Network-Based Statistical Insights into Speed Dating Match Success

Statistics Colloquium, Fall 2024

Lucy Ly • December 09, 2024

Contents

Background

Overview

Original Purpose

- Design an experiment to examine dating behavior differences between men and women

GENDER DIFFERENCES IN MATE SELECTION: EVIDENCE FROM A SPEED DATING EXPERIMENT*

RAYMOND FISMAN
SHEENA S. IYENGAR
EMIR KAMENICA
ITAMAR SIMONSON

We study dating behavior using data from a Speed Dating experiment where we generate random matching of subjects and create random variation in the number of potential partners. Our design allows us to directly observe individual decisions rather than just final matches. Women put greater weight on the intelligence and the race of partner, while men respond more to physical attractiveness. Moreover, men do not value women's intelligence or ambition when it exceeds their own. Also, we find that women exhibit a preference for men who grew up in affluent neighborhoods. Finally, male selectivity is invariant to group size, while female selectivity is strongly increasing in group size.

I. INTRODUCTION

The choice of a marriage partner is one of the most serious decisions people face. In contemporary Western societies, this decision usually follows a long learning period during which people engage in more informal and often polygamous relationships, i.e., dating, which is the topic of this paper. In particular, we analyze gender differences in dating preferences. As in all matching markets, determining dating preferences from equilibrium outcomes is difficult because a given correlation of attributes across partners is often consistent with various preference structures. We overcome this problem by studying dating behavior using an experimental Speed Dating market. In our experimental paradigm, subjects meet a number of potential mates (between 9 and 21, a number determined by the experimenters) for four minutes each, and have the opportunity to accept or reject each partner.[1] If both parties desire a future

* We are grateful to Lawrence Katz, Edward Glaeser, and three anonymous referees for valuable suggestions. We would also like to thank Matthew Gentzkow, David Laibson, Jesse Shapiro, and participants at seminars at Harvard University, Massachusetts Institute of Technology, and Stanford Institute for Theoretical Economics for insightful comments. Kamenica acknowledges support by the National Science Foundation Graduate Research Fellowship and the National Institute on Aging (Grant No. T32-AG00186). We are solely responsible for all mistakes.

1. Throughout the paper we will refer to the individual making the decision as *subject,* and the person being decided upon as *partner.*

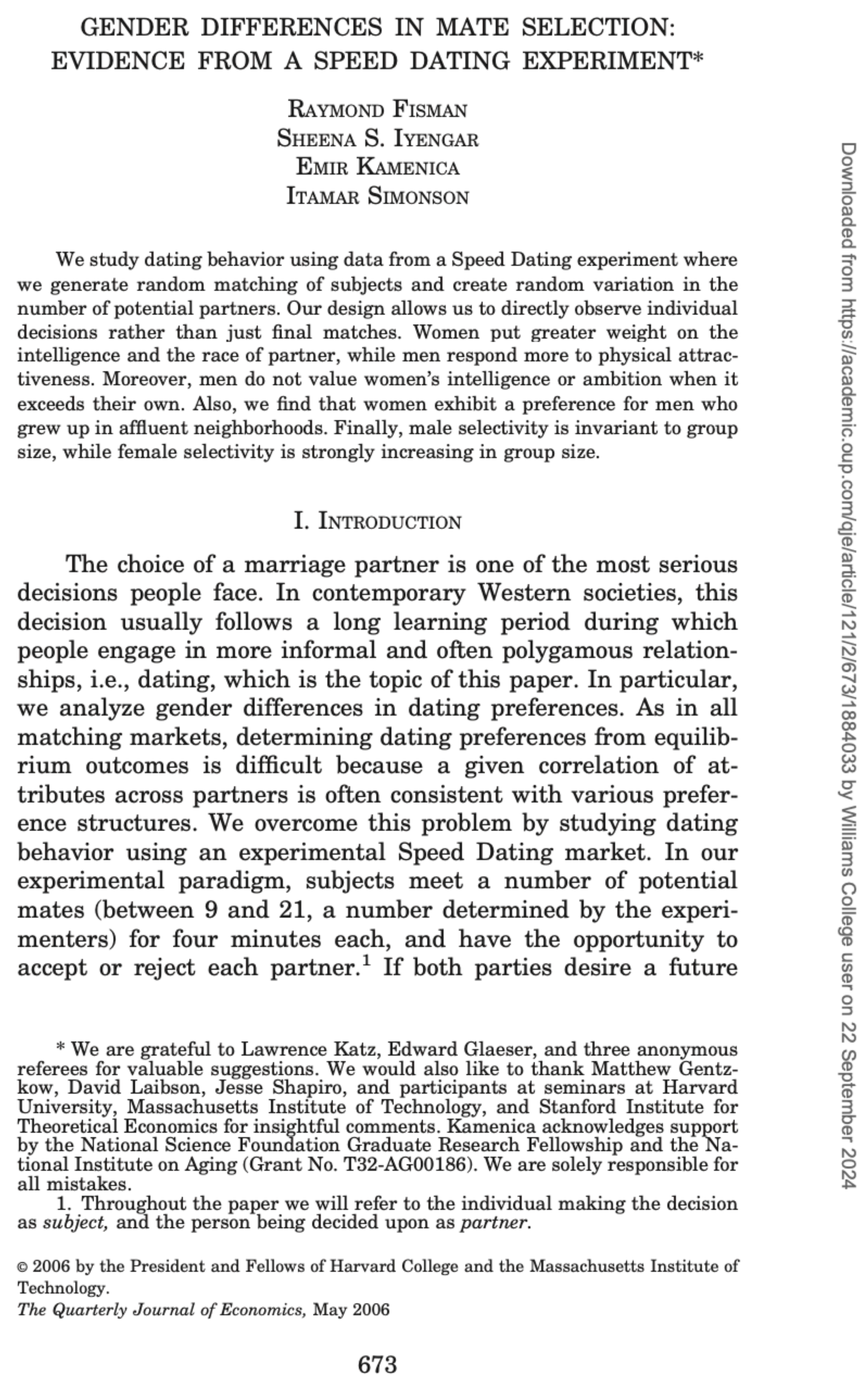
© 2006 by the President and Fellows of Harvard College and the Massachusetts Institute of Technology.
The Quarterly Journal of Economics, May 2006

Downloaded from https://academic.oup.com/qje/article/121/2/673/1884033 by Williams College user on 22 September 2024

Overview

Experimental Design

- Subjects:
  - Students in graduate and professional schools at Columbia University
- Setting:
  - 13 different weekday evenings across 2002 to 2004
  - 1-2 sessions (“waves”) per evening
  - Popular bar/restaurant near campus



GENDER DIFFERENCES IN MATE SELECTION: EVIDENCE FROM A SPEED DATING EXPERIMENT*

RAYMOND FISMAN
SHEENA S. IYENGAR
EMIR KAMENICA
ITAMAR SIMONSON

We study dating behavior using data from a Speed Dating experiment where we generate random matching of subjects and create random variation in the number of potential partners. Our design allows us to directly observe individual decisions rather than just final matches. Women put greater weight on the intelligence and the race of partner, while men respond more to physical attractiveness. Moreover, men do not value women’s intelligence or ambition when it exceeds their own. Also, we find that women exhibit a preference for men who grew up in affluent neighborhoods. Finally, male selectivity is invariant to group size, while female selectivity is strongly increasing in group size.

I. INTRODUCTION

The choice of a marriage partner is one of the most serious decisions people face. In contemporary Western societies, this decision usually follows a long learning period during which people engage in more informal and often polygamous relationships, i.e., dating, which is the topic of this paper. In particular, we analyze gender differences in dating preferences. As in all matching markets, determining dating preferences from equilibrium outcomes is difficult because a given correlation of attributes across partners is often consistent with various preference structures. We overcome this problem by studying dating behavior using an experimental Speed Dating market. In our experimental paradigm, subjects meet a number of potential mates (between 9 and 21, a number determined by the experimenters) for four minutes each, and have the opportunity to accept or reject each partner.[1] If both parties desire a future

* We are grateful to Lawrence Katz, Edward Glaeser, and three anonymous referees for valuable suggestions. We would also like to thank Matthew Gentzkow, David Laibson, Jesse Shapiro, and participants at seminars at Harvard University, Massachusetts Institute of Technology, and Stanford Institute for Theoretical Economics for insightful comments. Kamenica acknowledges support by the National Science Foundation Graduate Research Fellowship and the National Institute on Aging (Grant No. T32-AG00186). We are solely responsible for all mistakes.

1. Throughout the paper we will refer to the individual making the decision as *subject,* and the person being decided upon as *partner.*

© 2006 by the President and Fellows of Harvard College and the Massachusetts Institute of Technology.
The Quarterly Journal of Economics, May 2006

Downloaded from https://academic.oup.com/qje/article/121/2/673/1884033 by Williams College user on 22 September 2024

Overview

Experimental Design

- Other conditions:
  - Pre-event survey
  - Random assignment to sessions and partners
  - Number of partners per wave randomly chosen
    - 5 minimum, 21 maximum

GENDER DIFFERENCES IN MATE SELECTION: EVIDENCE FROM A SPEED DATING EXPERIMENT*

RAYMOND FISMAN
SHEENA S. IYENGAR
EMIR KAMENICA
ITAMAR SIMONSON

We study dating behavior using data from a Speed Dating experiment where we generate random matching of subjects and create random variation in the number of potential partners. Our design allows us to directly observe individual decisions rather than just final matches. Women put greater weight on the intelligence and the race of partner, while men respond more to physical attractiveness. Moreover, men do not value women's intelligence or ambition when it exceeds their own. Also, we find that women exhibit a preference for men who grew up in affluent neighborhoods. Finally, male selectivity is invariant to group size, while female selectivity is strongly increasing in group size.

I. INTRODUCTION

The choice of a marriage partner is one of the most serious decisions people face. In contemporary Western societies, this decision usually follows a long learning period during which people engage in more informal and often polygamous relationships, i.e., dating, which is the topic of this paper. In particular, we analyze gender differences in dating preferences. As in all matching markets, determining dating preferences from equilibrium outcomes is difficult because a given correlation of attributes across partners is often consistent with various preference structures. We overcome this problem by studying dating behavior using an experimental Speed Dating market. In our experimental paradigm, subjects meet a number of potential mates (between 9 and 21, a number determined by the experimenters) for four minutes each, and have the opportunity to accept or reject each partner.[1] If both parties desire a future

* We are grateful to Lawrence Katz, Edward Glaeser, and three anonymous referees for valuable suggestions. We would also like to thank Matthew Gentzkow, David Laibson, Jesse Shapiro, and participants at seminars at Harvard University, Massachusetts Institute of Technology, and Stanford Institute for Theoretical Economics for insightful comments. Kamenica acknowledges support by the National Science Foundation Graduate Research Fellowship and the National Institute on Aging (Grant No. T32-AG00186). We are solely responsible for all mistakes.

1. Throughout the paper we will refer to the individual making the decision as *subject,* and the person being decided upon as *partner*.

