Physical Distancing to Prevent Influenza-Like-Illness on College Campuses: the eX-FLU Trial

Brian Richardson

Department of Statistical Sciences, Wake Forest University

September 08, 2026

Scan for slides

Coauthors

Michael Hudgens, PhD
UNC Biostatistics

Allison Aiello, PhD
Columbia Epidemiology

eX-FLU Trial

- **eX-FLU**: trial to evaluate a physical distancing intervention on a college campus during flu season (Aiello et al., 2016; Zivich et al., 2020)

Design and methods of a social network isolation study for reducing respiratory infection transmission: The eX-FLU cluster randomized trial

Allison E. Aiello[a,*], Amanda M. Simanek[b,1], Marisa C. Eisenberg[c], Alison R. Walsh[c], Brian Davis[c], Erik Volz[d,1], Caroline Cheng[c], Jeanette J. Rainey[e], Amra Uzicanin[e], Hongjiang Gao[e], Nathaniel Osgood[f], Dylan Knowles[f], Kevin Stanley[f], Kara Tarter[c], Arnold S. Monto[c]

eX-FLU Trial

- **eX-FLU**: trial to evaluate a physical distancing intervention on a college campus during flu season (Aiello et al., 2016; Zivich et al., 2020)
- **Intervention**: encouragement to isolate in dorm for three days upon developing symptoms of **influenza-like illness** (ILI)

Design and methods of a social network isolation study for reducing respiratory infection transmission: The eX-FLU cluster randomized trial

Allison E. Aiello[a,*], Amanda M. Simanek[b,1], Marisa C. Eisenberg[c], Alison R. Walsh[c], Brian Davis[c], Erik Volz[d,1], Caroline Cheng[c], Jeanette J. Rainey[e], Amra Uzicanin[e], Hongjiang Gao[e], Nathaniel Osgood[f], Dylan Knowles[f], Kevin Stanley[f], Kara Tarter[c], Arnold S. Monto[c]

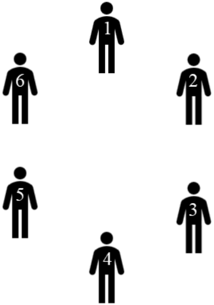
eX-FLU Trial

- **eX-FLU**: trial to evaluate a physical distancing intervention on a college campus during flu season (Aiello et al., 2016; Zivich et al., 2020)
- **Intervention**: encouragement to isolate in dorm for three days upon developing symptoms of **influenza-like illness** (ILI)
- **Central question**: does the encouragement-to-isolate intervention prevent ILI infection?

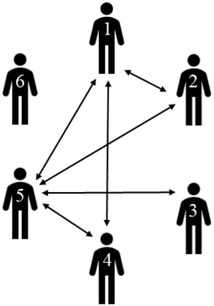
Design and methods of a social network isolation study for reducing respiratory infection transmission: The eX-FLU cluster randomized trial

Allison E. Aiello[a,*], Amanda M. Simanek[b,1], Marisa C. Eisenberg[c], Alison R. Walsh[c], Brian Davis[c], Erik Volz[d,1], Caroline Cheng[c], Jeanette J. Rainey[e], Amra Uzicanin[e], Hongjiang Gao[e], Nathaniel Osgood[f], Dylan Knowles[f], Kevin Stanley[f], Kara Tarter[c], Arnold S. Monto[c]

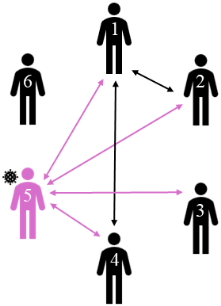
Transmission of Influenza-Like-Illness



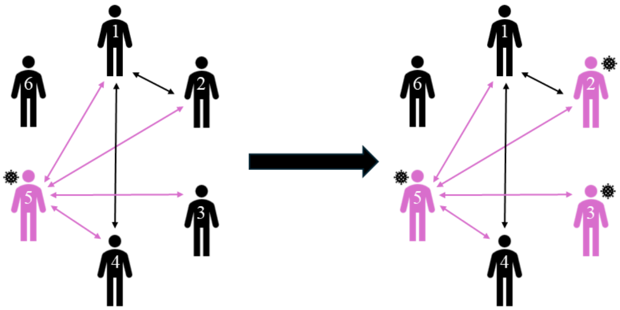
Transmission of Influenza-Like-Illness



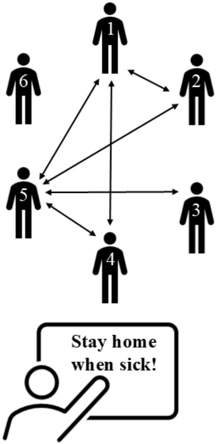
Transmission of Influenza-Like-Illness



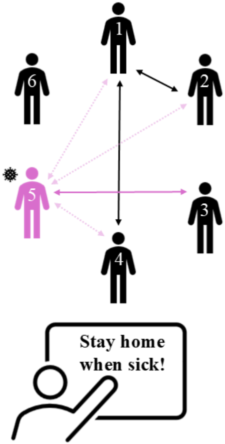
Transmission of Influenza-Like-Illness



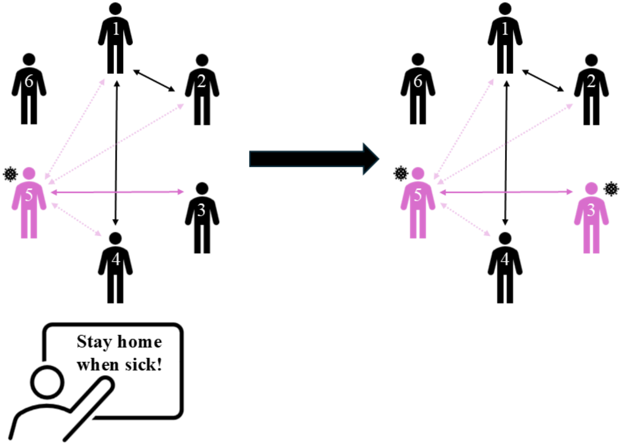
Example I: Intervention Affects Network and Infections



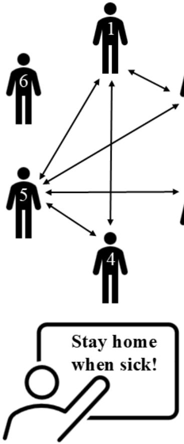
Example I: Intervention Affects Network and Infections



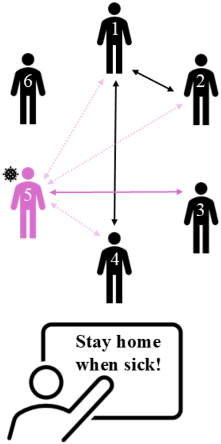
Example I: Intervention Affects Network and Infections



