion of shortest paths between all other pairs of nodes that the given node lies on

- A sits on no paths between others
- B sits on some paths:
  - \( A \rightarrow C = 1 \)
  - \( A \rightarrow D = 1 \)
  - \( C \rightarrow D = 0 \)
  - \( D \rightarrow C \)
  - \( C \rightarrow A \)
  - \( D \rightarrow A \)
Betweenness Centrality

Proportion of shortest paths between all other pairs of nodes that the given node lies on

- A sits on no paths between others
- B sits on some paths:
  - A→C = 1
  - A→D = 1
  - C→D = 0
  - D→C = 1
  - C→A
  - D→A
Betweenness Centrality

Proportion of shortest paths between all other pairs of nodes that the given node lies on

- A sits on no paths between others
- B sits on some paths:
  - $A \rightarrow C = 1$
  - $A \rightarrow D = 1$
  - $C \rightarrow D = 0$
  - $D \rightarrow C = 1$
  - $C \rightarrow A$
  - $D \rightarrow A$
Betweenness Centrality

Proportion of shortest paths between all other pairs of nodes that the given node lies on

- A sits on no paths between others
- B sits on some paths:
  - \( A \rightarrow C = 1 \)
  - \( A \rightarrow D = 1 \)
  - \( C \rightarrow D = 0 \)
  - \( D \rightarrow C = 1 \)
  - \( C \rightarrow A = 0 \)
  - \( D \rightarrow A \)
Betweenness Centrality

Proportion of shortest paths between all other pairs of nodes that the given node lies on

- A sits on no paths between others
- B sits on some paths:
  - $A \rightarrow C = 1$
  - $A \rightarrow D = 1$
  - $C \rightarrow D = 0$
  - $D \rightarrow C = 1$
  - $C \rightarrow A = 0$
  - $D \rightarrow A = 0$
Betweenness Centrality

Proportion of shortest paths between all other pairs of nodes that the given node lies on

- A sits on no paths between others
- B sits on some paths:
  - $A \rightarrow C = 1$
  - $A \rightarrow D = 1$
  - $C \rightarrow D = 0$
  - $D \rightarrow C = 1$
  - $C \rightarrow A = 0$
  - $D \rightarrow A = 0$
Betweenness Centrality

Proportion of shortest paths between all other pairs of nodes that the given node lies on

A → C = 0
A → D = 1/2
C → D = 0
D → C = 0
C → A = 0
D → A = 0
Betweenness Centrality

Proportion of shortest paths between all other pairs of nodes that the given node lies on

- B sits on some paths:
  - A → C
  - A → D
  - C → D
  - D → C
  - C → A
  - D → A
Betweenness Centrality

Proportion of shortest paths between all other pairs of nodes that the given node lies on

- B sits on some paths:
  - $A \rightarrow C = 0$
  - $A \rightarrow D$
  - $C \rightarrow D$
  - $D \rightarrow C$
  - $C \rightarrow A$
  - $D \rightarrow A$
Betweenness Centrality

Proportion of shortest paths between all other pairs of nodes that the given node lies on

- **B** sits on some paths:
  - A → C = 0
  - A → D
  - C → D
  - D → C
  - C → A
  - D → A
Betweenness Centrality

Proportion of shortest paths between all other pairs of nodes that the given node lies on

- B sits on some paths:
  - A→C = 0
  - A→D
  - C→D
  - D→C
  - C→A
  - D→A
Betweenness Centrality

Proportion of shortest paths between all other pairs of nodes that the given node lies on

- B sits on some paths:
  - A→C = 0
  - A→D = 1/2
  - C→D
  - D→C
  - C→A
  - D→A
Betweenness Centrality

Proportion of shortest paths between all other pairs of nodes that the given node lies on

- **B** sits on some paths:
  - \(A \to C = 0\)
  - \(A \to D = 1/2\)
  - \(C \to D = 0\)
  - \(D \to C = 0\)
  - \(C \to A = 0\)
  - \(D \to A = 0\)
Betweenness Centrality

Forrest Pitts 1978 “The River Trade Network of Russia, Revisited”
Betweenness Centrality

Forrest Pitts 1978 “The River Trade Network of Russia, Revisited”
Closeness Centrality

How far is each node from the other nodes in the graph
Closeness Centrality

How far is each node from the other nodes in the graph

\[ C(v) = \frac{|V| - 1}{\sum d(v,i)} \]

- A's Distance to:
  - B = 1
  - C = 2
  - D = 3

A's Closeness Centrality = \( \frac{1 + 2 + 3}{3} = 1.5 \)
Closeness Centrality

How far is each node from the other nodes in the graph

\[ C(v) = \frac{|V| - 1}{\sum d(v,i)} \]
Closeness Centrality

How far is each node from the other nodes in the graph

\[ C(v) = \frac{|V|-1}{\sum d(v,i)} \]

• A’s Distance to

How far is each node from the other nodes in the graph

- B = 1
- C = 2
- D = 3
- A’s Closeness Centrality = \[ \frac{1+2+3}{3} = 1.6667 \]
Closeness Centrality

How far is each node from the other nodes in the graph

- $C(v) = \frac{|V| - 1}{\sum d(v,i)}$
- A’s Distance to
  - B = 1
  - C = 2
  - D = 3
- A’s Closeness Centrality $= \frac{1+2+3}{3}=1.5$
Closeness Centrality

How far is each node from the other nodes in the graph

- $C(v) = \frac{|V|-1}{\sum d(v,i)}$
- A’s Distance to
  - $B = 1$
  - $C = 2$
Closeness Centrality

How far is each node from the other nodes in the graph

\[ C(v) = \frac{|V| - 1}{\sum d(v, i)} \]

- A’s Distance to
  - B = 1
  - C = 2
  - D = 3
Closeness Centrality

How far is each node from the other nodes in the graph

- \[ C(v) = \frac{|V|-1}{\sum d(v,i)} \]
- A’s Distance to
  - B = 1
  - C = 2
  - D = 3
- A’s Closeness Centrality
  \[ = \frac{3}{1+2+3} = .5 \]
Closeness Centrality

How far is each node from the other nodes in the graph

- \[ C(v) = \frac{|V|-1}{\sum d(v,i)} \]
- B’s Distance to

\[
\begin{align*}
\text{• } C(v) &= |V|-1 \sum d(v,i) \\
\text{• } B\text{’s Distance to}
\end{align*}
\]
Closeness Centrality

How far is each node from the other nodes in the graph

- \( C(v) = \frac{|V|-1}{\sum d(v,i)} \)
- B’s Distance to
  - C = 1
Closeness Centrality

How far is each node from the other nodes in the graph

\[ C(v) = \frac{|V|-1}{\sum_{i} d(v,i)} \]

- B's Distance to
  - B = \infty
  - B's Closeness Centrality = \frac{\infty+1+2}{3} = \infty

- C = 1
- D = 2
Closeness Centrality

How far is each node from the other nodes in the graph

- \( C(v) = \frac{|V|-1}{\sum d(v,i)} \)

- B’s Distance to
  - \( C = 1 \)
  - \( D = 2 \)
  - \( B = \infty \)
How far is each node from the other nodes in the graph

- \( C(v) = \frac{|V|-1}{\sum d(v,i)} \)
- B’s Distance to
  - C = 1
  - D = 2
  - B = \( \infty \)
- B’s Closeness Centrality
  \[
  \frac{3}{\infty+1+2} = \infty
  \]
Closeness Centrality – Alternate Measure

How far is each node from the other nodes in the graph

\[ C^2(v) = \frac{1}{|V|-1} \sum \frac{1}{d(v,i)} \]

- B’s Distance to
  - C = 1
  - D = 2
  - B = \infty

A

\[ \rightarrow \]

B

\[ \uparrow \]

\[ \rightarrow \]

C

\[ \downarrow \]

\[ \rightarrow \]

D
Closeness Centrality – Alternate Measure

How far is each node from the other nodes in the graph

- \( C^2(v) = \frac{1}{|V| - 1} \sum \frac{1}{d(v,i)} \)
- B’s Distance to
  - C = 1
  - D = 2
  - B = \( \text{inf} \)
- B’s Closeness Centrality
  \[
  \frac{\frac{1}{\text{inf}} + \frac{1}{1} + \frac{1}{2}}{3} = \frac{1}{2}
  \]
Page Rank and Eigenvector Centrality
Page Rank
Page Rank

A network diagram showing relationships between individuals with Page Rank scores.