© 2006 by the President and Fellows of Harvard College and the Massachusetts Institute of Technology.
The Quarterly Journal of Economics, May 2006

Downloaded from https://academic.oup.com/qje/article/121/2/673/1884033 by Williams College user on 22 September 2024

Event Format

Before

- Fill out pre-event survey:
  - Demographic information
  - Personal interests
  - Dating preferences
  - Perception of others' dating preferences (same- and opposite-sex)
  - Perception of self to others

Event Format

During

1. Check in and receive clipboard, pen, name tag
2. Meet partners:
    1. Seating and 4-minute meeting
    2. Fill out score card
    3. Rotate (repeat)

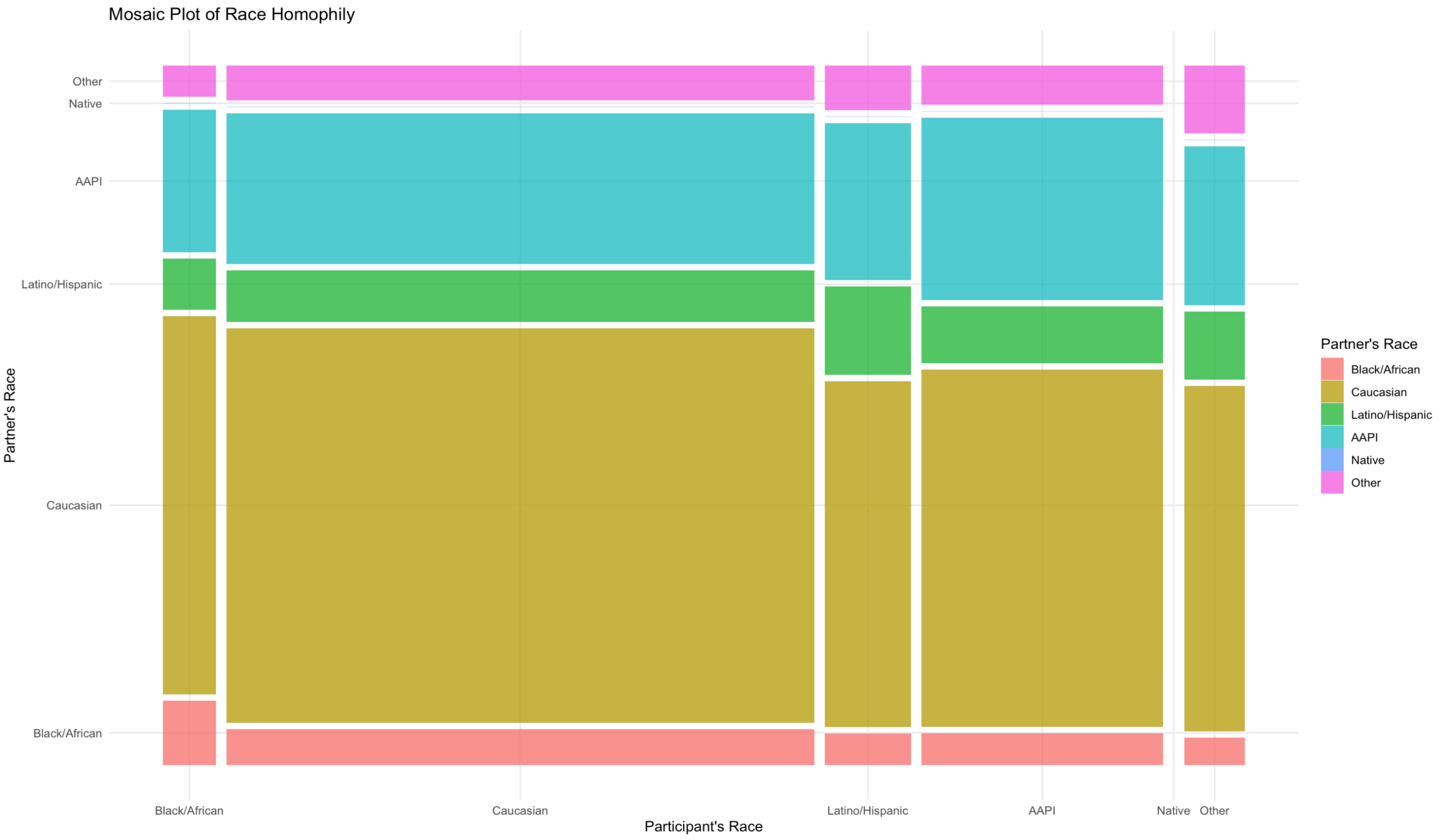
Exploratory Data Analysis

195

Variables Recorded

Grouping

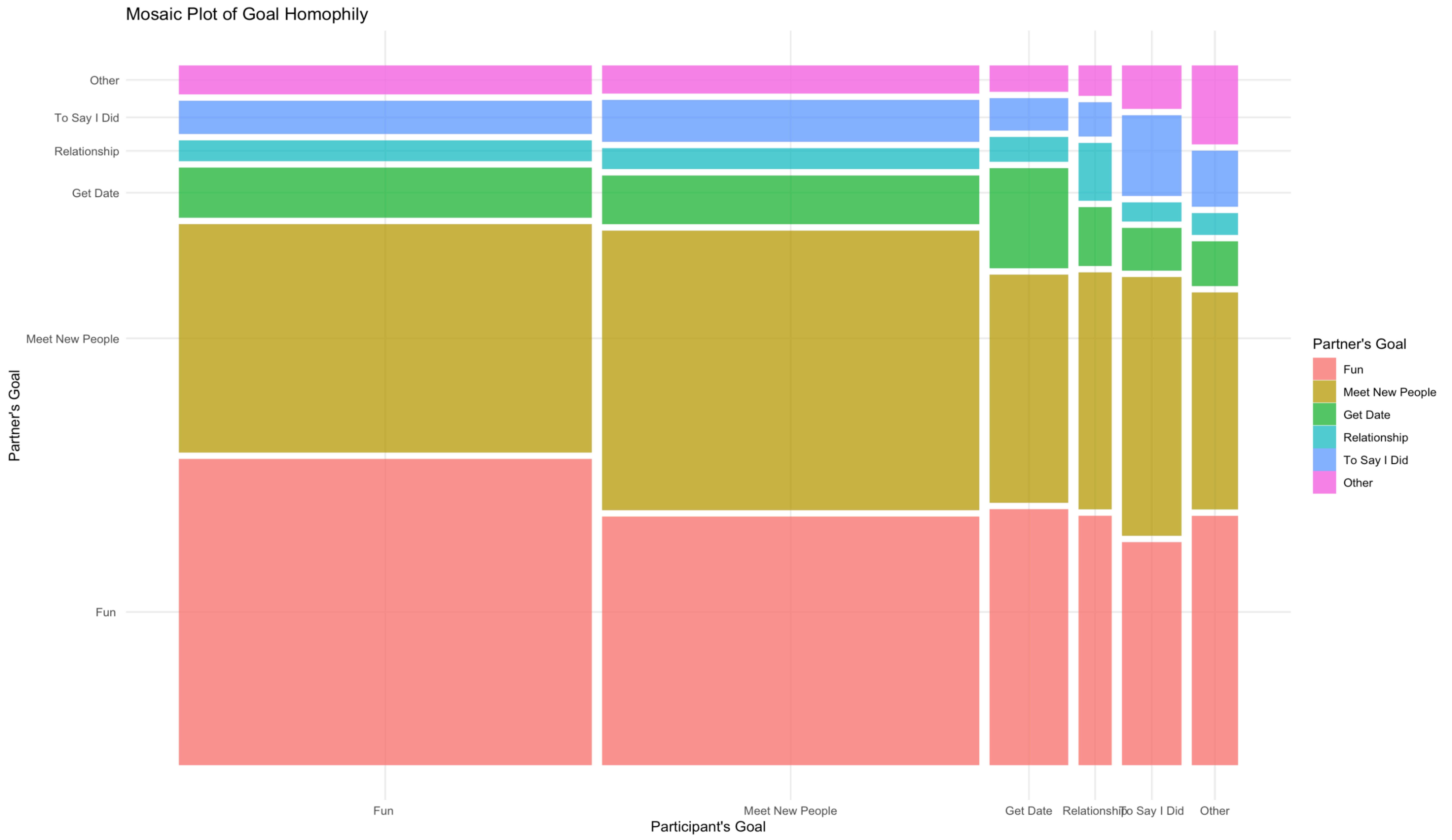
- Demographic information: **27** variables
- Individual-Partner Interaction: **34** variables
- Participant Interests and Activities: **17** variables
- Participant Preferences and Self-Perception: **82** variables
- Perceptions of Others' Preferences: **18** variables
- Mid-Event and Post-Event Preferences: **34** variables
- Expectations and Satisfaction: **5** variables
- Follow-up Outcomes: **5** variables



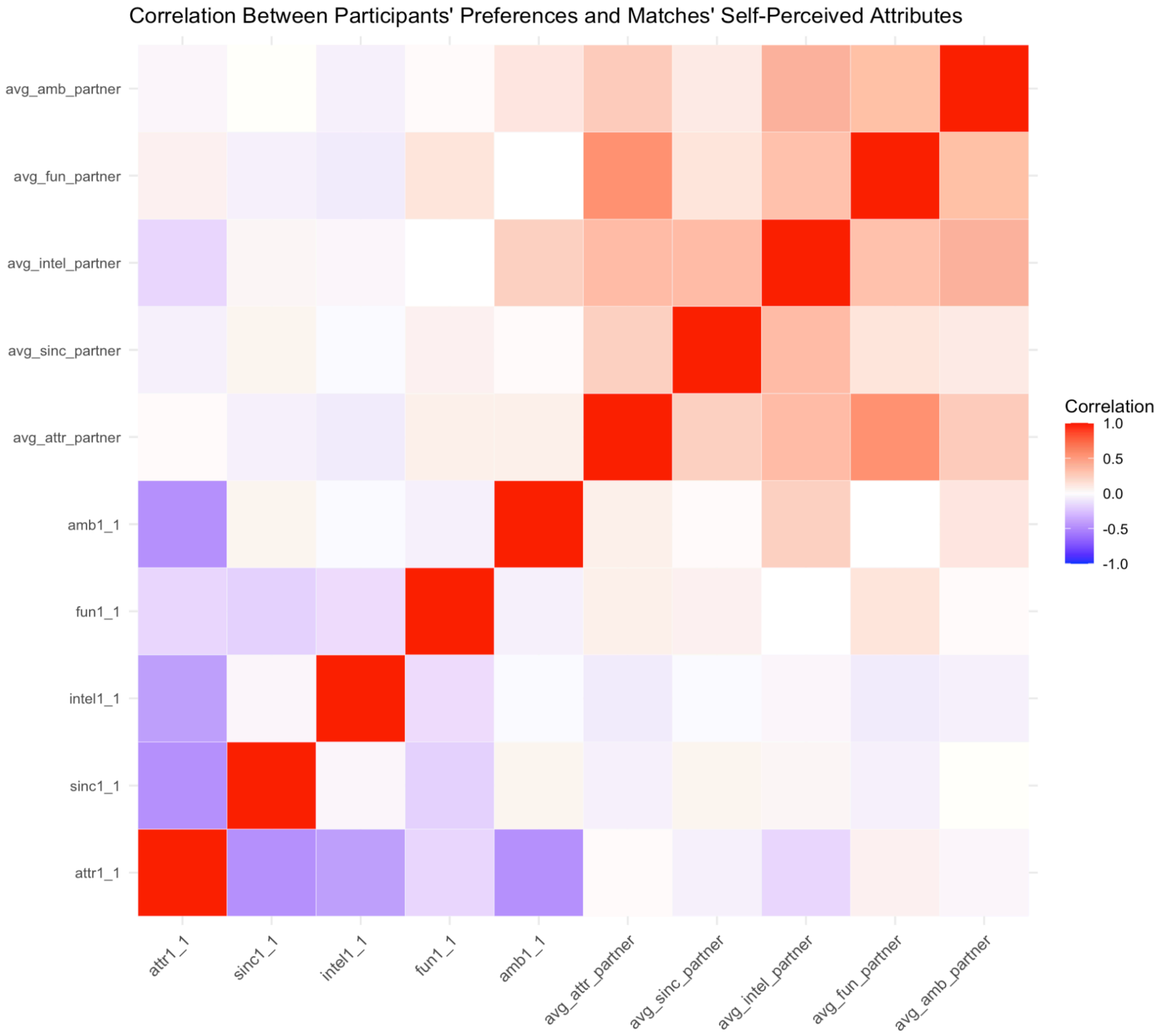
Mosaic Plot of Race Homophily
Other
Native
AAPI
Latino/Hispanic
Caucasian
Black/African
Partner's Race
Black/African
Caucasian
Latino/Hispanic
AAPI
Native
Other
Participant's Race
Partner's Race
Black/African
Caucasian
Latino/Hispanic
AAPI
Native
Other