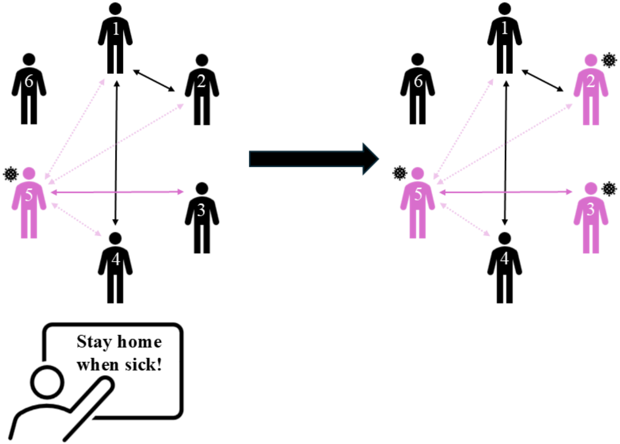
Example II: Intervention Affects Network, Not Infections



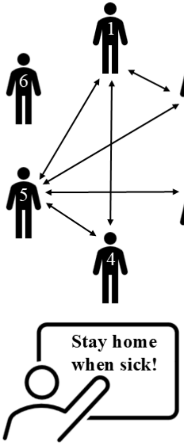
Example II: Intervention Affects Network, Not Infections



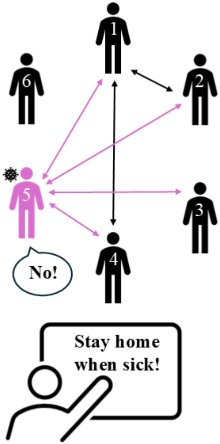
Example II: Intervention Affects Network, Not Infections



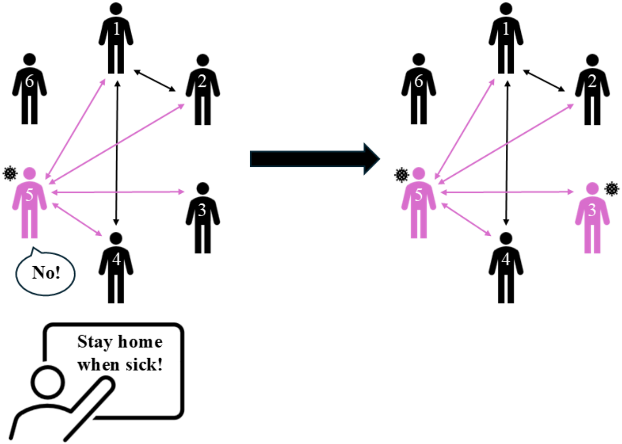
Example III: Intervention Affects Infections, Not Network



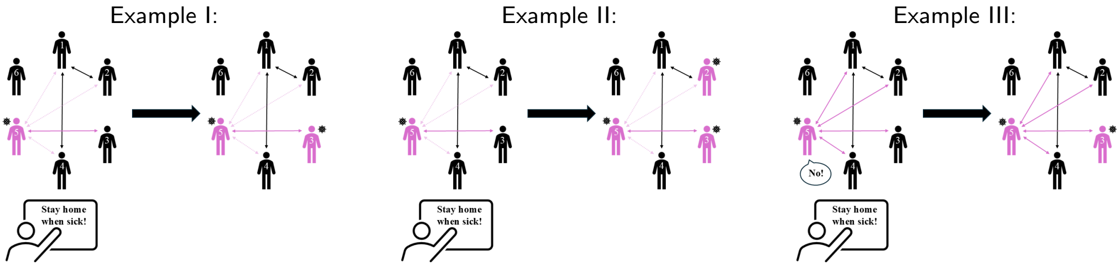
Example III: Intervention Affects Infections, Not Network



Example III: Intervention Affects Infections, Not Network



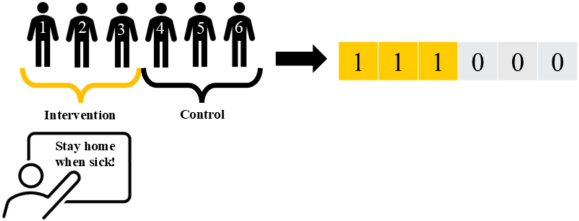
Central Question: Does the Intervention Prevent ILI Infection?



In Examples I and III, the answer is yes

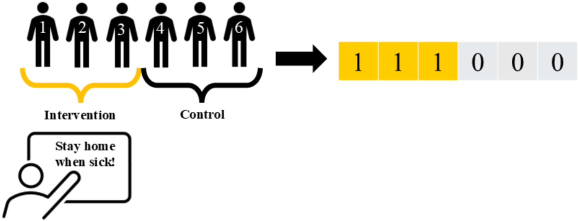
eX-FLU Observed Data

- Baseline randomization assignments $\mathbf{Z} = (Z_1, \ldots, Z_n) \in \mathcal{Z}$, for $Z_i = \mathbb{1}(\text{student } i \text{ gets intervention})$



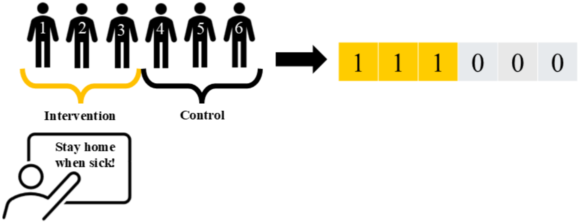
eX-FLU Observed Data

- Baseline randomization assignments $\mathbf{Z} = (Z_1, \ldots, Z_n) \in \mathcal{Z}$, for $Z_i = \mathbb{1}(\text{student } i \text{ gets intervention})$
  - randomization distribution $r(\mathbf{z})$



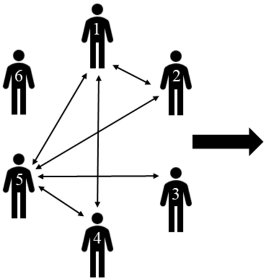
eX-FLU Observed Data

- Baseline randomization assignments $\mathbf{Z} = (Z_1, \ldots, Z_n) \in \mathcal{Z}$, for $Z_i = \mathbb{1}(\text{student } i \text{ gets intervention})$
  - randomization distribution $r(\mathbf{z})$
- At weeks $k = 1, \ldots, \tau$, we observe:



eX-FLU Observed Data

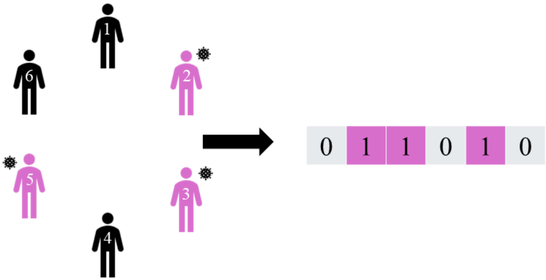
- Baseline randomization assignments
  $\mathbf{Z} = (Z_1, \ldots, Z_n) \in \mathcal{Z}$, for
  $Z_i = \mathbb{1}(\text{student } i \text{ gets intervention})$
  - randomization distribution $r(\mathbf{z})$
- At weeks $k = 1, \ldots, \tau$, we observe:
  - networks $\mathbf{A}^k = [A_{ij}^k]$, for
    $A_{ij}^k = \mathbb{1}(\text{students } i, j \text{ in contact at week } k)$