- Amy: 1
- Bryan: 1
- Erica: 1
- David: 1
- Carter: 1
Page Rank
Page Rank

Amy : 0
Bryan : 1+.5
Erica : 1
David : 1
Carter : 1+.5
Page Rank

Erica
1

Amy
0

Bryan
1+.5

Carter
1+.5

David
1

1+1
Page Rank

Diagram:

- Amy: 0
- Bryan: 0.5
- Erica: 1+1
- David: 1
- Carter: 1.5

Connections:
- Amy to Erica: 1
- Amy to David: 1
- Erica to David: 1
- Erica to Carter: 1
- David to Amy: 1
- Bryan to Carter: 0.5
- Bryan to Amy: 1
- Bryan to Erica: 0.5
- Carter to Erica: 0.5
- Carter to Bryan: 1
- Carter to David: 1
- David to Bryan: 0.5
- David to Carter: 0.5
Page Rank

- Amy: 0
- Bryan: 0+.5
- Erica: 1+1
- David: 1
- Carter: 1+.5
Page Rank

![Graph of Page Rank]

- Amy: 0
- Bryan: 0.5
- Erica: 1+1
- David: 1+1
- Carter: 0+.5
Page Rank

The diagram illustrates a network with nodes representing individuals and directed edges indicating citations. The numbers next to each node represent the PageRank score, which is a measure of the importance of a node in the network.

- **Amy** with a PageRank of 0
- **Bryan** with a PageRank of 0.5
- **Erica** with a PageRank of 1.1
- **David** with a PageRank of 1.1
- **Carter** with a PageRank of 1.5

The PageRank algorithm assigns more weight to nodes that are cited by other important nodes.

---

Source: SNA Intro Data Structures Descriptives Hypothesis Testing

Page: 33-133

Authors: Erica, Amy, Bryan, Carter, David
Page Rank

- Amy: $0+1$
- Bryan: $0+.5$
- Erica: $1+1$
- David: $0+1$
- Carter: $0+.5$
Page Rank

Erica

Amy

Bryan

David

Carter

0+1

0+.5

1+1

0+1

0+.5

1+1

0+.5

1+1

0+.5

0+1

0+1

0+.5
Page Rank
Graph Level Indices
Density

- Number of ties, expressed as a percentage of the number of possible ties
  - For directed graphs: \( \frac{E}{N(N-1)} \)
  - For undirected graphs: \( \frac{E}{N(N-1)/2} = \frac{2E}{3(N-1)} = \frac{4}{3} \)
Density

- Number of ties, expressed as a percentage of the number of possible ties
Density

- Number of ties, expressed as a percentage of the number of possible ties
- For directed graphs: \( \frac{E}{N(N-1)} \)
Density

- Number of ties, expressed as a percentage of the number of possible ties
- For directed graphs: \( \frac{E}{N(N-1)} \)
- For undirected graphs: \( \frac{E}{\frac{N(N-1)}{2}} \)
Density

- Number of ties, expressed as a percentage of the number of possible ties

- For directed graphs: \( \frac{E}{N(N-1)} \)

- For undirected graphs: \( \frac{E}{\frac{N(N-1)}{2}} \)
Density

- Number of ties, expressed as a percentage of the number of possible ties

- For directed graphs: \( \frac{E}{N(N-1)} \)

- For undirected graphs: \( \frac{E}{\frac{N(N-1)}{2}} \)

\[
\begin{align*}
\text{Density} & = \frac{2}{3(3-1)} = \frac{4}{6}
\end{align*}
\]
Graph Level Indices
Density

- Number of ties, expressed as a percentage of the number of possible ties
- For directed graphs: $E/N(N-1)$
- For undirected graphs: $E/N(N-1)/2 = 2(3-1)/2 = 4/6$
Density

- Number of ties, expressed as a percentage of the number of possible ties
Density

- Number of ties, expressed as a percentage of the number of possible ties
- For directed graphs: \( \frac{E}{N(N-1)} \)
Density

• Number of ties, expressed as a percentage of the number of possible ties

• For directed graphs: \( \frac{E}{N(N-1)} \)

• For undirected graphs: \( \frac{E}{2N(N-1)} \)
Density

- Number of ties, expressed as a percentage of the number of possible ties

- For directed graphs: $\frac{E}{N(N-1)}$

- For undirected graphs: $\frac{E}{\frac{N(N-1)}{2}}$
Density

- Number of ties, expressed as a percentage of the number of possible ties

- For directed graphs: \( \frac{E}{N(N-1)} \)

- For undirected graphs: \( \frac{E}{N(N-1)/2} \)

\[ \frac{2}{3(3-1)/2} = \frac{4}{6} \]
Mean Degree

The mean degree, $\bar{d}$, of all nodes in the graph is
Mean Degree

The mean degree, $\bar{d}$, of all nodes in the graph is

$$\bar{d}(n) = \frac{\sum_{i=1}^{N} d(n_i)}{N} = \frac{2E}{N}$$
Mean Degree

The mean degree, $\bar{d}$, of all nodes in the graph is

$$\bar{d}(n) = \frac{\sum_{i=1}^{N} d(n_i)}{N} = \frac{2E}{N}$$

$$= \frac{1+2+1}{3} = \frac{4}{3}$$
Size, Density, and Mean Degree

If we hold the mean degree constant, but vary size, what happens to density?
Size, Density, and Mean Degree

If we hold the mean degree constant, but vary size, what happens to density?
Centralization

• Extent to which centrality is concentrated on a single vertex
• Calculated as the sum of the differences between each node's centrality score and the maximum score
• Most centralized structure is usually a star network
Centralization

- Extent to which centrality is concentrated on a single vertex
Centralization

- **Extent to which centrality is concentrated on a single vertex**
- **Calculated as the sum of the differences between each node’s centrality score and the maximum score**
Centralization

- Extent to which centrality is concentrated on a single vertex
- Calculated as the sum of the differences between each node’s centrality score and the maximum score
- Most centralized structure is usually a star network
Bavelas Experiments

• People in positions passed messages to one another to solve a problem
• Studied the effect of structure on:
  • Efficiency
  • Leadership
  • Satisfaction
Bavelas Experiments

- People in positions passed messages to one another to solve a problem
Bavelas Experiments

- People in positions passed messages to one another to solve a problem
- Studied the effect of structure on
  - Efficiency
  - Leadership
  - Satisfaction
Bavelas Experiments

Circle

Line

Star

Y
Bavelas Experiments

Circle

Star
Bavelas Experiments

- Slowest to completion

Circle

Star
Bavelas Experiments

- Slowest to completion
- Most errors

Circle

Star
Bavelas Experiments

- Slowest to completion
- Most errors
- Most satisfied

Circle

Star
Bavelas Experiments

- Slowest to completion
- Most errors
- Most satisfied

Circle

- Fastest

Star
Bavelas Experiments

- Slowest to completion
- Most errors
- Most satisfied

Circle

- Fastest
- Fewest errors

Star
Bavelas Experiments

- Slowest to completion
- Most errors
- Most satisfied

- Fastest
- Fewest errors
- Most dissatisfied
Dyad Census
Dyad Census

Mutual (M)
Dyad Census

Mutual (M)

Assymmetric (A)
Dyad Census

- Mutual (M)

- Assymetric (A)

- Null (N)
Reciprocity

- **Dyadic**: the proportion of dyads that are symmetric
  \[ M + N / (M + A + N) \]

- **Dyadic non-null**: the proportion of non-null dyads that are reciprocal
  \[ M / (M + A) \]

- **Edgewise**: \[ 2 \times M / (2 	imes M + A) \]
Reciprocity

- Dyadic: the proportion of dyads that are symmetric

\[
\frac{M+N}{M+A+N}
\]
Reciprocity

- Dyadic: the proportion of dyads that are symmetric
  \[
  \frac{M + N}{M + A + N}
  \]
- Dyadic non-null: the proportion of non-null dyads that are reciprocal
  \[
  \frac{M}{M + A}
  \]
Reciprocity

- Dyadic: the proportion of dyads that are symmetric
  \[ \frac{M+N}{M+A+N} \]

- Dyadic non-null: the proportion of non-null dyads that are reciprocal
  \[ \frac{M}{M+A} \]