Race Differences Among Married Couples in 2002 (USA)

	Wife's Race			
Husband's Race	**White (%)**	**Black (%)**	**American Indian & Alaska Native (%)**	**Asian & Other Pacific Islander (%)**
White	85.7	0.2	0.5	0.9
Black	0.5	7.1	0.0	0.1
American Indian & Alaska Native	0.4	0.5	0.4	0.0
Asian & Other Pacific Islander	0.3	0.0	0.0	0.0



Mosaic Plot of Goal Homophily
Partner's Goal
Other
To Say I Did
Relationship
Get Date
Meet New People
Fun
Participant's Goal
Fun
Meet New People
Get Date
Relationship
To Say I Did
Other
Partner's Goal
Fun
Meet New People
Get Date
Relationship
To Say I Did
Other



Correlation Between Participants' Preferences and Matches' Self-Perceived Attributes
avg_amb_partner
avg_fun_partner
avg_intel_partner
avg_sinc_partner
avg_attr_partner
amb1_1
fun1_1
intel1_1
sinc1_1
attr1_1
attr1_1
sinc1_1
intel1_1
fun1_1
amb1_1
avg_attr_partner
avg_sinc_partner
avg_intel_partner
avg_fun_partner
avg_amb_partner
Correlation
1.0
0.5
0.0
-0.5
-1.0

Traditional Methods

Why use network analysis instead?

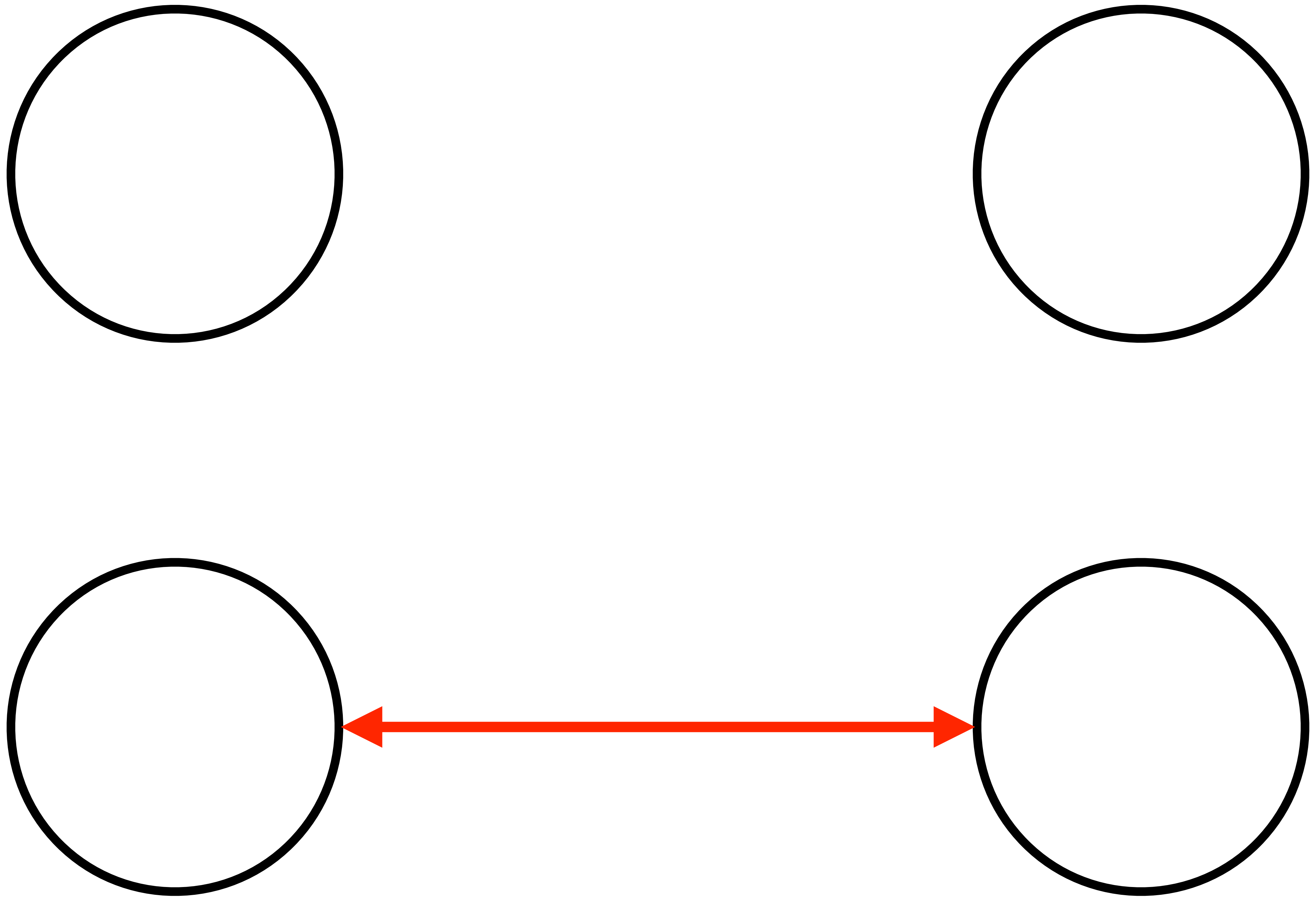
What have we already covered?

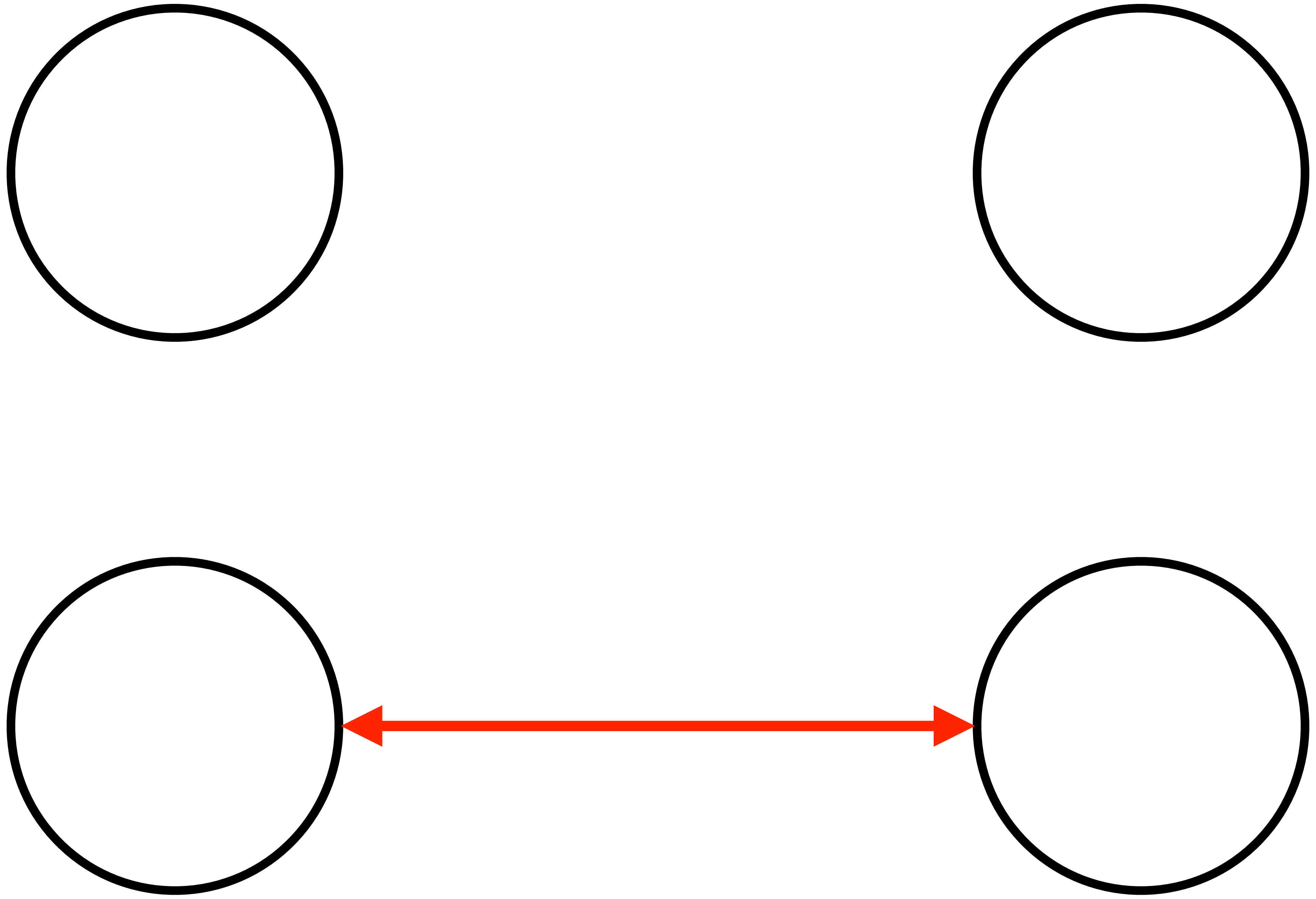
(In Williams statistics classes)

What have we already covered?