	1	2	3	4	5	6
1	0	1	0	1	1	0
2	1	0	0	0	1	0
3	0	0	0	0	1	0
4	1	0	0	0	1	0
5	1	1	1	1	0	0
6	0	0	0	0	0	0

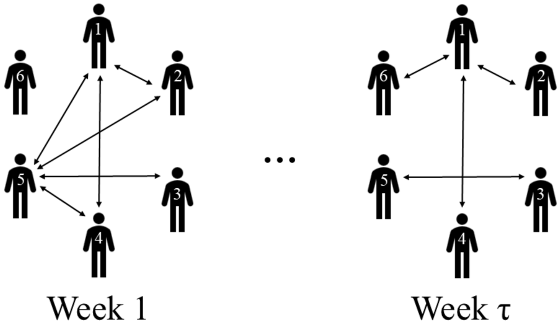
eX-FLU Observed Data

- Baseline randomization assignments $\mathbf{Z} = (Z_1, \ldots, Z_n) \in \mathcal{Z}$, for $Z_i = \mathbb{1}(\text{student } i \text{ gets intervention})$
  - randomization distribution $r(\mathbf{z})$
- At weeks $k = 1, \ldots, \tau$, we observe:
  - networks $\mathbf{A}^k = [A_{ij}^k]$, for $A_{ij}^k = \mathbb{1}(\text{students } i, j \text{ in contact at week } k)$
  - infections $\mathbf{Y}^k = (Y_1^k, \ldots, Y_n^k)$, for $Y_i^k = \mathbb{1}(\text{student } i \text{ infected at week } k)$



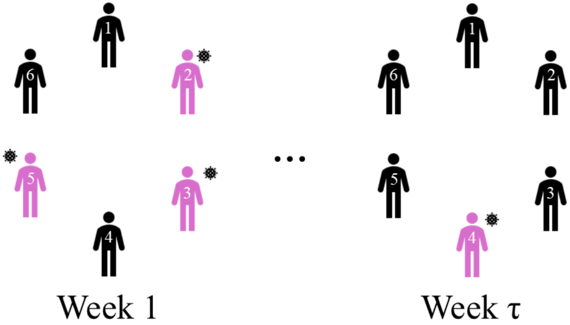
eX-FLU Observed Data

- Baseline randomization assignments $\mathbf{Z} = (Z_1, \ldots, Z_n) \in \mathcal{Z}$, for $Z_i = \mathbb{1}(\text{student } i \text{ gets intervention})$
  - randomization distribution $r(\mathbf{z})$
- At weeks $k = 1, \ldots, \tau$, we observe:
  - networks $\mathbf{A}^k = [A^k_{ij}]$, for $A^k_{ij} = \mathbb{1}(\text{students } i, j \text{ in contact at week } k)$
  - infections $\mathbf{Y}^k = (Y^k_1, \ldots, Y^k_n)$, for $Y^k_i = \mathbb{1}(\text{student } i \text{ infected at week } k)$
- Network history: $\overline{\mathbf{A}} = \{\mathbf{A}^k\}_{k=1}^{\tau}$

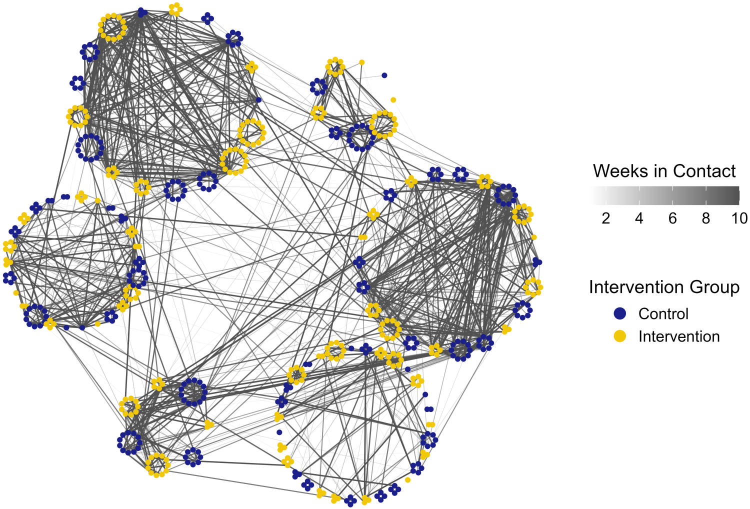


eX-FLU Observed Data

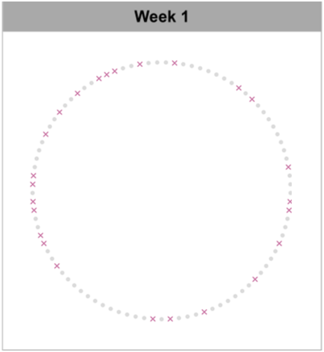
- Baseline randomization assignments $\mathbf{Z} = (Z_1, \ldots, Z_n) \in \mathcal{Z}$, for $Z_i = \mathbb{1}(\text{student } i \text{ gets intervention})$
  - randomization distribution $r(\mathbf{z})$
- At weeks $k = 1, \ldots, \tau$, we observe:
  - networks $\mathbf{A}^k = [A^k_{ij}]$, for $A^k_{ij} = \mathbb{1}(\text{students } i, j \text{ in contact at week } k)$
  - infections $\mathbf{Y}^k = (Y^k_1, \ldots, Y^k_n)$, for $Y^k_i = \mathbb{1}(\text{student } i \text{ infected at week } k)$
- Network history: $\overline{\mathbf{A}} = \{\mathbf{A}^k\}_{k=1}^{\tau}$
- Infection history: $\overline{\mathbf{Y}} = \{\mathbf{Y}^k\}_{k=1}^{\tau}$



eX-FLU Observed Data



eX-FLU Observed Data



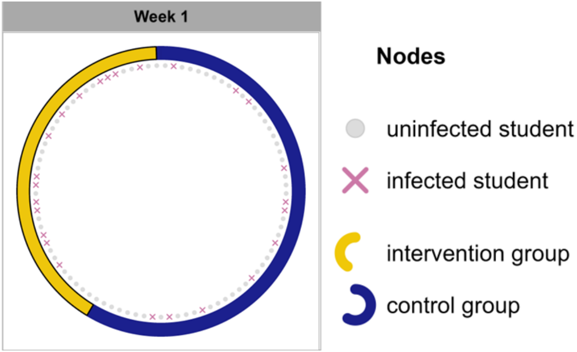
Nodes

uninfected student

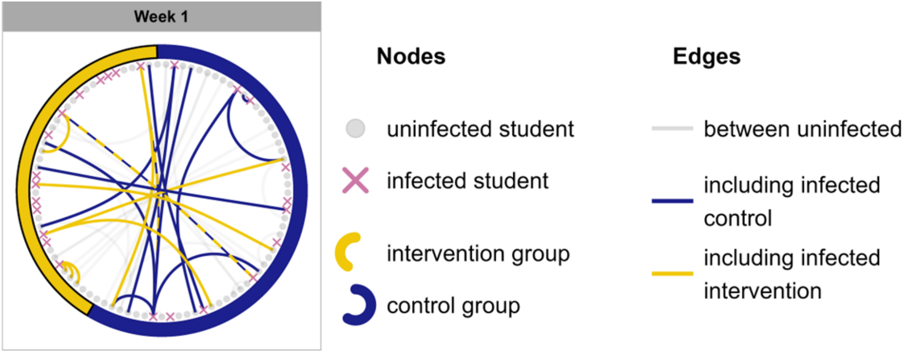
infected student

(93 out of 579 students with at least one infection)