- Edgewise: \[ \frac{2*M}{2*M+A} \]
Triad Census
Triad Census

003 012 102 111D 201 210 300

021D 111U 120D

021U 030T 120U

021C 030C 120C
Triad Census

16 different triad types

One row per network
Triad Census

16 different triad types

i,j cell is the number of triad type j in network i

One row per network
FIGURE 2. Singular value decomposition of triad census array, first two left singular vectors, multiplied by singular values, $N = 82$ networks.
Triad Census

**FIGURE 2.** Singular value decomposition of triad census array, first two left singular vectors, multiplied by singular values, \( N = 82 \) networks.
Triad Census

**FIGURE 2.** Singular value decomposition of triad census array, first two left singular vectors, multiplied by singular values, $N = 82$ networks.
Transitivity

• Usually calculated as the fraction of completed two-paths
• Related to Granovetter’s ‘forbidden triad’
• Can be directed or undirected
Transitivity

- Usually calculated as the fraction of completed two-paths
- Related to Granovetter's 'forbidden triad'
- Can be directed or undirected
Transitivity

- Usually calculated as the fraction of completed two-paths
- Related to Granovetter's 'forbidden triad'
- Can be directed or undirected
Transitivity

- Usually calculated as the fraction of completed two-paths
- Related to Granovetter's 'forbidden triad'
- Can be directed or undirected
Transitivity

• Usually calculated as the fraction of completed two-paths
• Related to Granovetter's 'forbidden triad'
• Can be directed or undirected
Transitivity

- Usually calculated as the fraction of completed two-paths

![Diagram showing transitivity with nodes 1, 2, and 3 connected in a triangle]
Transitivity

- Usually calculated as the fraction of completed two-paths
- Related to Granovetter’s ‘forbidden triad’
Transitivity

- Usually calculated as the fraction of completed two-paths
- Related to Grannovetter’s ‘forbidden triad’
- Can be directed or undirected
Forbidden Triad or Structural Hole?

Fig. 1.—Forbidden triad

Granovetter, Mark S. 1973. "The Strength of Weak Ties"
Forbidden Triad or Structural Hole?


![Forbidden Triad Diagram](image-url)
Forbidden Triad or Structural Hole?

Forbidden Triad or Structural Hole?

Attributes!

 Extensions

• Properties of nodes, edges, or even networks
• Pretty much anything you can measure could be an attribute
• Extension based on node attributes: Brokerage
• Extension based on edge attributes: Structural Balance
Extensions

Attributes!

- Properties of nodes, edges, or even networks
Attributes!

- Properties of nodes, edges, or even networks
- Pretty much anything you can measure could be an attribute
Extensions

Attributes!

- Properties of nodes, edges, or even networks
- Pretty much anything you can measure could be an attribute
- Extension based on node attributes: Brokerage
Extensions

Attributes!

- Properties of nodes, edges, or even networks
- Pretty much anything you can measure could be an attribute
- Extension based on node attributes: Brokerage
- Extension based on edge attributes: Structural Balance
Brokerage

- Brokerage is a process “by which intermediary actors facilitate transactions between other actors lacking access to or trust in one another” (Marsden 1982)
Brokerage

- Brokerage is a process “by which intermediary actors facilitate transactions between other actors lacking access to or trust in one another” (Marsden 1982)
- Brokers play a crucial role in knitting together diverse groups of people, organizations, parties
Brokerage

- Brokerage is a process “by which intermediary actors facilitate transactions between other actors lacking access to or trust in one another” (Marsden 1982)
- Brokers play a crucial role in knitting together diverse groups of people, organizations, parties
- Brokers can gain a lot – early access to information, prestige
Brokerage

- Brokerage is a process “by which intermediary actors facilitate transactions between other actors lacking access to or trust in one another” (Marsden 1982)
- Brokers play a crucial role in knitting together diverse groups of people, organizations, parties
- Brokers can gain a lot – early access to information, prestige
- But can also be distrusted by everyone
Brokerage: Formal Concept

In a network $N$ with edges $E$, 

\[ \text{Diagram with three nodes interconnected by edges.} \]
Brokerage: Formal Concept

In a network $N$ with edges $E$, node $j$ brokers
Brokerage: Formal Concept

In a network $N$ with edges $E$, node $j$ brokers nodes $i$ and $k$. 
Brokerage: Formal Concept

In a network $N$ with edges $E$, node $j$ brokers nodes $i$ and $k$ if $e_{ij} \in E$.
Brokerage: Formal Concept

In a network $N$ with edges $E$, node $j$ brokers nodes $i$ and $k$ if $e_{ij} \in E$ and $e_{jk} \in E$.
Brokerage: Formal Concept

In a network $N$ with edges $E$, node $j$ brokers nodes $i$ and $k$ if $e_{ij} \in E$ and $e_{jk} \in E$ but $e_{ik} \notin E$.
Brokerage: Formal Concept

Brokerage: Formal Concept


- formalized the concept
Brokerage: Formal Concept


• formalized the concept
• added a vertex attribute component
Brokerage: Formal Concept


- formalized the concept
- added a vertex attribute component
- compared empirical brokerage counts to counts from random graphs conditioned on the number of edges
Brokerage: One Mode

Coordinator
Brokerage: One Mode

Coordinator

Representative
Brokerage: One Mode

Coordinator

Representative

Gatekeeper
Brokerage: One Mode

Coordinator
Representative
Gatekeeper
Itinerant
Brokerage: One Mode

Coordinator  Representative  Gatekeeper  Itinerant  Liaison
Gould and Fernandez’ Findings

• the benefits of brokerage are mediated both by the type of organization (the node sets) and the type of brokerage chain
• non-governmental organizations were found to have more influence when they held any type of brokerage position
• governmental organizations gained influence only when they held “outsider” brokerage roles in itinerant and liaison chains
Gould and Fernandez’ Findings

- the benefits of brokerage are mediated both by the type of organization (the node sets) and the type of brokerage chain
Gould and Fernandez’ Findings

- the benefits of brokerage are mediated both by the type of organization (the node sets) and the type of brokerage chain
- non-governmental organizations were found to have more influence when they held any type of brokerage position
Gould and Fernandez’ Findings

- the benefits of brokerage are mediated both by the type of organization (the node sets) and the type of brokerage chain
- non-governmental organizations were found to have more influence when they held any type of brokerage position
- governmental organizations gained influence only when they held “outsider” brokerage roles in itinerant and liaison chains
Structural Balance

(a) $A, B, \text{ and } C$ are mutual friends: balanced.

(b) $A$ is friends with $B \text{ and } C$, but they don't get along with each other: not balanced.

(c) $A \text{ and } B$ are friends with $C$ as a mutual enemy: balanced.

(d) $A, B, \text{ and } C$ are mutual enemies: not balanced.
Bernoulli ‘Random’ Graphs

Bernoulli 'Random' Graphs

- A probability distribution for a success/failure
- Best known example is the coin flip
Bernoulli ‘Random’ Graphs

- A probability distribution for a success/failure
Bernoulli ‘Random’ Graphs

- A probability distribution for a success/failure
- Best known example is the coin flip
Bernoulli ‘Random’ Graphs
Bernoulli ‘Random’ Graphs
Bernoulli ‘Random’ Graphs
Bernoulli ‘Random’ Graphs
Bernoulli ‘Random’ Graphs

---

SNA
LJasny
Intro
Data Structures
Descriptives
Hypothesis Testing
Bernoulli ‘Random’ Graphs
Bernoulli ‘Random’ Graphs
Bernoulli ‘Random’ Graphs
Bernoulli ‘Random’ Graphs
Bernoulli ‘Random’ Graphs

![Histogram of degree distribution for Bernoulli random graphs]
What are the characteristics of real world?
Small Worlds

Small Worlds

Watts-Strogatz Model

• high clustering coefficient
• low diameter
Watts-Strogatz Model

- high clustering coefficient
Watts-Strogatz Model

- high clustering coefficient
- low diameter
Clustering Coefficient and Diameter

• The diameter of a graph is the longest shortest path between any pair of nodes.
• The clustering coefficient of a graph is the number of triangles divided by the number of two-paths.
Clustering Coefficient and Diameter

- The **diameter** of a graph is the longest shortest path between any pair nodes
Clustering Coefficient and Diameter

- The *diameter* of a graph is the longest shortest path between any pair nodes.
- The *clustering coefficient* of a graph is the number of triangles divided by the number of two-paths.
Clustering Coefficient Tangent