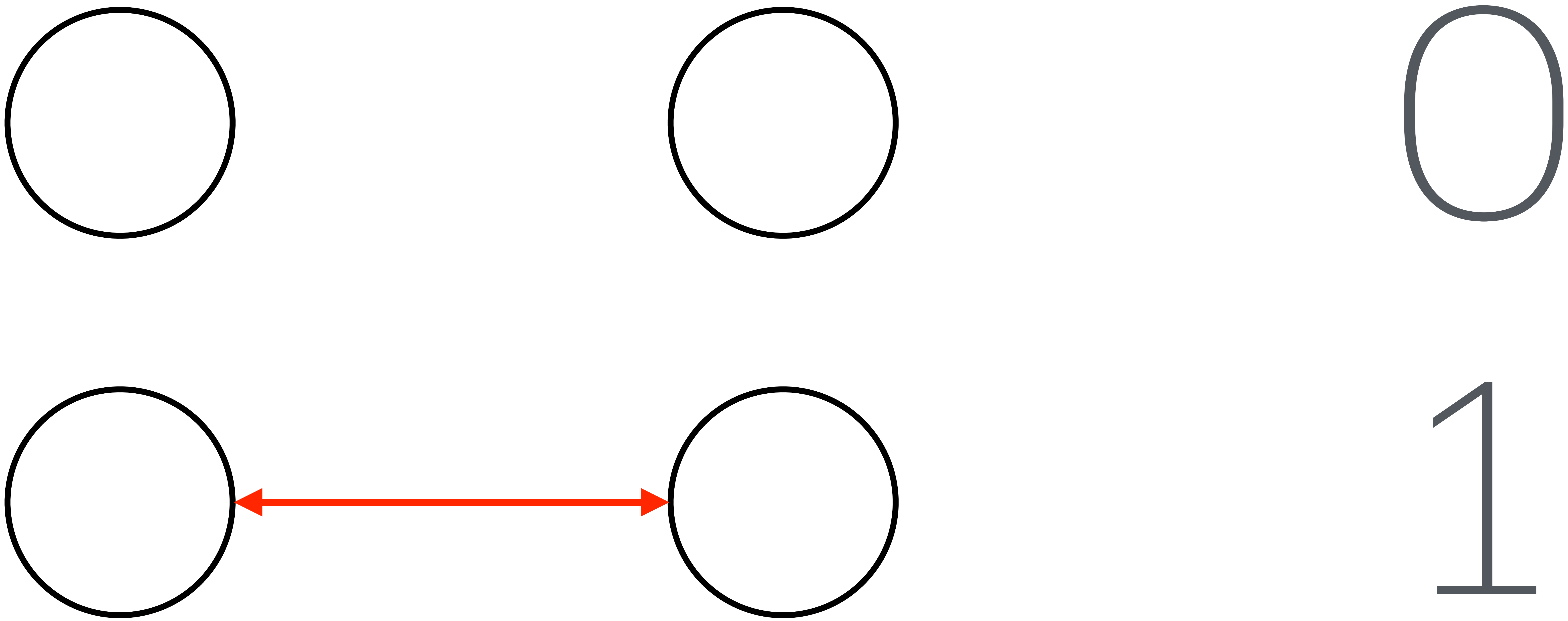
(In Williams statistics classes)

- Logistic Regression
- Decision Trees
- Random Forests, Support Vector Machines





0


0
1

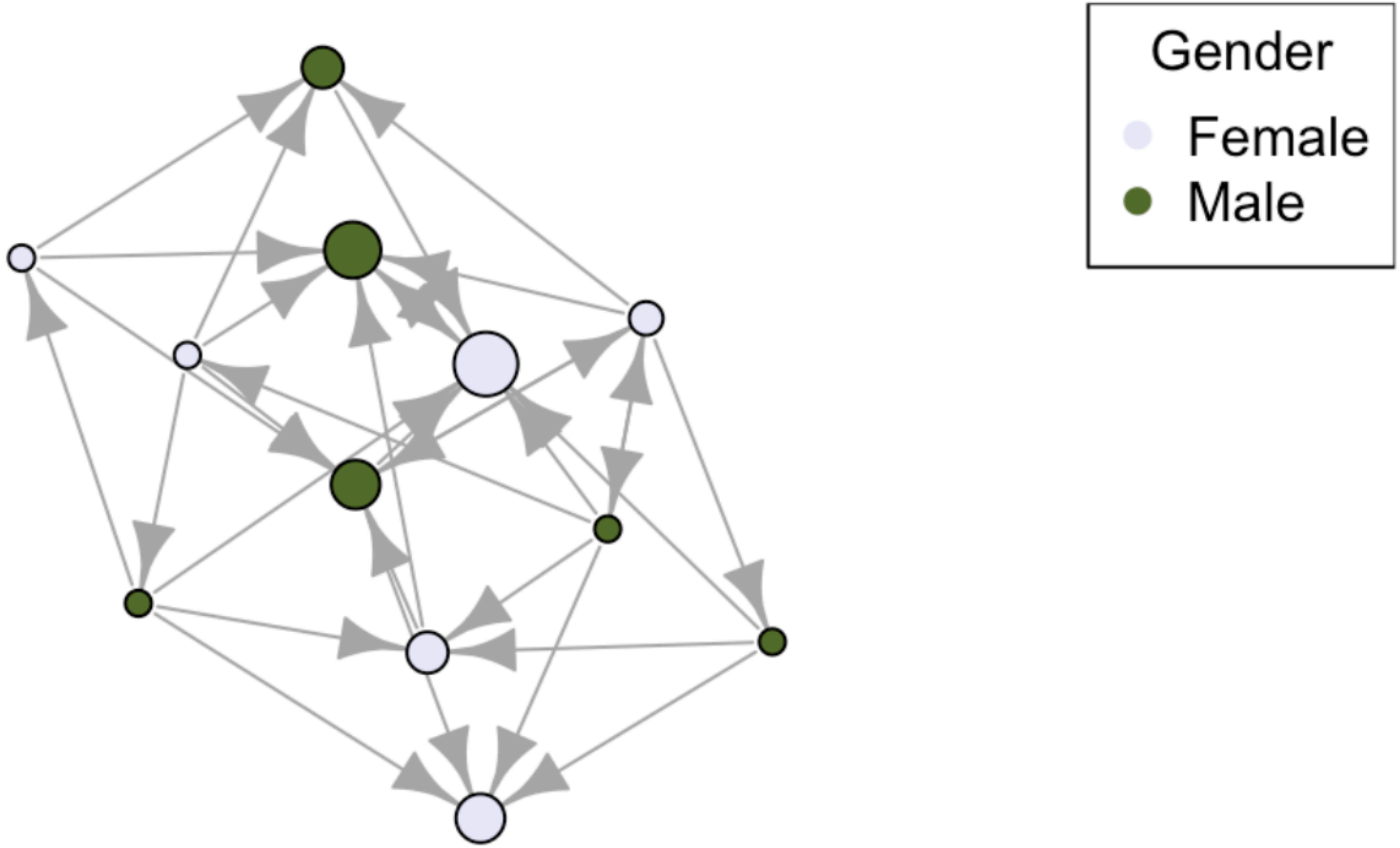
Why not these?

Why not these?

For one, assumption(s).

Second,

Second, relationships.


Gender
Female
Male

Exponential-family Random Graph Models (ERGMs)

A Network Analysis Approach

ERGM Fundamentals

Nodes, Edges, and Tie-Variables

$$N = \{1,\ldots,n\}$$

$$J = \{(i,j) : i,j \in N, i \neq j\}$$

$$X_{ij} = \begin{cases} 1 & \text{if } (i,j) \in E \\ 0 & \text{if } (i,j) \notin E \end{cases}$$

$$X = [X_{ij}], x = [x_{ij}]$$

ERGM Fundamentals

Dyadic Dependence Assumptions

- Bernoulli dependence assumption
  - Ties independent and identically distributed
- Dyad-independent assumption
  - Tie from i to j is dependent on tie from j to i
  - Dyads: pairs of tie-variables

Generalized Linear Model Perspective

ERGM General Form

$$P(Y = y) = \frac{\exp(\theta^\top g(y))h(y)}{k(\theta)}$$

Y : random variable for state of the network

y : particular realization of Y

$g(y)$: vector of model statistics for Y

$h(y)$: reference measure (defines baseline behavior of model when $\theta = 0$)

θ : vector of coefficients for statistics

$k(\theta)$: sum of numerator's value over set of all possible y

Generalized Linear Model Perspective

ERGM Probabilities at the Tie Level

$$\text{logit } P(Y_{ij} = 1 \mid y_{ij}^c) = \theta^\top \delta_{ij}(y)$$

- Y_{ij} : Random variable for the state of the actor pair i, j
- $y_{i,j}^c$: the entire network $\boldsymbol{y}$ except for y_{ij}
- $\delta_{ij}(y)$: vector of "change statistics" for each model term
  - $\delta_{ij}(y) = g(y_{ij}^+) - g(y_{ij}^-) = g(y_{ij} = 1) - g(y_{ij} = 0)$
    - y_{ij}^+ : y_{ij}^c along with $y_{ij} = 1$
    - y_{ij}^- : y_{ij}^c along with $y_{ij} = 0$

Generalized Linear Model Perspective

ERGM $g(y)$ Statistics

- Exogenous term examples:
  - Edges: model intercept
  - Nodal/actor attribute: $\delta_{\text{nodeocov}}(y_{ij}) = x_i$
    - e.g., age or sex
  - Edge covariate: $\delta_{\text{edgecov}}(y_{ij}) = x_{ij}$
    - e.g. how much i likes j from 1 to 10

Generalized Linear Model Perspective

ERGM $g(y)$ Statistics

- More examples:
  - Homophily: $\delta_{\text{nodematch}}(y_{ij}) = \begin{cases} 1 & \text{if } x_i = x_j \\ 0 & \text{if } x_i \neq x_j \end{cases}$
    - e.g. whether i and j are of the same race or field of study
  - Absolute differences: $\delta_{\text{absdiff}}(y_{ij}) = |x_i - x_j|$
    - e.g. absolute difference in how much i and j value music

Generalized Linear Model Perspective

Conditions for Equivalence to Logistic Regression

Generalized Linear Model Perspective

Conditions for Equivalence to Logistic Regression

- Dyadic independence

Generalized Linear Model Perspective

Conditions for Equivalence to Logistic Regression

- Dyadic independence
- Endogenous terms
  - Ex: mutuality/reciprocity
  - Incorporates structural dependencies

Generalized Linear Model Perspective

Conditions for Equivalence to Logistic Regression

- Dyadic independence
- Endogenous terms
  - Ex: mutuality/reciprocity
  - Incorporates structural dependencies
- Exogenous terms
  - Ex: node attributes, dyadic covariates

Generalized Linear Model Perspective

Equivalence Proof

$$P(Y = y) = \frac{\exp(\theta^\top g(y))h(y)}{k(\theta)} \to \frac{\exp(\theta^\top g(y))}{k(\theta)}, \; h \propto 1$$

Generalized Linear Model Perspective

Equivalence Proof

$$g(y) = \sum_{i,j} [y_{ij} g(y_{ij}^{+}) + (1 - y_{ij}) g(y_{ij}^{-})]$$

Generalized Linear Model Perspective

Equivalence Proof

$$P(Y = y) = \frac{\exp(\theta^\top \sum_{i,j} [y_{ij} g(y_{ij}^+) + (1 - y_{ij}) g(y_{ij}^-)])}{k(\theta)}$$