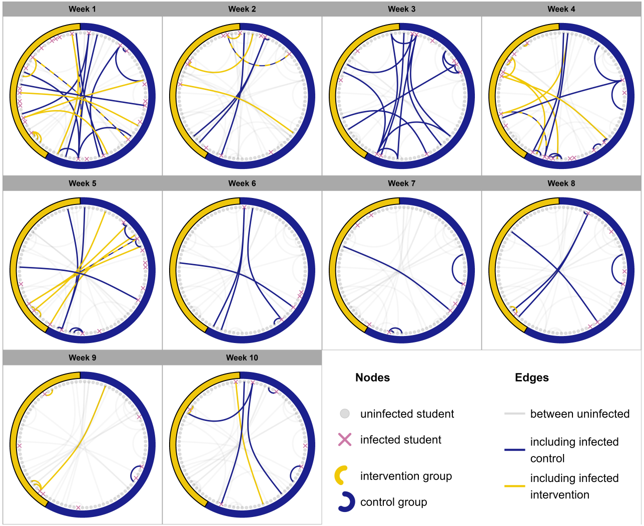
eX-FLU Observed Data



eX-FLU Observed Data



eX-FLU Observed Data



eX-FLU Potential Outcomes

eX-FLU Potential Outcomes

Potential outcomes are defined for the networks and ILI infections:

$$\overline{\mathbf{A}}(\mathbf{z}), \quad \overline{\mathbf{Y}}(\mathbf{z}) \quad \text{for } \mathbf{z} \in \mathcal{Z}$$

eX-FLU Potential Outcomes

Potential outcomes are defined for the networks and ILI infections:

$$\overline{\mathbf{A}}(\mathbf{z}), \quad \overline{\mathbf{Y}}(\mathbf{z}) \quad \text{for } \mathbf{z} \in \mathcal{Z}$$

- $\overline{\mathbf{A}}(\mathbf{z})$: sequence of networks that **would** occur if students **were** encouraged to isolate according to $\mathbf{z}$

eX-FLU Potential Outcomes

Potential outcomes are defined for the networks and ILI infections:

$$\overline{\mathbf{A}}(\mathbf{z}), \quad \overline{\mathbf{Y}}(\mathbf{z}) \quad \text{for } \mathbf{z} \in \mathcal{Z}$$

- $\overline{\mathbf{A}}(\mathbf{z})$: sequence of networks that **would** occur if students **were** encouraged to isolate according to $\mathbf{z}$
- $\overline{\mathbf{Y}}(\mathbf{z})$: sequence of infection statuses that **would** occur if students **were** encouraged to isolate according to $\mathbf{z}$

eX-FLU Potential Outcomes

Potential outcomes are defined for the networks and ILI infections:

$$\overline{\mathbf{A}}(\mathbf{z}), \quad \overline{\mathbf{Y}}(\mathbf{z}) \quad \text{for } \mathbf{z} \in \mathcal{Z}$$

- $\overline{\mathbf{A}}(\mathbf{z})$: sequence of networks that **would** occur if students **were** encouraged to isolate according to $\mathbf{z}$
- $\overline{\mathbf{Y}}(\mathbf{z})$: sequence of infection statuses that **would** occur if students **were** encouraged to isolate according to $\mathbf{z}$
- Both of these depend on the entire vector $\mathbf{z}$
  - potential outcomes for one student depend on intervention assignment of other students (**interference**) (Hudgens and Halloran, 2008)

eX-FLU Potential Outcomes

Potential outcomes are defined for the networks and ILI infections:

$$\overline{\mathbf{A}}(\mathbf{z}), \quad \overline{\mathbf{Y}}(\mathbf{z}) \quad \text{for } \mathbf{z} \in \mathcal{Z}$$

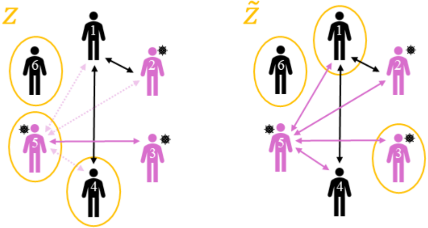
- $\overline{\mathbf{A}}(\mathbf{z})$: sequence of networks that **would** occur if students **were** encouraged to isolate according to $\mathbf{z}$
- $\overline{\mathbf{Y}}(\mathbf{z})$: sequence of infection statuses that **would** occur if students **were** encouraged to isolate according to $\mathbf{z}$
- Both of these depend on the entire vector $\mathbf{z}$
  - potential outcomes for one student depend on intervention assignment of other students (**interference**) (Hudgens and Halloran, 2008)
- Assume **causal consistency**:

$$\overline{\mathbf{A}} = \overline{\mathbf{A}}(\mathbf{Z}) \text{ and } \overline{\mathbf{Y}} = \overline{\mathbf{Y}}(\mathbf{Z})$$

eX-FLU Null Hypotheses

$$H_0^Y : \overline{\mathbf{Y}}(\mathbf{z}) = \overline{\mathbf{Y}}(\widetilde{\mathbf{z}}) \text{ for all } \mathbf{z}, \widetilde{\mathbf{z}} \in \mathcal{Z}$$

("the intervention has no effect on the infections")



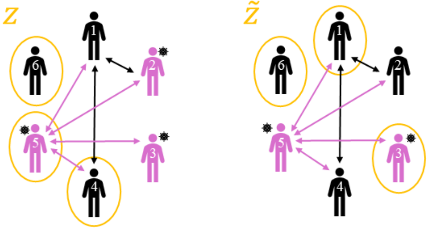
eX-FLU Null Hypotheses

$$H_0^Y : \overline{\mathbf{Y}}(\mathbf{z}) = \overline{\mathbf{Y}}(\widetilde{\mathbf{z}}) \text{ for all } \mathbf{z}, \widetilde{\mathbf{z}} \in \mathcal{Z}$$