• remember the triad census...
• your numerator is... the number of triads with 3 ties
• your denominator is... the number of triads with 2 ties PLUS 3*the number of triads with 3 ties
Clustering Coefficient Tangent

- remember the triad census...

\[
\frac{\text{# of triads with 3 ties}}{\text{# of triads with 2 ties} + 3 \times \text{# of triads with 3 ties}}
\]
Clustering Coefficient Tangent

- remember the triad census...
- your numerator is...
Clustering Coefficient Tangent

- remember the triad census...
- your numerator is... the number of triads with 3 ties
Cluster Coefficient Tangent

- remember the triad census...
- your numerator is... the number of triads with 3 ties
- your denominator is...
Clustering Coefficient Tangent

- remember the triad census...
- your numerator is... the number of triads with 3 ties
- your denominator is... the number of triads with 2 ties PLUS 3*the number of triads with 3 ties
Watts-Strogatz Model

- high clustering coefficient
- low diameter
Watts-Strogatz Model

- high clustering coefficient
- low diameter
- starts with a lattice structure
Watts-Strogatz Model

- high clustering coefficient
- low diameter
- starts with a lattice structure
- randomly re-wires ties
Watts-Strogatz Model

- Regular Network
- Small World Network
- Random Network

P=0 → Increasing Randomness → P=1
Watts-Strogatz Model
Watts-Strogatz Model

<table>
<thead>
<tr>
<th></th>
<th>Film Actors</th>
<th>Power Grid</th>
<th>C. Elegans</th>
</tr>
</thead>
<tbody>
<tr>
<td>actual</td>
<td>3.65</td>
<td>18.7</td>
<td>2.65</td>
</tr>
<tr>
<td>C</td>
<td>2.99</td>
<td>12.4</td>
<td>2.25</td>
</tr>
<tr>
<td>p</td>
<td>0.79</td>
<td>0.08</td>
<td>0.28</td>
</tr>
<tr>
<td>significance</td>
<td>0.00027</td>
<td>0.005</td>
<td>0.05</td>
</tr>
</tbody>
</table>
Watts-Strogatz Model

<table>
<thead>
<tr>
<th></th>
<th>$L_{actual}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Film Actors</td>
<td></td>
</tr>
<tr>
<td>Power Grid</td>
<td></td>
</tr>
<tr>
<td>C. Elegans</td>
<td></td>
</tr>
</tbody>
</table>
## Watts-Strogatz Model

<table>
<thead>
<tr>
<th>L\textsubscript{actual}</th>
<th>L\textsubscript{1}</th>
</tr>
</thead>
<tbody>
<tr>
<td>Film Actors</td>
<td></td>
</tr>
<tr>
<td>Power Grid</td>
<td></td>
</tr>
<tr>
<td>C. Elegans</td>
<td></td>
</tr>
</tbody>
</table>
## Watts-Strogatz Model

<table>
<thead>
<tr>
<th></th>
<th>$L_{\text{actual}}$</th>
<th>$L_1$</th>
<th>$C_{\text{actual}}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Film Actors</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Power Grid</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C. Elegans</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
# Watts-Strogatz Model

<table>
<thead>
<tr>
<th></th>
<th>$L_{\text{actual}}$</th>
<th>$L_1$</th>
<th>$C_{\text{actual}}$</th>
<th>$C_1$</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Film Actors</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Power Grid</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>C. Elegans</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
## Watts-Strogatz Model

<table>
<thead>
<tr>
<th></th>
<th>$L_{actual}$</th>
<th>$L_1$</th>
<th>$C_{actual}$</th>
<th>$C_1$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Film Actors</td>
<td>3.65</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Power Grid</td>
<td>18.7</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C. Elegans</td>
<td>2.65</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
## Watts-Strogatz Model

<table>
<thead>
<tr>
<th></th>
<th>$L_{\text{actual}}$</th>
<th>$L_1$</th>
<th>$C_{\text{actual}}$</th>
<th>$C_1$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Film Actors</td>
<td>3.65</td>
<td>2.99</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Power Grid</td>
<td>18.7</td>
<td>12.4</td>
<td></td>
<td></td>
</tr>
<tr>
<td>C. Elegans</td>
<td>2.65</td>
<td>2.25</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
## Watts-Strogatz Model

<table>
<thead>
<tr>
<th></th>
<th>$L_{actual}$</th>
<th>$L_1$</th>
<th>$C_{actual}$</th>
<th>$C_1$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Film Actors</td>
<td>3.65</td>
<td>2.99</td>
<td>0.79</td>
<td></td>
</tr>
<tr>
<td>Power Grid</td>
<td>18.7</td>
<td>12.4</td>
<td>0.08</td>
<td></td>
</tr>
<tr>
<td>C. Elegans</td>
<td>2.65</td>
<td>2.25</td>
<td>0.28</td>
<td></td>
</tr>
</tbody>
</table>
### Watts-Strogatz Model

<table>
<thead>
<tr>
<th></th>
<th>$L_{actual}$</th>
<th>$L_1$</th>
<th>$C_{actual}$</th>
<th>$C_1$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Film Actors</td>
<td>3.65</td>
<td>2.99</td>
<td>0.79</td>
<td>0.00027</td>
</tr>
<tr>
<td>Power Grid</td>
<td>18.7</td>
<td>12.4</td>
<td>0.08</td>
<td>0.005</td>
</tr>
<tr>
<td>C. Elegans</td>
<td>2.65</td>
<td>2.25</td>
<td>0.28</td>
<td>0.05</td>
</tr>
</tbody>
</table>
Watts-Strogatz Model

![Graph showing degree distribution](image-url)
Watts-Strogatz Model
Watts-Strogatz Model

![Graph showing degree distribution with frequency on the y-axis and degree on the x-axis.](image-url)
Barabasi-Albert Model

- Alternative network generating model
- Where Watts and Strogatz's model results in a world where everyone has approximately the same number of ties,
- Barabasi and Albert thought about a skewed distribution of ties
- Based on the idea of 'preferential attachment' aka 'rich get richer'
- Unlike Watts-Strogatz, this model starts with one node, add additional nodes one at a time
- Nodes 'preferentially' attach to those with higher degree
Barabasi-Albert Model

- Alternative network generating model
Barabasi-Albert Model

- Alternative network generating model
- Where Watts and Strogatz’s model results in a world where everyone has approximately the same number of ties,
Barabasi-Albert Model

- Alternative network generating model
- Where Watts and Strogatz’s model results in a world where everyone has approximately the same number of ties,
- Barabasi and Albert thought about a skewed distribution of ties
Barabasi-Albert Model

• Alternative network generating model
• Where Watts and Strogatz’s model results in a world where everyone has approximately the same number of ties,
• Barabasi and Albert thought about a skewed distribution of ties
• Based on the idea of ‘preferential attachment’ aka ’rich get richer’
Barabasi-Albert Model

- Alternative network generating model
- Where Watts and Strogatz’s model results in a world where everyone has approximately the same number of ties,
- Barabasi and Albert thought about a skewed distribution of ties
- Based on the idea of ‘preferential attachment’ aka ’rich get richer’
- Unlike Watts-Strogatz, this model starts with one node, add additional nodes one at a time
Barabasi-Albert Model

• Alternative network generating model
• Where Watts and Strogatz’s model results in a world where everyone has approximately the same number of ties,
• Barabasi and Albert thought about a skewed distribution of ties
• Based on the idea of ‘preferential attachment’ aka ’rich get richer’
• Unlike Watts-Strogatz, this model starts with one node, add additional nodes one at a time
• Nodes ‘preferentially’ attach to those with higher degree
Barabasi-Albert Model
Barabasi-Albert Model
Barabasi-Albert Model
Barabasi-Albert Model
Barabasi-Albert Model

![Diagram of Barabasi-Albert Model]
Barabasi-Albert Model

B
A
C
D

73-133
Barabasi-Albert Model
Barabasi-Albert Model
Barabasi-Albert Model

Code Time!