Generalized Linear Model Perspective

Equivalence Proof

$$P(Y=y) = \frac{\exp(\theta^\top \sum_{i,j} [y_{ij} g(y_{ij}^{+}) + (1 - y_{ij}) g(y_{ij}^{-})])}{k(\theta)}$$

$$= \frac{\exp(\theta^\top \sum_{i,j} [g(y_{ij}^{-}) + y_{ij}(g(y_{ij}^{+}) - g(y_{ij}^{-}))])}{k(\theta)}$$

Generalized Linear Model Perspective

Equivalence Proof

$$P(Y = y) = \frac{\exp(\theta^\top \sum_{i,j} [y_{ij} g(y_{ij}^+) + (1 - y_{ij}) g(y_{ij}^-)])}{k(\theta)}$$

$$= \frac{\exp(\theta^\top \sum_{i,j} [g(y_{ij}^-) + y_{ij}(g(y_{ij}^+) - g(y_{ij}^-))])}{k(\theta)}$$

$$= \frac{\exp(\theta^\top \sum_{i,j} [g(y_{ij}^-) + y_{ij}(\delta_{ij}(y))])}{k(\theta)}$$

Generalized Linear Model Perspective

Equivalence Proof

$$P(Y = y) = \frac{\exp(\theta^\top \sum_{i,j} [g(y_{\bar{ij}}) + y_{ij}(\delta_{ij}(y))])}{k(\theta)}$$

$$= \frac{\exp(\sum_{i,j} \theta^\top g(y_{\bar{ij}}) + \sum_{i,j} \theta^\top y_{ij}(\delta_{ij}(y))}{k(\theta)}$$

Generalized Linear Model Perspective

Equivalence Proof

$$P(Y=y) = \frac{\exp(\theta^\top \sum_{i,j}[g(y_{ij}^-) + y_{ij}(\delta_{ij}(y))])}{k(\theta)}$$

$$= \frac{\exp(\sum_{i,j} \theta^\top g(y_{ij}^-) + \sum_{i,j} \theta^\top y_{ij}(\delta_{ij}(y))}{k(\theta)}$$

$$= \frac{\prod_{i,j} \exp(\theta^\top g(y_{ij}^-))\exp(\theta^\top y_{ij}(\delta_{ij}(y)))}{k(\theta)}$$

Generalized Linear Model Perspective

Equivalence Proof

$$
\begin{aligned}
k(\theta) &= \sum_{y \in \{0,1\}^{n \times n}} \exp(\theta^\top g(y)) \\
&= \prod_{i,j} \left[\exp(\theta^\top g(y_{ij}^-)) + \exp(\theta^\top g(y_{ij}^+))\right]
\end{aligned}
$$

Generalized Linear Model Perspective

Equivalence Proof

$$P(Y = y) = \frac{\prod_{i,j} \exp(\theta^\top g(y_{ij}^-))\exp(\theta^\top y_{ij}(\delta_{ij}(y)))}{k(\theta)}$$

$$= \prod_{i,j} \frac{\exp(\theta^\top g(y_{ij}^-))\exp(\theta^\top y_{ij}(\delta_{ij}(y)))}{\exp(\theta^\top g(y_{ij}^-)) + \exp(\theta^\top g(y_{ij}^+))}$$

Generalized Linear Model Perspective

Equivalence Proof

$$\exp\left(\theta^\top y_{ij}\delta_{ij}(y)\right) = \begin{cases} \exp\left(\theta^\top \delta_{ij}(y)\right) & \text{if } y_{ij} = 1 \\ 1 & \text{if } y_{ij} = 0 \end{cases}$$

$$= y_{ij} \cdot \exp\left(\theta^\top \delta_{ij}(y)\right) + (1 - y_{ij}) \cdot 1$$

Generalized Linear Model Perspective

Equivalence Proof

$$P(Y = y) = \prod_{i,j} \left[\frac{\exp(\theta^\top g(y_{ij}^-))[y_{ij} \cdot \exp\left(\theta^\top \delta_{ij}(y)\right) + (1 - y_{ij})]}{\exp(\theta^\top g(y_{ij}^-)) + \exp(\theta^\top g(y_{ij}^+))} \right]$$

Generalized Linear Model Perspective

Equivalence Proof

$$P(Y=y) = \prod_{i,j} \left[\frac{\exp(\theta^\top g(y_{ij}^-))[y_{ij} \cdot \exp\left(\theta^\top \delta_{ij}(y)\right) + (1-y_{ij})]}{\exp(\theta^\top g(y_{ij}^-)) + \exp(\theta^\top g(y_{ij}^+))} \right]$$

$$= \prod_{i,j} \left[\frac{y_{ij} \cdot \exp\left(\theta^\top g(y_{ij}^-)\right) \cdot \exp\left(\theta^\top \delta_{ij}(y)\right) + (1-y_{ij}) \cdot \exp\left(\theta^\top g(y_{ij}^-)\right)}{\exp\left(\theta^\top g(y_{ij}^-)\right) + \exp\left(\theta^\top g(y_{ij}^+)\right)} \right]$$

$$
\begin{aligned}
P(Y=y) &= \prod_{i,j} \left[\frac{\exp(\theta^\top g(y_{ij}^-))[y_{ij} \cdot \exp\left(\theta^\top \delta_{ij}(y)\right) + (1-y_{ij})]}{\exp(\theta^\top g(y_{ij}^-)) + \exp(\theta^\top g(y_{ij}^+))} \right] \\
&= \prod_{i,j} \left[\frac{y_{ij} \cdot \exp\left(\theta^\top g(y_{ij}^-)\right) \cdot \exp\left(\theta^\top \delta_{ij}(y)\right) + (1-y_{ij}) \cdot \exp\left(\theta^\top g(y_{ij}^-)\right)}{\exp\left(\theta^\top g(y_{ij}^-)\right) + \exp\left(\theta^\top g(y_{ij}^+)\right)} \right] \\
&= \prod_{i,j} \left[\frac{y_{ij} \cdot \exp\left(\theta^\top g(y_{ij}^-)\right) \cdot \dfrac{\exp\left(\theta^\top g(y_{ij}^+)\right)}{\exp\left(\theta^\top g(y_{ij}^-)\right)} + (1-y_{ij}) \cdot \exp\left(\theta^\top g(y_{ij}^-)\right)}{\exp\left(\theta^\top g(y_{ij}^-)\right) + \exp\left(\theta^\top g(y_{ij}^+)\right)} \right]
\end{aligned}
$$

$$P(Y=y)=\prod_{i,j}\left[\frac{\exp(\theta^\top g(y_{ij}^-))[y_{ij}\cdot\exp\left(\theta^\top\delta_{ij}(y)\right)+(1-y_{ij})]}{\exp(\theta^\top g(y_{ij}^-))+\exp(\theta^\top g(y_{ij}^+))}\right]$$

$$=\prod_{i,j}\left[\frac{y_{ij}\cdot\exp\left(\theta^\top g(y_{ij}^-)\right)\cdot\exp\left(\theta^\top\delta_{ij}(y)\right)+(1-y_{ij})\cdot\exp\left(\theta^\top g(y_{ij}^-)\right)}{\exp\left(\theta^\top g(y_{ij}^-)\right)+\exp\left(\theta^\top g(y_{ij}^+)\right)}\right]$$

$$=\prod_{i,j}\left[\frac{y_{ij}\cdot\exp\left(\theta^\top g(y_{ij}^-)\right)\cdot\frac{\exp\left(\theta^\top g(y_{ij}^+)\right)}{\exp\left(\theta^\top g(y_{ij}^-)\right)}+(1-y_{ij})\cdot\exp\left(\theta^\top g(y_{ij}^-)\right)}{\exp\left(\theta^\top g(y_{ij}^-)\right)+\exp\left(\theta^\top g(y_{ij}^+)\right)}\right]$$

$$=\prod_{i,j}\left[\frac{y_{ij}\cdot\exp\left(\theta^\top g(y_{ij}^+)\right)+(1-y_{ij})\cdot\exp\left(\theta^\top g(y_{ij}^-)\right)}{\exp\left(\theta^\top g(y_{ij}^-)\right)+\exp\left(\theta^\top g(y_{ij}^+)\right)}\right]$$

$$P(Y=y)=\prod_{i,j}\left[\frac{y_{ij}\cdot\exp\left(\theta^\top g(y_{ij}^+)\right)+(1-y_{ij})\cdot\exp\left(\theta^\top g(y_{ij}^-)\right)}{\exp\left(\theta^\top g(y_{ij}^-)\right)+\exp\left(\theta^\top g(y_{ij}^+)\right)}\right]$$

$$=\prod_{i,j}\left[\frac{y_{ij}\cdot\exp\left(\theta^\top g(y_{ij}^+)\right)}{\exp\left(\theta^\top g(y_{ij}^-)\right)+\exp\left(\theta^\top g(y_{ij}^+)\right)}+\frac{(1-y_{ij})\cdot\exp\left(\theta^\top g(y_{ij}^-)\right)}{\exp\left(\theta^\top g(y_{ij}^-)\right)+\exp\left(\theta^\top g(y_{ij}^+)\right)}\right]$$

$$
\begin{aligned}
P(Y=y) &= \prod_{i,j}\left[\frac{y_{ij}\cdot\exp\left(\theta^\top g(y_{ij}^+)\right)+(1-y_{ij})\cdot\exp\left(\theta^\top g(y_{ij}^-)\right)}{\exp\left(\theta^\top g(y_{ij}^-)\right)+\exp\left(\theta^\top g(y_{ij}^+)\right)}\right] \\
&= \prod_{i,j}\left[\frac{y_{ij}\cdot\exp\left(\theta^\top g(y_{ij}^+)\right)}{\exp\left(\theta^\top g(y_{ij}^-)\right)+\exp\left(\theta^\top g(y_{ij}^+)\right)}+\frac{(1-y_{ij})\cdot\exp\left(\theta^\top g(y_{ij}^-)\right)}{\exp\left(\theta^\top g(y_{ij}^-)\right)+\exp\left(\theta^\top g(y_{ij}^+)\right)}\right] \\
&= \prod_{i,j}\left[y_{ij}\cdot\frac{\exp\left(\theta^\top g(y_{ij}^+)\right)}{\exp\left(\theta^\top g(y_{ij}^-)\right)+\exp\left(\theta^\top g(y_{ij}^+)\right)}+(1-y_{ij})\cdot\frac{\exp\left(\theta^\top g(y_{ij}^-)\right)}{\exp\left(\theta^\top g(y_{ij}^-)\right)+\exp\left(\theta^\top g(y_{ij}^+)\right)}\right]
\end{aligned}
$$

$$P(Y_{ij} = 1) = \frac{\exp\left(\theta^\top g(y_{ij}^{+})\right)}{\exp\left(\theta^\top g(y_{ij}^{-})\right) + \exp\left(\theta^\top g(y_{ij}^{+})\right)}$$

$$P(Y_{ij} = 0) = \frac{\exp\left(\theta^\top g(y_{ij}^{-})\right)}{\exp\left(\theta^\top g(y_{ij}^{-})\right) + \exp\left(\theta^\top g(y_{ij}^{+})\right)}$$