(“the intervention has no effect on the infections”)

$$H_0^A : \overline{\mathbf{A}}(\mathbf{z}) = \overline{\mathbf{A}}(\widetilde{\mathbf{z}}) \text{ for all } \mathbf{z}, \widetilde{\mathbf{z}} \in \mathcal{Z}$$

(“the intervention has no effect the networks”)



eX-FLU Null Hypotheses

$$H_0^Y : \overline{\mathbf{Y}}(\mathbf{z}) = \overline{\mathbf{Y}}(\widetilde{\mathbf{z}}) \text{ for all } \mathbf{z}, \widetilde{\mathbf{z}} \in \mathcal{Z}$$

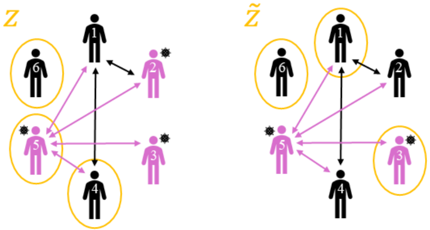
(“the intervention has no effect on the infections”)

$$H_0^A : \overline{\mathbf{A}}(\mathbf{z}) = \overline{\mathbf{A}}(\widetilde{\mathbf{z}}) \text{ for all } \mathbf{z}, \widetilde{\mathbf{z}} \in \mathcal{Z}$$

(“the intervention has no effect the networks”)

$$H_0^\sharp = H_0^Y \cap H_0^A : \overline{\mathbf{Y}}(\mathbf{z}) = \overline{\mathbf{Y}}(\widetilde{\mathbf{z}}) \text{ and } \overline{\mathbf{A}}(\mathbf{z}) = \overline{\mathbf{A}}(\widetilde{\mathbf{z}}) \text{ for all } \mathbf{z}, \widetilde{\mathbf{z}} \in \mathcal{Z}$$

(“the intervention has no effect on the infections or on the networks”)



eX-FLU Null Hypotheses

$$H_0^Y : \overline{\mathbf{Y}}(\mathbf{z}) = \overline{\mathbf{Y}}(\widetilde{\mathbf{z}}) \text{ for all } \mathbf{z}, \widetilde{\mathbf{z}} \in \mathcal{Z}$$

$$\text{(``the intervention has no effect on the infections'')}$$

$$H_0^A : \overline{\mathbf{A}}(\mathbf{z}) = \overline{\mathbf{A}}(\widetilde{\mathbf{z}}) \text{ for all } \mathbf{z}, \widetilde{\mathbf{z}} \in \mathcal{Z}$$

$$\text{(``the intervention has no effect the networks'')}$$

$$H_0^\sharp = H_0^Y \cap H_0^A : \overline{\mathbf{Y}}(\mathbf{z}) = \overline{\mathbf{Y}}(\widetilde{\mathbf{z}}) \text{ and } \overline{\mathbf{A}}(\mathbf{z}) = \overline{\mathbf{A}}(\widetilde{\mathbf{z}}) \text{ for all } \mathbf{z}, \widetilde{\mathbf{z}} \in \mathcal{Z}$$

$$\text{(``the intervention has no effect on the infections or on the networks'')}$$

Previous analyses of the eX-FLU trial (Alexandria et al., 2023) tested $H_0^\sharp$, but no analyses have tested H_0^Y

Testing $H_0^\sharp$

"How unlikely are the observed data under $H_0^\sharp$?"

Testing $H_0^\sharp$

"How unlikely are the observed data under $H_0^\sharp$?" We can answer this using a **randomization test** (Fisher, 1936; Alexandria et al., 2023; Zhang and Zhao, 2023; Ritzwoller et al., 2024)

Testing $H_0^\sharp$

"How unlikely are the observed data under $H_0^\sharp$?" We can answer this using a **randomization test** (Fisher, 1936; Alexandria et al., 2023; Zhang and Zhao, 2023; Ritzwoller et al., 2024)

- Choose a test statistic $T = T(\mathbf{Z}, \overline{\mathbf{A}}, \overline{\mathbf{Y}})$

Testing $H_0^\sharp$

"How unlikely are the observed data under $H_0^\sharp$?" We can answer this using a **randomization test** (Fisher, 1936; Alexandria et al., 2023; Zhang and Zhao, 2023; Ritzwoller et al., 2024)

- Choose a test statistic $T = T(\mathbf{Z}, \overline{\mathbf{A}}, \overline{\mathbf{Y}})$
- Find the distribution of T under $H_0^\sharp : \overline{\mathbf{Y}}(\mathbf{z}) = \overline{\mathbf{Y}}(\widetilde{\mathbf{z}})$ and $\overline{\mathbf{A}}(\mathbf{z}) = \overline{\mathbf{A}}(\widetilde{\mathbf{z}})$ for all $\mathbf{z}, \widetilde{\mathbf{z}} \in \mathcal{Z}$

Testing $H_0^\sharp$

"How unlikely are the observed data under $H_0^\sharp$?" We can answer this using a **randomization test** (Fisher, 1936; Alexandria et al., 2023; Zhang and Zhao, 2023; Ritzwoller et al., 2024)

- Choose a test statistic $T = T(\mathbf{Z}, \overline{\mathbf{A}}, \overline{\mathbf{Y}})$
- Find the distribution of T under $H_0^\sharp : \overline{\mathbf{Y}}(\mathbf{z}) = \overline{\mathbf{Y}}(\widetilde{\mathbf{z}})$ and $\overline{\mathbf{A}}(\mathbf{z}) = \overline{\mathbf{A}}(\widetilde{\mathbf{z}})$ for all $\mathbf{z}, \widetilde{\mathbf{z}} \in \mathcal{Z}$

Randomization	**Networks**	**Infections**	**Test Statistic**
$\mathbf{z}_1$			
$\mathbf{z}_2$			
$\dots$			
$\mathbf{z}_{\lvert\mathcal{Z}\rvert}$			

Testing $H_0^\sharp$

"How unlikely are the observed data under $H_0^\sharp$?" We can answer this using a **randomization test** (Fisher, 1936; Alexandria et al., 2023; Zhang and Zhao, 2023; Ritzwoller et al., 2024)

- Choose a test statistic $T = T(\mathbf{Z}, \overline{\mathbf{A}}, \overline{\mathbf{Y}})$
- Find the distribution of T under $H_0^\sharp : \overline{\mathbf{Y}}(\mathbf{z}) = \overline{\mathbf{Y}}(\widetilde{\mathbf{z}})$ and $\overline{\mathbf{A}}(\mathbf{z}) = \overline{\mathbf{A}}(\widetilde{\mathbf{z}})$ for all $\mathbf{z}, \widetilde{\mathbf{z}} \in \mathcal{Z}$