- Section 4
Hypothesis Testing
Relating Node level indices to covariates

- Node Level Indices: centrality measures, brokerage, constraint
Relating Node level indices to covariates

• Node Level Indices: centrality measures, brokerage, constraint

• Node Covariates: measures of power, career advancement, gender – really anything you want to study that varies at the node level
Emergent Multi-Organizational Networks (EMON) Dataset

• 7 case studies of EMONs in the context of search and rescue activities from Drabek et. al. (1981)
Emergent Multi-Organizational Networks (EMON) Dataset

• 7 case studies of EMONs in the context of search and rescue activities from Drabek et. al. (1981)
• Ties between organizations are self-reported levels of communication coded from 1 to 4 with 1 as most frequent
Emergent Multi-Organizational Networks (EMON) Dataset

Attribute Data

- Command Rank Score (CRS): mean rank (reversed) for prominence in the command structure
- Decision Rank Score (DRS): mean rank (reversed) for prominence in decision making process
- Paid Staff: number of paid employees
- Volunteer Staff: number of volunteer staff
- Sponsorship: organization type (City, County, State, Federal, or Private)
Correlation between DRS and Degree?

- Subsample of Mutually Reported "Continuous Communication" in Texas EMON
- Degree is shown in color (darker is bigger)
- DRS in size
- Empirical correlation $\rho = 0.86$
Correlation between DRS and Degree?

- Subsample of Mutually Reported “Continuous Communication” in Texas EMON
Correlation between DRS and Degree?

- Subsample of Mutually Reported “Continuous Communication” in Texas EMON
- Degree is shown in color (darker is bigger)
Correlation between DRS and Degree?

- Subsample of Mutually Reported “Continuous Communication” in Texas EMON
- Degree is shown in color (darker is bigger)
- DRS in size
Correlation between DRS and Degree?

- Subsample of Mutually Reported “Continuous Communication” in Texas EMON
- Degree is shown in color (darker is bigger)
- DRS in size
- Empirical correlation $\rho = 0.86$
Correlation between DRS and Degree?

\[ \rho = 0.86 \]
Correlation between DRS and Degree?

\[ \rho = 0.86 \]
Correlation between DRS and Degree?

\[ \rho = 0.86 \]

\[ \rho = -0.07 \]
Correlation between DRS and Degree?

$\rho = 0.86$

$\rho = -0.07$

$\rho = -0.12$

$\rho = -0.39$
Correlation between DRS and Degree?

\[ \rho = 0.86 \]

\[ \rho = -0.07 \]

\[ \rho = -0.12 \]
Correlation between DRS and Degree?

\[ \rho = 0.86 \]

\[ \rho = -0.07 \]

\[ \rho = -0.12 \]

\[ \rho = -0.39 \]
Correlation between DRS and Degree?

\[ \rho_{\text{obs}} = 0.86 \]

\[ \Pr(\rho \geq \rho_{\text{obs}}) = 3 \times 10^{-5} \]

\[ \Pr(\rho < \rho_{\text{obs}}) = 0.9999 \]
Correlation between DRS and Degree?

\[ \rho_{obs} = 0.86 \]
Correlation between DRS and Degree?

\[ \rho_{\text{obs}} = 0.86 \]

\[ Pr(\rho \geq \rho_{\text{obs}}) = 3e^{-5} \]
Correlation between DRS and Degree?

\[ Pr(\rho < \rho_{obs}) = 0.9999 \]

\[ Pr(\rho \geq \rho_{obs}) = 3e^{-5} \]
Regression?

• Can use Node Level Indices as independent variables in a regression
• Big assumption: position predicts the properties of those who hold them
• Conditioning on NLI values, so dependence in accounted for assuming no error in the network
• NLIs as dependent variables more problematic due to autocorrelation
• Can use Node Level Indices as independent variables in a regression
Regression?

- Can use Node Level Indices as independent variables in a regression
- Big assumption: *position* predicts the *properties of those who hold them*
Regression?

- Can use Node Level Indices as independent variables in a regression
- Big assumption: *position* predicts the *properties of those who hold them*
- Conditioning on NLI values, so dependence in accounted for *assuming no error in the network*
Regression?

- Can use Node Level Indices as independent variables in a regression
- Big assumption: *position* predicts the *properties of those who hold them*
- Conditioning on NLI values, so dependence in accounted for *assuming no error in the network*
- NLIs as dependent variables more problematic due to autocorrelation
Code Time!

• 5.1-5.3
 Quadratic Assignment Procedure

Marriage

Business
Quadratic Assignment Procedure

Marriage

Business
Graph Correlation

• Simple way of comparing graphs on the same vertex set by element

\[ g_{cor}(\begin{bmatrix} 1 & 1 \\ 1 & 0 \end{bmatrix}, \begin{bmatrix} 1 & 1 \\ 2 & 2 \end{bmatrix}) = cor(\begin{bmatrix} 1 & 1 \\ 1 & 0 \end{bmatrix}, \begin{bmatrix} 1 & 1 \\ 2 & 2 \end{bmatrix}) \]
Graph Correlation

- Simple way of comparing graphs on the same vertex set by element
Graph Correlation

- Simple way of comparing graphs on the same vertex set by element
- \( gcor \left( \begin{bmatrix} 1 & 1 \\ 1 & 0 \end{bmatrix}, \begin{bmatrix} 1 & 1 \\ 2 & 2 \end{bmatrix} \right) = cor([1, 1, 1, 0], [1, 1, 2, 2]) \)
Do business ties coincide with marriages?

\[ \rho = 0.372 \]
Do business ties coincide with marriages?

Marriage

Business
Do business ties coincide with marriages?

\[ \rho = 0.169 \]
Do business ties coincide with marriages?

\[
\rho = -0.034
\]
Do business ties coincide with marriages?

Marriage

\[ \rho = -0.101 \]

Business
QAP Test

Estimated Density of QAP Replications

Test Statistic
QAP Test

$\rho_{\text{obs}} = 0.372$
QAP Test

\[ \rho_{\text{obs}} = 0.372 \]

\[ Pr(\rho \geq \rho_{\text{obs}}) = 0.001 \]
QAP Test

\[ Pr(\rho < \rho_{obs}) = 0.999 \]

\[ \rho_{obs} = 0.372 \]

\[ Pr(\rho \geq \rho_{obs}) = .001 \]
Network Regression

• Family of models predicting social ties
• Special case of standard OLS regression
• Dependent variable is a network adjacency matrix

\[ E_{ij} = \beta_0 + \beta_1 X_{1ij} + \beta_2 X_{2ij} + \cdots + \beta_\rho X_{\rho ij} \]

- Where \( E \) is the expectation operator (analagous to "mean" or "average")
- \( Y_{ij} \) is the value from \( i \) to \( j \) on the dependent relation with adjacency matrix \( Y \)
- \( X_{kij} \) is the value of the \( k \)th predictor for the \((i,j)\) ordered pair, and \( \beta_0, \ldots, \beta_\rho \) are coefficients
Network Regression

- Family of models predicting social ties
Network Regression

- Family of models predicting social ties
  - Special case of standard OLS regression
Network Regression

- Family of models predicting social ties
  - Special case of standard OLS regression
  - Dependent variable is a network adjacency matrix
Network Regression

- Family of models predicting social ties
  - Special case of standard OLS regression
  - Dependent variable is a network adjacency matrix
- $\mathbb{E}Y_{ij} = \beta_0 + \beta_1 X_{1ij} + \beta_2 X_{2ij} + \cdots + \beta_\rho X_{\rho ij}$

- Where $\mathbb{E}$ is the expectation operator (analagous to "mean" or "average")
- $Y_{ij}$ is the value from $i$ to $j$ on the dependent relation with adjacency matrix $Y$
- $X_{kij}$ is the value of the $k$th predictor for the $(i,j)$ ordered pair, and $\beta_0, \ldots, \beta_\rho$ are coefficients
Network Regression

- Family of models predicting social ties
  - Special case of standard OLS regression
  - Dependent variable is a network adjacency matrix
- $EY_{ij} = \beta_0 + \beta_1 X_{1ij} + \beta_2 X_{2ij} + \cdots + \beta_{\rho} X_{\rho ij}$
  - Where $E$ is the expectation operator (analogous to “mean” or “average”)
Network Regression

- Family of models predicting social ties
  - Special case of standard OLS regression
  - Dependent variable is a network adjacency matrix
- $\mathbf{E}Y_{ij} = \beta_0 + \beta_1 X_{1ij} + \beta_2 X_{2ij} + \cdots + \beta_\rho X_{\rho ij}$
  - Where $\mathbf{E}$ is the expectation operator (analogous to “mean” or “average”)
  - $Y_{ij}$ is the value from $i$ to $j$ on the dependent relation with adjacency matrix $Y$
Network Regression