$$P(Y=y)=\prod_{i,j}\left[y_{ij}\cdot\frac{\exp\left(\theta^\top g(y_{ij}^{+})\right)}{\exp\left(\theta^\top g(y_{ij}^{-})\right)+\exp\left(\theta^\top g(y_{ij}^{+})\right)}+(1-y_{ij})\cdot\frac{\exp\left(\theta^\top g(y_{ij}^{-})\right)}{\exp\left(\theta^\top g(y_{ij}^{-})\right)+\exp\left(\theta^\top g(y_{ij}^{+})\right)}\right]$$

$$=\prod_{i,j}\left[y_{ij}\cdot P(Y_{ij}=1)+(1-y_{ij})\cdot P(Y_{ij}=0)\right]$$

Results

Modeling

Setup: Wave 1

- First: Logistic regression
  - 43 predictors
  - Both stepwise selection
- Second: ERGM

```
                 Estimate Std. Error z value Pr(>|z|)    
(Intercept)      -22.15263    6.21593  -3.564 0.000365 ***
imprace           -0.65334    0.21165  -3.087 0.002023 ** 
expnum             1.23394    0.29374   4.201 2.66e-05 ***
attr1_1           -0.18268    0.04677  -3.906 9.38e-05 ***
shar1_1            0.71385    0.17990   3.968 7.25e-05 ***
amb3_1            -1.43806    0.39737  -3.619 0.000296 ***
like               1.04042    0.57540   1.808 0.070582 .  
prob               1.03813    0.36124   2.874 0.004056 ** 
attr               0.82238    0.43071   1.909 0.056214 .  
intel             -1.04105    0.48768  -2.135 0.032785 *  
fun                0.63616    0.34089   1.866 0.062019 .  
shar               1.04935    0.40502   2.591 0.009574 ** 
same_fieldTRUE    -3.46116    1.11771  -3.097 0.001957 ** 
absdiff_hiking     0.54711    0.30791   1.777 0.075591 .  
absdiff_tv         0.45452    0.22352   2.033 0.042003 *  
absdiff_theater    0.78887    0.29810   2.646 0.008137 ** 
absdiff_concerts  -0.65720    0.29829  -2.203 0.027579 *  
absdiff_shopping   0.52508    0.22460   2.338 0.019396 *  
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 274.83  on 199  degrees of freedom
Residual deviance:  54.99  on 182  degrees of freedom
AIC: 90.99
```

Logistic Regression

```
                                  Estimate Std. Error MCMC % z value Pr(>|z|)    
edges                            -22.15263    6.21593      0  -3.564 0.000365 ***
nodeocov.imprace                  -0.65334    0.21165      0  -3.087 0.002023 ** 
nodeocov.expnum                    1.23394    0.29374      0   4.201  < 1e-04 ***
nodeocov.attr1_1                  -0.18268    0.04677      0  -3.906  < 1e-04 ***
nodeocov.shar1_1                   0.71385    0.17990      0   3.968  < 1e-04 ***
nodeocov.amb3_1                   -1.43806    0.39737      0  -3.619 0.000296 ***
edgecov.edge_matrices[["like"]]    1.04042    0.57540      0   1.808 0.070582 .  
edgecov.edge_matrices[["prob"]]    1.03813    0.36124      0   2.874 0.004056 ** 
edgecov.edge_matrices[["attr"]]    0.82238    0.43071      0   1.909 0.056214 .  
edgecov.edge_matrices[["intel"]]  -1.04105    0.48768      0  -2.135 0.032785 *  
edgecov.edge_matrices[["fun"]]     0.63616    0.34089      0   1.866 0.062019 .  
edgecov.edge_matrices[["shar"]]    1.04935    0.40502      0   2.591 0.009574 ** 
nodematch.gender                      -Inf    0.00000      0    -Inf  < 1e-04 ***
nodematch.field_cd                -3.46116    1.11771      0  -3.097 0.001957 ** 
absdiff.hiking                     0.54711    0.30791      0   1.777 0.075591 .  
absdiff.tv                         0.45452    0.22352      0   2.033 0.042003 *  
absdiff.theater                    0.78887    0.29810      0   2.646 0.008137 ** 
absdiff.concerts                  -0.65720    0.29829      0  -2.203 0.027579 *  
absdiff.shopping                   0.52508    0.22460      0   2.338 0.019396 *  
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

     Null Deviance: 277.26  on 200  degrees of freedom
 Residual Deviance:  54.99  on 181  degrees of freedom
 
AIC: 90.99  BIC: 150.4  (Smaller is better. MC Std. Err. = 0)
```

ERGM

Modeling

Endogenous Term

```
                                   Estimate Std. Error MCMC % z value Pr(>|z|)    
edges                              -22.46068    6.35034      0  -3.537 0.000405 ***
mutual                              -0.87428    0.89060      0  -0.982 0.326259    
nodeocov.imprace                    -0.70108    0.22532      0  -3.111 0.001862 ** 
nodeocov.expnum                      1.25310    0.29958      0   4.183  < 1e-04 ***
nodeocov.attr1_1                    -0.18428    0.04738      0  -3.889 0.000101 ***
nodeocov.shar1_1                     0.72520    0.18642      0   3.890 0.000100 ***
nodeocov.amb3_1                     -1.46643    0.43192      0  -3.395 0.000686 ***
edgecov.edge_matrices[["like"]]      0.94998    0.56563      0   1.680 0.093053 .  
edgecov.edge_matrices[["prob"]]      1.05583    0.37114      0   2.845 0.004443 ** 
edgecov.edge_matrices[["attr"]]      0.85371    0.43544      0   1.961 0.049931 *  
edgecov.edge_matrices[["intel"]]    -1.01567    0.50902      0  -1.995 0.046008 *  
edgecov.edge_matrices[["fun"]]       0.75939    0.36812      0   2.063 0.039121 *  
edgecov.edge_matrices[["shar"]]      1.05114    0.40013      0   2.627 0.008614 ** 
nodematch.gender                        -Inf    0.00000      0    -Inf  < 1e-04 ***
nodematch.field_cd                  -3.50931    1.16451      0  -3.014 0.002582 ** 
absdiff.hiking                       0.52793    0.30507      0   1.731 0.083535 .  
absdiff.tv                           0.46147    0.22268      0   2.072 0.038231 *  
absdiff.theater                      0.80261    0.29681      0   2.704 0.006848 ** 
absdiff.concerts                    -0.59531    0.31713      0  -1.877 0.060493 .  
absdiff.shopping                     0.52031    0.23836      0   2.183 0.029049 *  
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

Modeling

Endogenous Term

	Estimate	Std. Error	MCMC %	z value	Pr(>|z|)
mutual	-0.87428	0.89060	0	-0.982	0.326259

Modeling

Dependence Question

```
        Pearson's Chi-squared test with Yates' continuity correction

data:  cont_table
X-squared = 0, df = 1, p-value = 1

data:  cont_table2
X-squared = 14.878, df = 1, p-value = 0.0001147

data:  cont_table3
X-squared = 0.94029, df = 1, p-value = 0.3322
data:  cont_table4
X-squared = 2.6938, df = 1, p-value = 0.1007

data:  cont_table5
X-squared = 4.7416, df = 1, p-value = 0.02944

data:  cont_table6
X-squared = 1.28, df = 1, p-value = 0.2579

data:  cont_table7
X-squared = 2.0978, df = 1, p-value = 0.1475
data:  cont_table8
X-squared = 0.019176, df = 1, p-value = 0.8899

data:  cont_table9
X-squared = 11.413, df = 1, p-value = 0.0007293

data:  cont_table10
X-squared = 6.3678, df = 1, p-value = 0.01162
```

```
data:  cont_table11
X-squared = 4.5159, df = 1, p-value = 0.03358
data:  cont_table12
X-squared = 0.3498, df = 1, p-value = 0.5542

data:  cont_table13
X-squared = 6.5832, df = 1, p-value = 0.01029

data:  cont_table14
X-squared = 0.080355, df = 1, p-value = 0.7768
data:  cont_table15
X-squared = 4.0851, df = 1, p-value = 0.04326
data:  cont_table16
X-squared = 1.8778, df = 1, p-value = 0.1706
data:  cont_table17
X-squared = 0.85158, df = 1, p-value = 0.3561
data:  cont_table18
X-squared = 13.585, df = 1, p-value = 0.000228

data:  cont_table19
X-squared = 16.845, df = 1, p-value = 4.055e-05
data:  cont_table20
X-squared = 3.36, df = 1, p-value = 0.0668
data:  cont_table21
X-squared = 2.3431, df = 1, p-value = 0.1258
```

Modeling

Under Verified Dependence: Wave 21

```
                                          Estimate Std. Error MCMC % z value Pr(>|z|)    
edges                                     -5.42208    1.30373      0  -4.159  < 1e-04 ***
nodeofactor.race.Black/African American   -2.42903    0.96887      0  -2.507 0.012173 *  
nodeofactor.race.European/Caucasian-American -1.29112    0.30064      0  -4.295  < 1e-04 ***
nodeofactor.race.Latino/Hispanic American -1.62022    0.51943      0  -3.119 0.001813 ** 
nodeofactor.race.Other                     0.23341    0.69667      0   0.335 0.737595    
nodeofactor.goal.To get a date            -1.28539    0.90901      0  -1.414 0.157347    
nodeofactor.goal.To meet new people        0.73637    0.26938      0   2.734 0.006264 ** 
nodeofactor.goal.To say I did it           1.57114    0.59783      0   2.628 0.008587 ** 
nodeofactor.go_out.Several times a week    0.53722    0.35384      0   1.518 0.128948    
nodeofactor.go_out.Several times a year   -0.38486    1.37018      0  -0.281 0.778800    
nodeofactor.go_out.Twice a month           1.25233    0.52303      0   2.394 0.016649 *  
nodeofactor.go_out.Twice a week           -1.17094    0.37571      0  -3.117 0.001830 ** 
nodeocov.imprace                           0.21571    0.05651      0   3.817 0.000135 ***
nodeocov.sinc1_1                           0.14171    0.02507      0   5.653  < 1e-04 ***
nodeocov.intel1_1                          0.05324    0.02487      0   2.141 0.032290 *  
nodeocov.amb1_1                           -0.18398    0.03216      0  -5.720  < 1e-04 ***
nodeocov.shar1_1                           0.03289    0.02255      0   1.458 0.144780    
nodeocov.sinc3_1                          -1.15556    0.14526      0  -7.955  < 1e-04 ***
nodeocov.intel3_1                          0.24973    0.15283      0   1.634 0.102263    
nodeocov.amb3_1                            0.19262    0.08971      0   2.147 0.031789 *  
edgecov.edge_matrices[["like"]]            0.59411    0.10494      0   5.661  < 1e-04 ***
edgecov.edge_matrices[["prob"]]            0.32799    0.07114      0   4.610  < 1e-04 ***
edgecov.edge_matrices[["attr"]]            0.52142    0.07237      0   7.205  < 1e-04 ***
edgecov.edge_matrices[["sinc"]]           -0.35947    0.08024      0  -4.480  < 1e-04 ***
edgecov.edge_matrices[["fun"]]             0.25605    0.08741      0   2.929 0.003399 ** 
edgecov.edge_matrices[["shar"]]            0.36418    0.07123      0   5.113  < 1e-04 ***
nodematch.gender                              -Inf    0.00000      0    -Inf  < 1e-04 ***
absdiff.clubbing                          -0.09172    0.04463      0  -2.055 0.039881 *  
absdiff.concerts                          -0.11840    0.06662      0  -1.777 0.075550 .  
absdiff.shopping                          -0.10671    0.04959      0  -2.152 0.031408 *  
absdiff.yoga                              -0.11128    0.04949      0  -2.249 0.024526 *  
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