Randomization	Networks	Infections	Test Statistic
$\mathbf{z}_1$	$\overline{\mathbf{A}}(\mathbf{z}_1) =?$	$\overline{\mathbf{Y}}(\mathbf{z}_1) =?$	$T(\mathbf{z}_1, ?, ?)$
$\mathbf{z}_2$	$\overline{\mathbf{A}}(\mathbf{z}_2) =?$	$\overline{\mathbf{Y}}(\mathbf{z}_1) =?$	$T(\mathbf{z}_1, ?, ?)$
$\dots$	$\dots$	$\dots$	$\dots$
$\mathbf{z}_{\lvert\mathcal{Z}\rvert}$	$\overline{\mathbf{A}}(\mathbf{z}_{\lvert\mathcal{Z}\rvert}) =?$	$\overline{\mathbf{Y}}(\mathbf{z}_{\lvert\mathcal{Z}\rvert}) =?$	$T(\mathbf{z}_{\lvert\mathcal{Z}\rvert}, ?, ?)$

Testing $H_0^\sharp$

"How unlikely are the observed data under $H_0^\sharp$?" We can answer this using a **randomization test** (Fisher, 1936; Alexandria et al., 2023; Zhang and Zhao, 2023; Ritzwoller et al., 2024)

- Choose a test statistic $T = T(\mathbf{Z}, \overline{\mathbf{A}}, \overline{\mathbf{Y}})$
- Find the distribution of T under $H_0^\sharp : \overline{\mathbf{Y}}(\mathbf{z}) = \overline{\mathbf{Y}}(\widetilde{\mathbf{z}})$ and $\overline{\mathbf{A}}(\mathbf{z}) = \overline{\mathbf{A}}(\widetilde{\mathbf{z}})$ for all $\mathbf{z}, \widetilde{\mathbf{z}} \in \mathcal{Z}$

Randomization	**Networks**	**Infections**	**Test Statistic**
$\mathbf{z}_1$	$\overline{\mathbf{A}}(\mathbf{z}_1) = \overline{\mathbf{A}}$	$\overline{\mathbf{Y}}(\mathbf{z}_1) = \overline{\mathbf{Y}}$	$T(\mathbf{z}_1, \overline{\mathbf{A}}, \overline{\mathbf{Y}})$
$\mathbf{z}_2$	$\overline{\mathbf{A}}(\mathbf{z}_2) = \overline{\mathbf{A}}$	$\overline{\mathbf{Y}}(\mathbf{z}_2) = \overline{\mathbf{Y}}$	$T(\mathbf{z}_2, \overline{\mathbf{A}}, \overline{\mathbf{Y}})$
$\ldots$	$\ldots$	$\ldots$	$\ldots$
$\mathbf{z}_{\lvert\mathcal{Z}\rvert}$	$\overline{\mathbf{A}}(\mathbf{z}_{\lvert\mathcal{Z}\rvert}) = \overline{\mathbf{A}}$	$\overline{\mathbf{Y}}(\mathbf{z}_{\lvert\mathcal{Z}\rvert}) = \overline{\mathbf{Y}}$	$T(\mathbf{z}_{\lvert\mathcal{Z}\rvert}, \overline{\mathbf{A}}, \overline{\mathbf{Y}})$

Testing $H_0^\sharp$

"How unlikely are the observed data under $H_0^\sharp$?" We can answer this using a **randomization test** (Fisher, 1936; Alexandria et al., 2023; Zhang and Zhao, 2023; Ritzwoller et al., 2024)

- Choose a test statistic $T = T(\mathbf{Z}, \overline{\mathbf{A}}, \overline{\mathbf{Y}})$
- Find the distribution of T under $H_0^\sharp : \overline{\mathbf{Y}}(\mathbf{z}) = \overline{\mathbf{Y}}(\tilde{\mathbf{z}})$ and $\overline{\mathbf{A}}(\mathbf{z}) = \overline{\mathbf{A}}(\tilde{\mathbf{z}})$ for all $\mathbf{z}, \tilde{\mathbf{z}} \in \mathcal{Z}$
  - This is possible since $\overline{\mathbf{A}}(\mathbf{z}), \overline{\mathbf{Y}}(\mathbf{z})$ are **imputable** under $H_0^\sharp$

Randomization	Networks	Infections	Test Statistic
$\mathbf{z}_1$	$\overline{\mathbf{A}}(\mathbf{z}_1) = \overline{\mathbf{A}}$	$\overline{\mathbf{Y}}(\mathbf{z}_1) = \overline{\mathbf{Y}}$	$T(\mathbf{z}_1, \overline{\mathbf{A}}, \overline{\mathbf{Y}})$
$\mathbf{z}_2$	$\overline{\mathbf{A}}(\mathbf{z}_2) = \overline{\mathbf{A}}$	$\overline{\mathbf{Y}}(\mathbf{z}_2) = \overline{\mathbf{Y}}$	$T(\mathbf{z}_2, \overline{\mathbf{A}}, \overline{\mathbf{Y}})$
$\dots$	$\dots$	$\dots$	$\dots$
$\mathbf{z}_{\lvert\mathcal{Z}\rvert}$	$\overline{\mathbf{A}}(\mathbf{z}_{\lvert\mathcal{Z}\rvert}) = \overline{\mathbf{A}}$	$\overline{\mathbf{Y}}(\mathbf{z}_{\lvert\mathcal{Z}\rvert}) = \overline{\mathbf{Y}}$	$T(\mathbf{z}_{\lvert\mathcal{Z}\rvert}, \overline{\mathbf{A}}, \overline{\mathbf{Y}})$

Testing $H_0^\sharp$

"How unlikely are the observed data under $H_0^\sharp$?" We can answer this using a **randomization test** (Fisher, 1936; Alexandria et al., 2023; Zhang and Zhao, 2023; Ritzwoller et al., 2024)

- Choose a test statistic $T = T(\mathbf{Z}, \overline{\mathbf{A}}, \overline{\mathbf{Y}})$
- Find the distribution of T under $H_0^\sharp : \overline{\mathbf{Y}}(\mathbf{z}) = \overline{\mathbf{Y}}(\widetilde{\mathbf{z}})$ and $\overline{\mathbf{A}}(\mathbf{z}) = \overline{\mathbf{A}}(\widetilde{\mathbf{z}})$ for all $\mathbf{z}, \widetilde{\mathbf{z}} \in \mathcal{Z}$
  - This is possible since $\overline{\mathbf{A}}(\mathbf{z}), \overline{\mathbf{Y}}(\mathbf{z})$ are **imputable** under $H_0^\sharp \implies H_0^\sharp$ is **sharp**