- Family of models predicting social ties
  - Special case of standard OLS regression
  - Dependent variable is a network adjacency matrix

\[ \mathbf{E}Y_{ij} = \beta_0 + \beta_1 X_{1ij} + \beta_2 X_{2ij} + \cdots + \beta_\rho X_{\rho ij} \]

- Where \( \mathbf{E} \) is the expectation operator (analogous to “mean” or “average”)
- \( Y_{ij} \) is the value from \( i \) to \( j \) on the dependent relation with adjacency matrix \( Y \)
- \( X_{kij} \) is the value of the \( k \)th predictor for the \((i, j)\) ordered pair, and \( \beta_0, \ldots, \beta_\rho \) are coefficients
Data Prep

• Dependent variable is an adjacency matrix
• Standard case: dichotomous data
• Valued case
• Independent variables also in adjacency matrix form
• Always takes matrix form, but may be based on vector data
• eg. simple adjacency matrix, sender/receiver effects, attribute differences, elements held in common
Data Prep

- Dependent variable is an adjacency matrix
Data Prep

- Dependent variable is an adjacency matrix
  - Standard case: dichotomous data
• Dependent variable is an adjacency matrix
  • Standard case: dichotomous data
  • Valued case
• Dependent variable is an adjacency matrix
  • Standard case: dichotomous data
  • Valued case
• Independent variables also in adjacency matrix form
Data Prep

- Dependent variable is an adjacency matrix
  - Standard case: dichotomous data
  - Valued case
- Independent variables also in adjacency matrix form
  - Always takes matrix form, but may be based on vector data
Data Prep

- **Dependent variable is an adjacency matrix**
  - Standard case: dichotomous data
  - Valued case

- **Independent variables also in adjacency matrix form**
  - Always takes matrix form, but may be based on vector data
  - eg. simple adjacency matrix, sender/receiver effects, attribute differences, elements held in common
Code Time!

• Sections 5.4 and 5.5
Network Autocorrelation Models

• Family of models for estimating how covariates relate to each other through ties
• Special case of standard OLS regression
• Dependent variable is a vertex attribute

\[ y = (I - \Theta W)^{-1} (X\beta + (I - \psi Z)^{-1} v) \]

• where \( \Theta \) is the matrix for the Auto-Regressive weights
• and \( \psi \) is the matrix for the Moving Average weights
Network Autocorrelation Models

- Family of models for estimating how covariates relate to each other through ties
Network Autocorrelation Models

- Family of models for estimating how covariates relate to each other through ties
  - Special case of standard OLS regression

\[ y = (I - \Theta W)^{-1} (X\beta + (I - \psi Z)^{-1} v) \]

- where \( \Theta \) is the matrix for the Auto-Regressive weights
- and \( \psi \) is the matrix for the Moving Average weights
Network Autocorrelation Models

- Family of models for estimating how covariates relate to each other through ties
  - Special case of standard OLS regression
  - Dependent variable is a vertex attribute
Network Autocorrelation Models

- Family of models for estimating how covariates relate to each other through ties
  - Special case of standard OLS regression
  - Dependent variable is a vertex attribute
- \( y = (I - \Theta W)^{-1}(X\beta + (I - \psi Z)^{-1}v) \)
Network Autocorrelation Models

- Family of models for estimating how covariates relate to each other through ties
  - Special case of standard OLS regression
  - Dependent variable is a vertex attribute
- \( y = (I - \Theta W)^{-1} (X \beta + (I - \psi Z)^{-1}v) \)
  - where \( \Theta \) is the matrix for the Auto-Regressive weights
Network Autocorrelation Models

- Family of models for estimating how covariates relate to each other through ties
  - Special case of standard OLS regression
  - Dependent variable is a vertex attribute
- \( y = (I - \Theta W)^{-1}(X\beta + (I - \psi Z)^{-1}v) \)
  - where \( \Theta \) is the matrix for the Auto-Regressive weights
  - and \( \psi \) is the matrix for the Moving Average weights
The Classical Regression Model
The Classical Regression Model

\[ \mathbf{X}_i \beta \]
The Classical Regression Model

\[ y_i = X_i \beta + \epsilon_i \]
The Classical Regression Model

\[ Y_i = X_i \beta + \epsilon_i \]
Adding Network AR Effects

\[ y_i = X_i \beta + \epsilon_i \]

\[ y_j = X_j \beta + \epsilon_j \]

\[ y_k = X_k \beta + \epsilon_k \]
Adding Network AR Effects

\[ y_i = X_i \beta + \epsilon_i \]

\[ y_j = X_j \beta + \epsilon_j \]

\[ y_k = X_k \beta + \epsilon_k \]
Adding Network AR Effects

$$y_i = X_i \beta + \epsilon_i$$

$$y_j = X_j \beta + \epsilon_j$$

$$y_k = X_k \beta + \epsilon_k$$
Adding Network MA Effects

\[ y_i = X_i \beta + \epsilon_i \]

\[ y_j = X_j \beta + \epsilon_j \]

\[ y_k = X_k \beta + \epsilon_k \]
Adding Network MA Effects

\[ Y_i = X_i \beta + \epsilon_i \]

\[ Y_j = X_j \beta + \epsilon_j \]

\[ Y_k = X_k \beta + \epsilon_k \]
Network ARMA Model

\[ y_i = X_i \beta + \epsilon_i \]

\[ y_j = X_j \beta + \epsilon_j \]

\[ y_k = X_k \beta + \epsilon_k \]
Network ‘Resonance’
Network ‘Resonance’
Network ‘Resonance’

\[ \epsilon_k \quad \epsilon_i \quad \epsilon_j \quad \epsilon_l \]
Network ‘Resonance’

$\epsilon_i \ldots$  

$\epsilon_j \ldots$  

$\epsilon_k \ldots$  

$\epsilon_l \ldots$
Network ‘Resonance’
Network ‘Resonance’
Inference with the Network Autocorrelation Model

• Usually observe $y$, $X$, and $Z$, and/or $Z$
• Want to infer $\beta$, $\theta$, and $\phi$
• Need each $I-W$, $I-Z$ invertible for solution to exist
• Error in disturbance autocorrelation, $v$, assumed as iid, $v_i \sim N(0,\sigma^2)$
• Standard errors based on the inverse information matrix at the MLE
• Compare models in the usual way (e.g., AIC, BIC)
Inference with the Network Autocorrelation Model

- Usually observe $y$, $X$, and $Z$ and/or $Z$, want to infer $\beta$, $\theta$, and $\phi$
Inference with the Network Autocorrelation Model

- Usually observe $y$, $X$, and $Z$ and/or $Z$, want to infer $\beta$, $\theta$, and $\phi$
- Need each $I - W$, $I - Z$ invertible for solution to exist
Inference with the Network Autocorrelation Model

- Usually observe $y$, $X$, and $Z$ and/or $Z$, want to infer $\beta$, $\theta$, and $\phi$
- Need each $I - W$, $I - Z$ invertible for solution to exist
- error in disturbance autocorrelation, $v$, assumed as iid, $v_i \sim N(0, \sigma^2)$
Inference with the Network Autocorrelation Model

- Usually observe $y$, $X$, and $Z$ and/or $Z$, want to infer $\beta$, $\theta$, and $\phi$
- Need each $I - W$, $I - Z$ invertible for solution to exist
- error in disturbance autocorrelation, $v$, assumed as iid, $v_i \sim N(0, \sigma^2)$
- Standard errors based on the inverse information matrix at the MLE
Inference with the Network Autocorrelation Model

• Usually observe $y$, $X$, and $Z$ and/or $Z$, want to infer $\beta$, $\theta$, and $\phi$

• Need each $I - W$, $I - Z$ invertible for solution to exist

• error in disturbance autocorrelation, $v$, assumed as iid, $v_i \sim N(0, \sigma^2)$

• Standard errors based on the inverse information matrix at the MLE

• Compare models in the usual way (eg AIC, BIC)
Choosing the Weight Matrix

• crucial modeling issue to choose the right form
• standard adjacency matrix
• row-normalized adjacency matrix
• structural equivalence distance

Many suggestions given by Leenders 2002
Choosing the Weight Matrix

- crucial modeling issue to choose the right form
Choosing the Weight Matrix

- crucial modeling issue to choose the right form
  - standard adjacency matrix

Many suggestions given by Leenders 2002
Choosing the Weight Matrix

- crucial modeling issue to choose the right form
  - standard adjacency matrix
  - row-normalized adjacency matrix
Choosing the Weight Matrix