Modeling

Under Verified Dependence: Wave 21

- Same analysis as before
- Differences:
    - Changes in terms, coefficient signs, coefficient significances between Wave 1 and Wave 21
- Key similarities:
    - ↑ shar1_1, attr, fun, shar, like, prob

Modeling

Reciprocation

```
                                          Estimate Std. Error MCMC % z value Pr(>|z|)    
edges                                     -5.36473    1.29879      0  -4.131  < 1e-04 ***
mutual                                    -0.13990    0.24718      0  -0.566 0.571407    
nodeofactor.race.Black/African American   -2.38224    0.96429      0  -2.470 0.013494 *  
nodeofactor.race.European/Caucasian-American -1.28183    0.29527      0  -4.341  < 1e-04 ***
nodeofactor.race.Latino/Hispanic American -1.62590    0.52069      0  -3.123 0.001793 ** 
nodeofactor.race.Other                     0.22973    0.69623      0   0.330 0.741433    
nodeofactor.goal.To get a date            -1.24418    0.90997      0  -1.367 0.171537    
nodeofactor.goal.To meet new people        0.74530    0.27281      0   2.732 0.006297 ** 
nodeofactor.goal.To say I did it           1.59118    0.61501      0   2.587 0.009675 ** 
nodeofactor.go_out.Several times a week    0.54146    0.35150      0   1.540 0.123456    
nodeofactor.go_out.Several times a year   -0.41935    1.36481      0  -0.307 0.758644    
nodeofactor.go_out.Twice a month           1.26880    0.53689      0   2.363 0.018116 *  
nodeofactor.go_out.Twice a week           -1.14037    0.38488      0  -2.963 0.003048 ** 
nodeocov.imprace                           0.21771    0.05728      0   3.801 0.000144 ***
nodeocov.sinc1_1                           0.14129    0.02517      0   5.614  < 1e-04 ***
nodeocov.intel1_1                          0.05312    0.02517      0   2.110 0.034829 *  
nodeocov.amb1_1                           -0.18547    0.03273      0  -5.666  < 1e-04 ***
nodeocov.shar1_1                           0.03428    0.02243      0   1.529 0.126332    
nodeocov.sinc3_1                          -1.16378    0.14755      0  -7.888  < 1e-04 ***
nodeocov.intel3_1                          0.25319    0.15098      0   1.677 0.093538 .  
nodeocov.amb3_1                            0.18902    0.08973      0   2.107 0.035145 *  
edgecov.edge_matrices[["like"]]            0.58861    0.10484      0   5.614  < 1e-04 ***
edgecov.edge_matrices[["prob"]]            0.33193    0.07244      0   4.582  < 1e-04 ***
edgecov.edge_matrices[["attr"]]            0.51933    0.07256      0   7.157  < 1e-04 ***
edgecov.edge_matrices[["sinc"]]           -0.35765    0.08035      0  -4.451  < 1e-04 ***
edgecov.edge_matrices[["fun"]]             0.26349    0.09090      0   2.899 0.003748 ** 
edgecov.edge_matrices[["shar"]]            0.36618    0.07207      0   5.081  < 1e-04 ***
nodematch.gender                              -Inf    0.00000      0    -Inf  < 1e-04 ***
absdiff.clubbing                          -0.09572    0.04584      0  -2.088 0.036780 *  
absdiff.concerts                          -0.11968    0.06671      0  -1.794 0.072821 .  
absdiff.shopping                          -0.10697    0.05047      0  -2.119 0.034052 *  
absdiff.yoga                              -0.10936    0.05095      0  -2.146 0.031856 *  
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

Modeling

Reciprocation

```
                                         Estimate Std. Error MCMC % z value Pr(>|z|)
edges                                    -5.36473    1.29879      0  -4.131  < 1e-04 ***
mutual                                   -0.13990    0.24718      0  -0.566 0.571407
nodeofactor.race.Black/African American  -2.38224    0.96429      0  -2.470 0.013494 *
nodeofactor.race.European/Caucasian-American -1.28183    0.29527      0  -4.341  < 1e-04 ***
nodeofactor.race.Latino/Hispanic American -1.62590    0.52069      0  -3.123 0.001793 **
nodeofactor.race.Other                    0.22973    0.69623      0   0.330 0.741433
nodeofactor.goal.To get a date           -1.24418    0.90997      0  -1.367 0.171537
nodeofactor.goal.To meet new people       0.74530    0.27281      0   2.732 0.006297 **
nodeofactor.goal.To say I did it          1.59118    0.61501      0   2.587 0.009675 **
nodeofactor.go_out.Several times a week   0.54146    0.35150      0   1.540 0.123456
nodeofactor.go_out.Several times a year  -0.41935    1.36481      0  -0.307 0.758644
nodeofactor.go_out.Twice a month          1.26880    0.53689      0   2.363 0.018116 *
nodeofactor.go_out.Twice a week          -1.14037    0.38488      0  -2.963 0.003048 **
nodeocov.imprace                          0.21771    0.05728      0   3.801 0.000144 ***
nodeocov.sinc1_1                          0.14129    0.02517      0   5.614  < 1e-04 ***
nodeocov.intel1_1                         0.05312    0.02517      0   2.110 0.034829 *
nodeocov.amb1_1                          -0.18547    0.03273      0  -5.666  < 1e-04 ***
nodeocov.shar1_1                          0.03428    0.02243      0   1.529 0.126332
nodeocov.sinc3_1                         -1.16378    0.14755      0  -7.888  < 1e-04 ***
nodeocov.intel3_1                         0.25319    0.15098      0   1.677 0.093538 .
nodeocov.amb3_1                           0.18902    0.08973      0   2.107 0.035145 *
edgecov.edge_matrices[["like"]]           0.58861    0.10484      0   5.614  < 1e-04 ***
edgecov.edge_matrices[["prob"]]           0.33193    0.07244      0   4.582  < 1e-04 ***
edgecov.edge_matrices[["attr"]]           0.51933    0.07256      0   7.157  < 1e-04 ***
edgecov.edge_matrices[["sinc"]]          -0.35765    0.08035      0  -4.451  < 1e-04 ***
edgecov.edge_matrices[["fun"]]            0.26349    0.09090      0   2.899 0.003748 **
edgecov.edge_matrices[["shar"]]           0.36618    0.07207      0   5.081  < 1e-04 ***
nodematch.gender                             -Inf    0.00000      0    -Inf  < 1e-04 ***
absdiff.clubbing                         -0.09572    0.04584      0  -2.088 0.036780 *
absdiff.concerts                         -0.11968    0.06671      0  -1.794 0.072821 .
absdiff.shopping                         -0.10697    0.05047      0  -2.119 0.034052 *
absdiff.yoga                             -0.10936    0.05095      0  -2.146 0.031856 *
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```


```
edgecov.edge_matrices[["prob"]]  0.33193    0.07244      0   4.582  < 1e-04 ***
```

Modeling

Reciprocation

```
                                               Estimate Std. Error MCMC % z value Pr(>|z|)
edges                                          -5.36473    1.29879      0  -4.131  < 1e-04 ***
mutual                                         -0.13990    0.24718      0  -0.566 0.571407
nodeofactor.race.Black/African American        -2.38224    0.96429      0  -2.470 0.013494 *
nodeofactor.race.European/Caucasian-American -1.28183    0.29527      0  -4.341  < 1e-04 ***
nodeofactor.race.Latino/Hispanic American      -1.62590    0.52069      0  -3.123 0.001793 **
nodeofactor.race.Other                          0.22973    0.69623      0   0.330 0.741433
nodeofactor.goal.To get a date                 -1.24418    0.90997      0  -1.367 0.171537
nodeofactor.goal.To meet new people             0.74530    0.27281      0   2.732 0.006297 **
nodeofactor.goal.To say I did it                1.59118    0.61501      0   2.587 0.009675 **
nodeofactor.go_out.Several times a week         0.54146    0.35150      0   1.540 0.123456
nodeofactor.go_out.Several times a year        -0.41935    1.36481      0  -0.307 0.758644
nodeofactor.go_out.Twice a month                1.26880    0.53689      0   2.363 0.018116 *
nodeofactor.go_out.Twice a week                -1.14037    0.38488      0  -2.963 0.003048 **
nodeocov.imprace                                0.21771    0.05728      0   3.801 0.000144 ***
nodeocov.sinc1_1                                0.14129    0.02517      0   5.614  < 1e-04 ***
nodeocov.intel1_1                               0.05312    0.02517      0   2.110 0.034829 *
nodeocov.amb1_1                                -0.18547    0.03273      0  -5.666  < 1e-04 ***
nodeocov.shar1_1                                0.03428    0.02243      0   1.529 0.126332
nodeocov.sinc3_1                               -1.16378    0.14755      0  -7.888  < 1e-04 ***
nodeocov.intel3_1                               0.25319    0.15098      0   1.677 0.093538 .
nodeocov.amb3_1                                 0.18902    0.08973      0   2.107 0.035145 *
edgecov.edge_matrices[["like"]]                 0.58861    0.10484      0   5.614  < 1e-04 ***
edgecov.edge_matrices[["prob"]]                 0.33193    0.07244      0   4.582  < 1e-04 ***
edgecov.edge_matrices[["attr"]]                 0.51933    0.07256      0   7.157  < 1e-04 ***
edgecov.edge_matrices[["sinc"]]                -0.35765    0.08035      0  -4.451  < 1e-04 ***
edgecov.edge_matrices[["fun"]]                  0.26349    0.09090      0   2.899 0.003748 **
edgecov.edge_matrices[["shar"]]                 0.36618    0.07207      0   5.081  < 1e-04 ***
nodematch.gender                                   -Inf    0.00000      0    -Inf  < 1e-04 ***
absdiff.clubbing                               -0.09572    0.04584      0  -2.088 0.036780 *
absdiff.concerts                               -0.11968    0.06671      0  -1.794 0.072821 .
absdiff.shopping                               -0.10697    0.05047      0  -2.119 0.034052 *
absdiff.yoga                                   -0.10936    0.05095      0  -2.146 0.031856 *
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

mutual -0.13990 0.24718 0 -0.566 0.571407

edgecov.edge_matrices[["prob"]] 0.33193 0.07244 0 4.582 < 1e-04 ***

Modeling

Reciprocation