- crucial modeling issue to choose the right form
  - standard adjacency matrix
  - row-normalized adjacency matrix
  - structural equivalence distance
Choosing the Weight Matrix

- crucial modeling issue to choose the right form
  - standard adjacency matrix
  - row-normalized adjacency matrix
  - structural equivalence distance
- Many suggestions given by Leenders 2002
Data Prep

- Dependent variable is a vertex attribute
- Covariates are in matrix form with one column per attribute
- Can include an intercept term by adding a column of 1s
- Weight matrices for both AR and MA terms in matrix form
- Can include multiple weight matrices (as a list) for both AR and MA
Data Prep

- Dependent variable is a vertex attribute
Data Prep

- Dependent variable is a vertex attribute
- Covariates are in matrix form with one column per attribute
Data Prep

- Dependent variable is a vertex attribute
- Covariates are in matrix form with one column per attribute
- Can include an intercept term by adding a column of 1s
• Dependent variable is a vertex attribute
• Covariates are in matrix form with one column per attribute
• Can include an intercept term by adding a column of 1s
• Weight matrices for both AR and MA terms in matrix form
Data Prep

- Dependent variable is a vertex attribute
- Covariates are in matrix form with one column per attribute
- Can include an intercept term by adding a column of 1s
- Weight matrices for both AR and MA terms in matrix form
- Can include multiple weight matrices (as a list) for both AR and MA
Leenders 2002
Variables

• Dependent variable: proportion of support in a parish for democratic presidential candidate Kennedy in the 1960 elections
Variables

- Dependent variable: proportion of support in a parish for democratic presidential candidate Kennedy in the 1960 elections
- Covariates:
Variables

- Dependent variable: proportion of support in a parish for democratic presidential candidate Kennedy in the 1960 elections
- Covariates:
  - $B$ is the percentage of African American residents in the parish
Variables

- Dependent variable: proportion of support in a parish for democratic presidential candidate Kennedy in the 1960 elections
- Covariates:
  - $B$ is the percentage of African American residents in the parish
  - $C$ is the percentage of Catholic residents in the parish
Variables

- Dependent variable: proportion of support in a parish for democratic presidential candidate Kennedy in the 1960 elections
- Covariates:
  - $B$ is the percentage of African American residents in the parish
  - $C$ is the percentage of Catholic residents in the parish
  - $U$ is the percentage of the parish considered urban
Variables

- Dependent variable: proportion of support in a parish for democratic presidential candidate Kennedy in the 1960 elections
- Covariates:
  - $B$ is the percentage of African American residents in the parish
  - $C$ is the percentage of Catholic residents in the parish
  - $U$ is the percentage of the parish considered urban
  - $BPE$ is a measure of 'black political equality'
Variables

- Dependent variable: proportion of support in a parish for democratic presidential candidate Kennedy in the 1960 elections
- Covariates:
  - $B$ is the percentage of African American residents in the parish
  - $C$ is the percentage of Catholic residents in the parish
  - $U$ is the percentage of the parish considered urban
  - $BPE$ is a measure of 'black political equality'
- Weight matrix ($\rho$): simple contiguity network
Table 3
Network effects model for the Louisiana voting data

<table>
<thead>
<tr>
<th></th>
<th>OLS</th>
<th>( w_{ij}^{[1]} )</th>
<th>( w_{ij}^{[2]} )</th>
<th>( w_{ij}^{[6]} )</th>
<th>( w_{ij}^{[9]} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \rho )</td>
<td>-</td>
<td>0.31* (0.10)</td>
<td>0.07 (0.06)</td>
<td>0.12 (0.25)</td>
<td>0.04 (0.12)</td>
</tr>
<tr>
<td>Constant</td>
<td>21.03* (4.40)</td>
<td>13.87* (4.67)</td>
<td>19.83* (4.34)</td>
<td>16.78 (10.06)</td>
<td>19.80* (5.62)</td>
</tr>
<tr>
<td>B</td>
<td>0.01 (0.08)</td>
<td>-0.00 (0.07)</td>
<td>0.00 (0.08)</td>
<td>0.01 (0.08)</td>
<td>0.01 (0.08)</td>
</tr>
<tr>
<td>C</td>
<td>0.30* (0.04)</td>
<td>0.22* (0.05)</td>
<td>0.28* (0.04)</td>
<td>0.29* (0.05)</td>
<td>0.29 (0.05)</td>
</tr>
<tr>
<td>U</td>
<td>-0.11* (0.04)</td>
<td>-0.10* (0.04)</td>
<td>-0.11* (0.04)</td>
<td>-0.11* (0.04)</td>
<td>-0.11* (0.04)</td>
</tr>
<tr>
<td>BPE</td>
<td>0.39* (0.06)</td>
<td>0.30* (0.06)</td>
<td>0.37* (0.06)</td>
<td>0.38* (0.06)</td>
<td>0.38* (0.06)</td>
</tr>
</tbody>
</table>

* \( P < 0.05 \).

Table 4
Network disturbances model for the Louisiana voting data

<table>
<thead>
<tr>
<th></th>
<th>( w_{ij}^{[1]} )</th>
<th>( w_{ij}^{[2]} )</th>
<th>( w_{ij}^{[6]} )</th>
<th>( w_{ij}^{[9]} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \rho )</td>
<td>0.69* (0.10)</td>
<td>0.53* (0.13)</td>
<td>0.22 (0.42)</td>
<td>0.74* (0.15)</td>
</tr>
<tr>
<td>Constant</td>
<td>26.99* (4.50)</td>
<td>24.98* (4.22)</td>
<td>21.52* (4.30)</td>
<td>24.51* (5.06)</td>
</tr>
<tr>
<td>B</td>
<td>-0.11 (0.07)</td>
<td>-0.07 (0.07)</td>
<td>-0.00 (0.08)</td>
<td>-0.09 (0.08)</td>
</tr>
<tr>
<td>C</td>
<td>0.37* (0.05)</td>
<td>0.35* (0.04)</td>
<td>0.31* (0.04)</td>
<td>0.38* (0.04)</td>
</tr>
<tr>
<td>U</td>
<td>-0.07* (0.03)</td>
<td>0.08* (0.03)</td>
<td>-0.11* (0.04)</td>
<td>-0.10* (0.04)</td>
</tr>
<tr>
<td>BPE</td>
<td>0.24* (0.06)</td>
<td>0.30* (0.06)</td>
<td>0.38* (0.06)</td>
<td>0.29* (0.06)</td>
</tr>
</tbody>
</table>

* \( P < 0.05 \).
<table>
<thead>
<tr>
<th>Weight matrix</th>
<th>AIC</th>
<th>Order within model</th>
<th>Overall order</th>
</tr>
</thead>
<tbody>
<tr>
<td>(w_{ij}^{[1]})</td>
<td>439.12</td>
<td>1</td>
<td>3</td>
</tr>
<tr>
<td>(w_{ij}^{[2]})</td>
<td>445.52</td>
<td>2</td>
<td>5</td>
</tr>
<tr>
<td>(w_{ij}^{[9]})</td>
<td>446.78</td>
<td>4</td>
<td>8</td>
</tr>
<tr>
<td>(w_{ij}^{[6]})</td>
<td>446.44</td>
<td>3</td>
<td>6</td>
</tr>
<tr>
<td>Network disturbances model</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(w_{ij}^{[1]})</td>
<td>431.92</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>(w_{ij}^{[2]})</td>
<td>436.33</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>(w_{ij}^{[9]})</td>
<td>446.69</td>
<td>4</td>
<td>7</td>
</tr>
<tr>
<td>(w_{ij}^{[6]})</td>
<td>440.95</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>OLS</td>
<td>–</td>
<td>–</td>
<td>9</td>
</tr>
</tbody>
</table>
Code Time!

- Section 5.6
Baseline Models

• treats social structure as maximally random given some fixed constraints
• methodological premise from Mayhew
• identify potentially constraining factors
• compare observed properties to baseline model
• useful even when baseline model is not ‘realistic’
Baseline Models

- treats social structure as maximally random given some fixed constraints
Baseline Models

- treats social structure as maximally random given some fixed constraints
- methodological premise from Mayhew
Baseline Models

- treats social structure as maximally random given some fixed constraints
- methodological premise from Mayhew
  - identify potentially constraining factors
Baseline Models

- treats social structure as maximally random given some fixed constraints
- methodological premise from Mayhew
  - identify potentially constraining factors
  - compare observed properties to baseline model
Baseline Models

- treats social structure as maximally random given some fixed constraints
- methodological premise from Mayhew
  - identify potentially constraining factors
  - compare observed properties to baseline model
  - useful even when baseline model is not ‘realistic’
Types of Baseline Hypotheses
Types of Baseline Hypotheses

Empirical Network
Types of Baseline Hypotheses

Empirical Network
Types of Baseline Hypotheses

Empirical Network
Types of Baseline Hypotheses

Empirical Network

... etc
Types of Baseline Hypotheses

Empirical Network
Types of Baseline Hypotheses

Empirical Network
Types of Baseline Hypotheses

Empirical Network
Types of Baseline Hypotheses

Empirical Network

... etc
Types of Baseline Models

• Size: given the number of individuals, all structures are equally likely
• Number of edges/probability of an edge: given the number of individuals and interactions (aka Erdős-Rényi random graphs)
• Dyad census: given number of individuals, mutuals, asymmetric, and null relationships
• Degree distribution: given the number of individuals and each individual's outgoing/incoming ties
• Number of triangles: not implemented due to complexity – with ERGM, can condition on the expected number of triangles
Types of Baseline Models

- **Size**: given the number of individuals, all structures are equally likely
Types of Baseline Models

- **Size**: given the number of individuals, all structures are equally likely
- **Number of edges/probability of an edge**: given the number of individuals and interactions (aka Erdös-Renyi random graphs)
Types of Baseline Models

- **Size**: given the number of individuals, all structures are equally likely
- **Number of edges/probability of an edge**: given the number of individuals and interactions (aka Erdős-Renyi random graphs)
- **Dyad census**: given number of individuals, mutuals, asymmetric, and null relationships
Types of Baseline Models

- **Size:** given the number of individuals, all structures are equally likely
- **Number of edges/probability of an edge:** given the number of individuals and interactions (aka Erdös-Renyi random graphs)
- **Dyad census:** given number of individuals, mutuals, asymmetric, and null relationships
- **Degree distribution:** given the number of individuals and each individual’s outgoing/incoming ties
Types of Baseline Models

- **Size:** given the number of individuals, all structures are equally likely
- **Number of edges/probability of an edge:** given the number of individuals and interactions (aka Erdös-Renyi random graphs)
- **Dyad census:** given number of individuals, mutuals, asymmetric, and null relationships
- **Degree distribution:** given the number of individuals and each individual’s outgoing/incoming ties
- **Number of triangles:** not implemented due to complexity – with ERGM, can condition on the *expected* number of triangles
Method

• Select a test statistic (graph correlation, reciprocity, transitivity...)
• Select a baseline hypothesis (what you're conditioning on)
• Simulate from the baseline hypothesis
• For each simulation, recalculate the test statistic
• Compare empirical value to null distribution, just as in standard statistical testing
Method

- Select a test statistic (graph correlation, reciprocity, transitivity...)

118-133
Method

• Select a test statistic (graph correlation, reciprocity, transitivity...)
• Select a baseline hypothesis (what you’re conditioning on)
Method

- Select a test statistic (graph correlation, reciprocity, transitivity...)
- Select a baseline hypothesis (what you’re conditioning on)
- Simulate from the baseline hypothesis
Method

• Select a test statistic (graph correlation, reciprocity, transitivity . . . )
• Select a baseline hypothesis (what you’re conditioning on)
• Simulate from the baseline hypothesis
• For each simulation, recalculate the test statistic
Method

- Select a test statistic (graph correlation, reciprocity, transitivity...)
- Select a baseline hypothesis (what you’re conditioning on)
- Simulate from the baseline hypothesis
- For each simulation, recalculate the test statistic
- Compare empirical value to null distribution, just as in standard statistical testing
Transitivity in the Hurricane Frederic EMON
Transitivity in the Hurricane Frederic EMON

\[ \rho = 0.475 \]

indicates that roughly half the time that \( i \rightarrow j \rightarrow k \), \( i \rightarrow j \).
Transitivity in the Hurricane Frederic EMON

- $\rho = 0.475$
- indicates that roughly half the time that $i \rightarrow j \rightarrow k$, $i \rightarrow j$
Example

Univariate CUG Test

CUG Replicates
Conditioning: size Reps: 1000
Univariate CUG Test

CUG Replicates
Conditioning: edges Reps: 1000
Caution!

• Your selection of baseline model controls what hypothesis you're testing.
• Changing the model can greatly change the results.
Caution!

- Your selection of baseline model controls what hypothesis you’re testing
- Changing the model can greatly change the results
Donations and voting patterns in the US 111th House of Representatives
Example

Donations and voting patterns in the US 111th House of Representatives

- Two two-mode matrices
Donations and voting patterns in the US 111th House of Representatives

• Two two-mode matrices
  • Representatives by donors
  • Representatives by votes
Donations and voting patterns in the US 111th House of Representatives

- Two two-mode matrices
  - Representatives by donors
  - Representatives by votes
- Convert to one-mode matrices
Example

Donations and voting patterns in the US 111th House of Representatives

- Two two-mode matrices
  - Representatives by donors
  - Representatives by votes
- Convert to one-mode matrices
  - Similarity in donations among representatives
  - Similarity in voting among representatives
Example

Donations and voting patterns in the US 111th House of Representatives

- Two two-mode matrices
  - Representatives by donors
  - Representatives by votes
- Convert to one-mode matrices
  - Similarity in donations among representatives
  - Similarity in voting among representatives
- What is the correlation between donations and voting?
• Consider the baseline model conditioning on degree distribution
Example

- Consider the baseline model conditioning on degree distribution
- In the two mode case this conditions on:
Example

- Consider the baseline model conditioning on degree distribution
- In the two mode case this conditions on:
  - the number of donations each donor makes
  - the number of donations each Representative receives
Example

- Consider the baseline model conditioning on degree distribution
- In the two mode case this conditions on:
  - the number of donations each donor makes
  - the number of donations each Representative receives
- In the one mode case:
Example

- Consider the baseline model conditioning on degree distribution
- In the two mode case this conditions on:
  - the number of donations each donor makes
  - the number of donations each Representative receives
- In the one mode case:
  - the similarity in donations received
Example

Lorien Jasny
Baseline Models for Two Mode Social Network Data
Example

Lorien Jasny
Baseline Models for Two Mode Social Network Data
Studying Network Dynamics with MDS
Hamming Distance

• Distance between two matrices, A and B, is equal to the number of dyads that would need to change for A to be equivalent to B
• Need a one-to-one mapping of vertices in A and B

Distance = 2
Hamming Distance

- Distance between two matrices, A and B, is equal to the number of dyads that would need to change for A to be equivalent to B
Hamming Distance

- Distance between two matrices, A and B, is equal to the number of dyads that would need to change for A to be equivalent to B
- Need a one-to-one mapping of vertices in A and B
Hamming Distance

- Distance between two matrices, A and B, is equal to the number of dyads that would need to change for A to be equivalent to B
- Need a one-to-one mapping of vertices in A and B
Hamming Distance

- Distance between two matrices, A and B, is equal to the number of dyads that would need to change for A to be equivalent to B
- Need a one-to-one mapping of vertices in A and B
Hamming Distance

- Distance between two matrices, A and B, is equal to the number of dyads that would need to change for A to be equivalent to B
- Need a one-to-one mapping of vertices in A and B
Hamming Distance

- Distance between two matrices, A and B, is equal to the number of dyads that would need to change for A to be equivalent to B
- Need a one-to-one mapping of vertices in A and B

Distance = 2
Example: Hamming Distance

One column per network

One row per network
Example: Hamming Distance

One column per network

\[
\begin{array}{c}
i,j \text{ cell is the} \\
\text{Hamming Distance} \\
between networks i and j
\end{array}
\]
Example: Hamming Distance
Example: Hamming Distance

- reveals qualitative dynamics
Example: Hamming Distance

- reveals qualitative dynamics
Example: Hamming Distance

- reveals qualitative dynamics
Example: Hamming Distance

- reveals qualitative dynamics
- pace of change
Example: Hamming Distance

- reveals qualitative dynamics
- pace of change
Example: Hamming Distance

- reveals qualitative dynamics
- pace of change
Example: Hamming Distance

- don’t over interpret curvature
- works well with valued data

Butts and Cross
Change and External Events
Journal of Social Structure, 2009
Code Time!

- Sections 6 and 7