```
                                          Estimate Std. Error MCMC % z value Pr(>|z|)
edges                                     -5.36473    1.29879      0  -4.131  < 1e-04 ***
mutual                                    -0.13990    0.24718      0  -0.566 0.571407
nodeofactor.race.Black/African American   -2.38224    0.96429      0  -2.470 0.013494 *
nodeofactor.race.European/Caucasian-American -1.28183    0.29527      0  -4.341  < 1e-04 ***
nodeofactor.race.Latino/Hispanic American -1.62590    0.52069      0  -3.123 0.001793 **
nodeofactor.race.Other                     0.22973    0.69623      0   0.330 0.741433
nodeofactor.goal.To get a date            -1.24418    0.90997      0  -1.367 0.171537
nodeofactor.goal.To meet new people        0.74530    0.27281      0   2.732 0.006297 **
nodeofactor.goal.To say I did it           1.59118    0.61501      0   2.587 0.009675 **
nodeofactor.go_out.Several times a week    0.54146    0.35150      0   1.540 0.123456
nodeofactor.go_out.Several times a year   -0.41935    1.36481      0  -0.307 0.758644
nodeofactor.go_out.Twice a month           1.26880    0.53689      0   2.363 0.018116 *
nodeofactor.go_out.Twice a week           -1.14037    0.38488      0  -2.963 0.003048 **
nodeocov.imprace                           0.21771    0.05728      0   3.801 0.000144 ***
nodeocov.sinc1_1                           0.14129    0.02517      0   5.614  < 1e-04 ***
nodeocov.intel1_1                          0.05312    0.02517      0   2.110 0.034829 *
nodeocov.amb1_1                           -0.18547    0.03273      0  -5.666  < 1e-04 ***
nodeocov.shar1_1                           0.03428    0.02243      0   1.529 0.126332
nodeocov.sinc3_1                          -1.16378    0.14755      0  -7.888  < 1e-04 ***
nodeocov.intel3_1                          0.25319    0.15098      0   1.677 0.093538 .
nodeocov.amb3_1                            0.18902    0.08973      0   2.107 0.035145 *
edgecov.edge_matrices[["like"]]            0.58861    0.10484      0   5.614  < 1e-04 ***
edgecov.edge_matrices[["prob"]]            0.33193    0.07244      0   4.582  < 1e-04 ***
edgecov.edge_matrices[["attr"]]            0.51933    0.07256      0   7.157  < 1e-04 ***
edgecov.edge_matrices[["sinc"]]           -0.35765    0.08035      0  -4.451  < 1e-04 ***
edgecov.edge_matrices[["fun"]]             0.26349    0.09090      0   2.899 0.003748 **
edgecov.edge_matrices[["shar"]]            0.36618    0.07207      0   5.081  < 1e-04 ***
nodematch.gender                              -Inf    0.00000      0    -Inf  < 1e-04 ***
absdiff.clubbing                          -0.09572    0.04584      0  -2.088 0.036780 *
absdiff.concerts                          -0.11968    0.06671      0  -1.794 0.072821 .
absdiff.shopping                          -0.10697    0.05047      0  -2.119 0.034052 *
absdiff.yoga                              -0.10936    0.05095      0  -2.146 0.031856 *
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

mutual -0.13990 0.24718 0 -0.566 0.571407

edgecov.edge_matrices[["prob"]] 0.33193 0.07244 0 4.582 < 1e-04 ***

Discussion

Model Improvements

- Explore other endogenous terms
- All 21 waves
- Other dependence assumptions (Markov)
- Interaction terms

Variable Selection

- Choose those most supported by sociological studies

Question of Dependence

	Estimate	Std. Error	MCMC %	z value	Pr(>|z|)	
edges	-1.82014	0.07645	0	-23.81	<1e-04	***
mutual	1.41831	0.18590	0	7.63	<1e-04	***

- Influence of reciprocity
  - Mutual and Edges term alone: conditionally dependent
  - With other predictors: conditionally independent

Back to the Original Research

Were their assumptions (or lack thereof) correct?

- Linear regression
- Assumptions not mentioned (outside of economic ones)

Bibliography

Fisman, R., Iyengar, S. S., Kamenica, E., & Simonson, I. (2006). Gender Differences in Mate Selection: Evidence From a Speed Dating Experiment. *The Quarterly Journal of Economics, 121(2)*, 673–697. https://doi.org/10.1162/qjec.2006.121.2.673

U.S. Census Bureau (2002). *Table FG4. Married Couple Family Groups, By Presence Of Own Children In Specific Age Groups, And Age, Earnings, Education, And Race And Hispanic Origin Of Both Spouses: March 2002*. Retrieved from https://www.census.gov/data/tables/2002/demo/families/families-living-arrangements.html

Proof A

Proof for normalizing constant $k(\theta)$

The normalizing constant $k(\theta)$ is defined by the following:

$$k(\theta) = \sum_{y \in 0,1^{n \times n}} \exp(\theta^\top g(y))$$

For each dyad (i, j), the change statistics are defined as $\delta_{ij}(y) = g(y_{ij}^+) - g(y_{ij}^-)$. Using this definition, and under dyadic independence, the network $g(y)$ can be expressed as a sum over individual dyads as such that:

$$\begin{aligned} g(y) &= \sum_{i,j} \left[y_{ij} g(y_{ij}^+) + (1 - y_{ij}) g(y_{ij}^-) \right] \\ &= \\ &= \sum_{i,j} \left[g(y_{ij}^-) + y_{ij} \delta_{ij}(y) \right] \end{aligned}$$

Substituting this into $k(\theta)$, we get the following:

$$
\begin{aligned}
k(\theta) &= \sum_{y \in \{0,1\}^{n \times n}} \exp\left(\theta^\top \sum_{i,j} \left[g(y_{\bar{ij}}) + y_{ij}\delta_{ij}(y)\right]\right) \\
&= \sum_{y} \exp\left(\sum_{i,j} \theta^\top g(y_{\bar{ij}}) + \sum_{i,j} \theta^\top y_{ij}\delta_{ij}(y)\right)
\end{aligned}
$$

The exponential of a sum could be rewritten as a product, so we factorize as in the following:

$$
\begin{aligned}
k(\theta) &= \sum_{y} \exp\left(\sum_{i,j} \theta^\top g(y_{\bar{ij}}) + \sum_{i,j} \theta^\top y_{ij}\delta_{ij}(y)\right) \\
&= \sum_{y} \prod_{i,j} \exp\left(\theta^\top g(y_{\bar{ij}})\right) \cdot \exp\left(\theta^\top y_{ij}\delta_{ij}(y)\right)
\end{aligned}
$$

We can separate the sums over the dyads under dyadic independence:

$$
\begin{aligned}
k(\theta) &= \sum_{y} \prod_{i,j} \exp\left(\theta^\top g(y_{ij}^-)\right) \cdot \exp\left(\theta^\top y_{ij} \delta_{ij}(y)\right) \\
&= \prod_{i,j} \left[\sum_{y_{ij} \in \{0,1\}} \exp\left(\theta^\top g(y_{ij}^-)\right) \cdot \exp\left(\theta^\top y_{ij} \delta_{ij}(y)\right) \right]
\end{aligned}
$$

Now for each dyad (i, j), we can calculate the sum over two possible cases, $y_{ij} = 1$ and $y_{ij} = 0$.

For $y_{ij} = 0$, $\exp\left(\theta^\top g(y_{ij}^-)\right) \cdot \exp\left(\theta^\top \cdot 0 \cdot \delta_{ij}(y)\right) = \exp\left(\theta^\top g(y_{ij}^-)\right) \cdot 1 = \exp\left(\theta^\top g(y_{ij}^-)\right)$.

For $y_{ij} = 1$, using the definition of the change statistic, we derive the following:

$$
\begin{aligned}
\exp\left(\theta^\top g(y_{ij}^-)\right) \cdot \exp\left(\theta^\top \cdot 1 \cdot \delta_{ij}(y)\right) &= \exp\left(\theta^\top g(y_{ij}^-)\right) \cdot \exp\left(\theta^\top \delta_{ij}(y)\right) \\
&= \exp\left(\theta^\top g(y_{ij}^-)\right) \cdot \exp\left(\theta^\top g(y_{ij}^+) - \theta^\top g(y_{ij}^-)\right) \\
&= \exp\left(\theta^\top g(y_{ij}^-)\right) \cdot \frac{\exp\left(\theta^\top g(y_{ij}^+)\right)}{\exp\left(\theta^\top g(y_{ij}^-)\right)} \\
&= \exp\left(\theta^\top g(y_{ij}^+)\right)
\end{aligned}
$$

Therefore, for each dyad (i, j), the sum comes down to the following:

$$\sum_{y_{ij} \in \{0,1\}} \exp\left(\theta^\top g(y_{ij}^-)\right) \cdot \exp\left(\theta^\top y_{ij} \delta_{ij}(y)\right) = \exp\left(\theta^\top g(y_{ij}^-)\right) + \exp\left(\theta^\top g(y_{ij}^+)\right)$$

Plugging this back into $k(\theta)$:

$$\begin{aligned} k(\theta) &= \prod_{i,j} \left[\sum_{y_{ij} \in \{0,1\}} \exp\left(\theta^\top g(y_{ij}^-)\right) \cdot \exp\left(\theta^\top y_{ij} \delta_{ij}(y)\right) \right] \\ &= \prod_{i,j} \left[\exp\left(\theta^\top g(y_{ij}^-)\right) + \exp\left(\theta^\top g(y_{ij}^+)\right) \right] \end{aligned}$$

Proof B

Proof for $P(Y_{ij} = 1)$ and $P(Y_{ij} = 0)$

Previously we have derived the following:

$$P(Y = y) = \prod_{i,j} \frac{\exp(\theta^\top g(y_{ij}^-)) \exp(\theta^\top y_{ij}(\delta_{ij}(y)))}{\exp(\theta^\top g(y_{ij}^-)) + \exp(\theta^\top g(y_{ij}^+))}$$

If we plug in case by case, using

$$\exp\left(\theta^\top y_{ij} \delta_{ij}(y)\right) = \begin{cases} \exp\left(\theta^\top \delta_{ij}(y)\right) & \text{if } y_{ij} = 1 \\ 1 & \text{if } y_{ij} = 0 \end{cases}$$

we derive the following:

$$
\begin{aligned}
P(Y_{ij}=1) &= \frac{\exp(\theta^\top g(y_{ij}^-)) \cdot \exp\left(\theta^\top \delta_{ij}(y)\right)}{\exp(\theta^\top g(y_{ij}^-)) + \exp(\theta^\top g(y_{ij}^+))} \\
&= \frac{\exp(\theta^\top g(y_{ij}^-)) \cdot \exp\left(\theta^\top \left(g(y_{ij}^+) - g(y_{ij}^-)\right)\right)}{\exp(\theta^\top g(y_{ij}^-)) + \exp(\theta^\top g(y_{ij}^+))} \\
&= \frac{\exp(\theta^\top g(y_{ij}^+))}{\exp(\theta^\top g(y_{ij}^-)) + \exp(\theta^\top g(y_{ij}^+))}
\end{aligned}
$$

$$
\begin{aligned}
P(Y_{ij}=0) &= \frac{\exp(\theta^\top g(y_{ij}^-)) \cdot 1}{\exp(\theta^\top g(y_{ij}^-)) + \exp(\theta^\top g(y_{ij}^+))} \\
&= \frac{\exp(\theta^\top g(y_{ij}^-))}{\exp(\theta^\top g(y_{ij}^-)) + \exp(\theta^\top g(y_{ij}^+))}
\end{aligned}
$$