Randomization	Networks	Infections	Test Statistic
$\mathbf{z}_1$	$\overline{\mathbf{A}}(\mathbf{z}_1) = \overline{\mathbf{A}}$	$\overline{\mathbf{Y}}(\mathbf{z}_1) = \overline{\mathbf{Y}}$	$T(\mathbf{z}_1, \overline{\mathbf{A}}, \overline{\mathbf{Y}})$
$\mathbf{z}_2$	$\overline{\mathbf{A}}(\mathbf{z}_2) = \overline{\mathbf{A}}$	$\overline{\mathbf{Y}}(\mathbf{z}_2) = \overline{\mathbf{Y}}$	$T(\mathbf{z}_2, \overline{\mathbf{A}}, \overline{\mathbf{Y}})$
$\dots$	$\dots$	$\dots$	$\dots$
$\mathbf{z}_{\lvert\mathcal{Z}\rvert}$	$\overline{\mathbf{A}}(\mathbf{z}_{\lvert\mathcal{Z}\rvert}) = \overline{\mathbf{A}}$	$\overline{\mathbf{Y}}(\mathbf{z}_{\lvert\mathcal{Z}\rvert}) = \overline{\mathbf{Y}}$	$T(\mathbf{z}_{\lvert\mathcal{Z}\rvert}, \overline{\mathbf{A}}, \overline{\mathbf{Y}})$

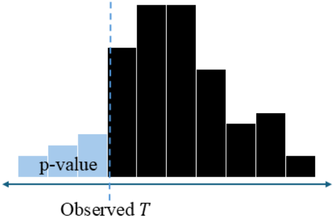
Testing $H_0^\sharp$

"How unlikely are the observed data under $H_0^\sharp$?" We can answer this using a **randomization test** (Fisher, 1936; Alexandria et al., 2023; Zhang and Zhao, 2023; Ritzwoller et al., 2024)

- Choose a test statistic $T = T(\mathbf{Z}, \overline{\mathbf{A}}, \overline{\mathbf{Y}})$
- Find the distribution of T under $H_0^\sharp : \overline{\mathbf{Y}}(\mathbf{z}) = \overline{\mathbf{Y}}(\widetilde{\mathbf{z}})$ and $\overline{\mathbf{A}}(\mathbf{z}) = \overline{\mathbf{A}}(\widetilde{\mathbf{z}})$ for all $\mathbf{z}, \widetilde{\mathbf{z}} \in \mathcal{Z}$
  - This is possible since $\overline{\mathbf{A}}(\mathbf{z}), \overline{\mathbf{Y}}(\mathbf{z})$ are **imputable** under $H_0^\sharp \implies H_0^\sharp$ is **sharp**
- Compute a p-value: $\rho^\sharp = F_T(T|H_0^\sharp)$, where

$$F_T(t|H_0^\sharp) = \Pr\left(T \le t|H_0^\sharp\right) = \sum_{\mathbf{z} \in \mathcal{Z}} r(\mathbf{z}) \mathbb{1}\{T(\mathbf{z}, \overline{\mathbf{A}}, \overline{\mathbf{Y}}) \le t\}$$

  is the **cumulative distribution function** (CDF) of T



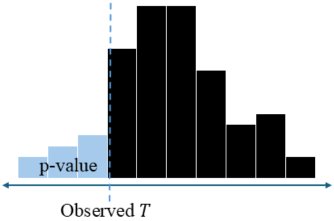
Testing $H_0^\sharp$

"How unlikely are the observed data under $H_0^\sharp$?" We can answer this using a **randomization test** (Fisher, 1936; Alexandria et al., 2023; Zhang and Zhao, 2023; Ritzwoller et al., 2024)

- Choose a test statistic $T = T(\mathbf{Z}, \overline{\mathbf{A}}, \overline{\mathbf{Y}})$
- Find the distribution of T under $H_0^\sharp : \overline{\mathbf{Y}}(\mathbf{z}) = \overline{\mathbf{Y}}(\tilde{\mathbf{z}})$ and $\overline{\mathbf{A}}(\mathbf{z}) = \overline{\mathbf{A}}(\tilde{\mathbf{z}})$ for all $\mathbf{z}, \tilde{\mathbf{z}} \in \mathcal{Z}$
  - This is possible since $\overline{\mathbf{A}}(\mathbf{z}), \overline{\mathbf{Y}}(\mathbf{z})$ are **imputable** under $H_0^\sharp \implies H_0^\sharp$ is **sharp**
- Compute a p-value: $\rho^\sharp = F_T(T|H_0^\sharp)$, where

$$F_T(t|H_0^\sharp) = \Pr\left(T \le t|H_0^\sharp\right) = \sum_{\mathbf{z}\in\mathcal{Z}} r(\mathbf{z})\mathbb{1}\{T(\mathbf{z}, \overline{\mathbf{A}}, \overline{\mathbf{Y}}) \le t\}$$

  is the **cumulative distribution function** (CDF) of T
- Reject $H_0^\sharp$ if $\rho^\sharp \le 0.05$



Testing $H_0^\sharp$

- This test of $H_0^\sharp$ controls the **type I error rate** exactly:

$$\Pr\left(\rho^\sharp \leq \alpha | H_0^\sharp\right) \leq \alpha$$

for any $\alpha \in [0, 1]$, for any sample size, and for any choice of test statistic

Testing $H_0^\sharp$

- This test of $H_0^\sharp$ controls the **type I error rate** exactly:

$$\Pr\left(\rho^\sharp \leq \alpha | H_0^\sharp\right) \leq \alpha$$

  for any $\alpha \in [0, 1]$, for any sample size, and for any choice of test statistic

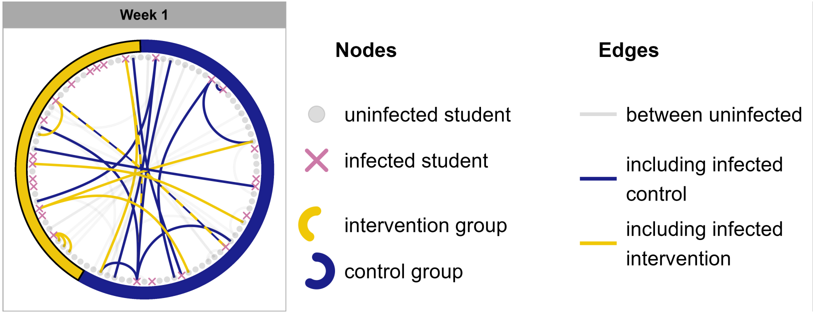
- The **power** of this test:

$$\Pr\left(\rho^\sharp \leq \alpha | H_1^\sharp\right)$$

  depends on the choice of test statistic

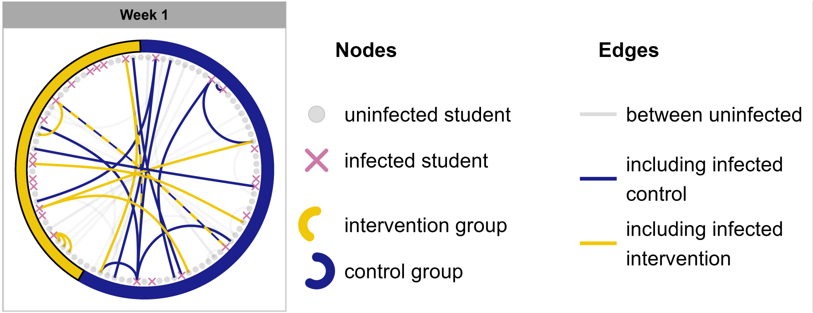
Choice of Test Statistic

- 93 (out of 579) students with at least one infection



Choice of Test Statistic

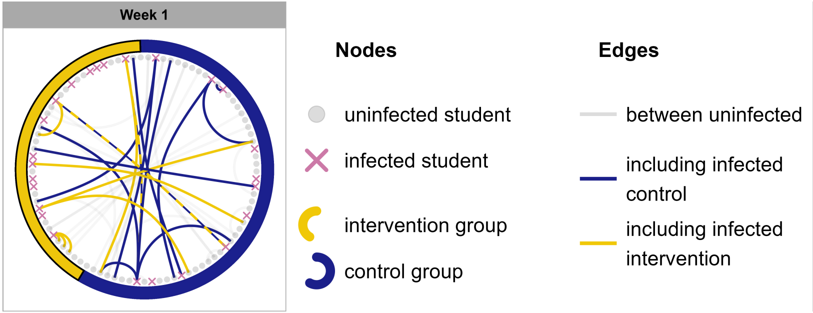
- 93 (out of 579) students with at least one infection
- Bold edges represent possible transmission events



Choice of Test Statistic

- 93 (out of 579) students with at least one infection
- Bold edges represent possible transmission events
  - Proportion of possible transmission events attributable to students in the intervention group:

$$T = \frac{\text{number of yellow edges}}{\text{total number of edges}} = \frac{\sum_{k=2}^{\tau} \sum_{i=1}^{n} \sum_{j \neq i} Z_i Y_i^{k-1} A_{ij}^{k-1}}{\sum_{k=2}^{\tau} \sum_{i=1}^{n} \sum_{j \neq i} Y_i^{k-1} A_{ij}^{k-1}} = 0.359$$



Testing H_0^Y

"How unlikely are the observed data under H_0^Y?"

Testing H_0^Y

"How unlikely are the observed data under H_0^Y?"

- Choose a test statistic $T = T(\mathbf{Z}, \overline{\mathbf{A}}, \overline{\mathbf{Y}})$

Testing H_0^Y

"How unlikely are the observed data under H_0^Y?"

- Choose a test statistic $T = T(\mathbf{Z}, \overline{\mathbf{A}}, \overline{\mathbf{Y}})$
- Find the distribution of T under $H_0^Y : \overline{\mathbf{Y}}(\mathbf{z}) = \overline{\mathbf{Y}}(\widetilde{\mathbf{z}})$ for all $\mathbf{z}, \widetilde{\mathbf{z}} \in \mathcal{Z}$

Testing H_0^Y

"How unlikely are the observed data under H_0^Y?"

- Choose a test statistic $T = T(\mathbf{Z}, \overline{\mathbf{A}}, \overline{\mathbf{Y}})$
- Find the distribution of T under $H_0^Y : \overline{\mathbf{Y}}(\mathbf{z}) = \overline{\mathbf{Y}}(\widetilde{\mathbf{z}})$ for all $\mathbf{z}, \widetilde{\mathbf{z}} \in \mathcal{Z}$

Randomization	**Networks**	**Infections**	**Test Statistic**
$\mathbf{z}_1$			
$\mathbf{z}_2$			
$\ldots$			
$\mathbf{z}_{\lvert\mathcal{Z}\rvert}$			

Testing H_0^Y

"How unlikely are the observed data under H_0^Y?"

- Choose a test statistic $T = T(\mathbf{Z}, \overline{\mathbf{A}}, \overline{\mathbf{Y}})$
- Find the distribution of T under H_0^Y : $\overline{\mathbf{Y}}(\mathbf{z}) = \overline{\mathbf{Y}}(\widetilde{\mathbf{z}})$ for all $\mathbf{z}, \widetilde{\mathbf{z}} \in \mathcal{Z}$

Randomization	Networks	Infections	Test Statistic
$\mathbf{z}_1$	$\overline{\mathbf{A}}(\mathbf{z}_1) = ?$	$\overline{\mathbf{Y}}(\mathbf{z}_1) = ?$	$T(\mathbf{z}_1, ?, ?)$
$\mathbf{z}_2$	$\overline{\mathbf{A}}(\mathbf{z}_2) = ?$	$\overline{\mathbf{Y}}(\mathbf{z}_1) = ?$	$T(\mathbf{z}_1, ?, ?)$
$\dots$	$\dots$	$\dots$	$\dots$
$\mathbf{z}_{\lvert\mathcal{Z}\rvert}$	$\overline{\mathbf{A}}(\mathbf{z}_{\lvert\mathcal{Z}\rvert}) = ?$	$\overline{\mathbf{Y}}(\mathbf{z}_{\lvert\mathcal{Z}\rvert}) = ?$	$T(\mathbf{z}_{\lvert\mathcal{Z}\rvert}, ?, ?)$

Testing H_0^Y

"How unlikely are the observed data under H_0^Y?"

- Choose a test statistic $T = T(\mathbf{Z}, \overline{\mathbf{A}}, \overline{\mathbf{Y}})$
- Find the distribution of T under $H_0^Y : \overline{\mathbf{Y}}(\mathbf{z}) = \overline{\mathbf{Y}}(\widetilde{\mathbf{z}})$ for all $\mathbf{z}, \widetilde{\mathbf{z}} \in \mathcal{Z}$

Randomization	**Networks**	**Infections**	**Test Statistic**
$\mathbf{z}_1$	$\overline{\mathbf{A}}(\mathbf{z}_1) = ?$	$\overline{\mathbf{Y}}(\mathbf{z}_1) = \overline{\mathbf{Y}}$	$T(\mathbf{z}_1, ?, \overline{\mathbf{Y}})$
$\mathbf{z}_2$	$\overline{\mathbf{A}}(\mathbf{z}_2) = ?$	$\overline{\mathbf{Y}}(\mathbf{z}_2) = \overline{\mathbf{Y}}$	$T(\mathbf{z}_2, ?, \overline{\mathbf{Y}})$
...	...	...	...
$\mathbf{z}_{\lvert\mathcal{Z}\rvert}$	$\overline{\mathbf{A}}(\mathbf{z}_{\lvert\mathcal{Z}\rvert}) = ?$	$\overline{\mathbf{Y}}(\mathbf{z}_{\lvert\mathcal{Z}\rvert}) = \overline{\mathbf{Y}}$	$T(\mathbf{z}_{\lvert\mathcal{Z}\rvert}, ?, \overline{\mathbf{Y}})$

Testing H_0^Y

"How unlikely are the observed data under H_0^Y?"

- Choose a test statistic $T = T(\mathbf{Z}, \overline{\mathbf{A}}, \overline{\mathbf{Y}})$
- Find the distribution of T under $H_0^Y : \overline{\mathbf{Y}}(\mathbf{z}) = \overline{\mathbf{Y}}(\widetilde{\mathbf{z}})$ for all $\mathbf{z}, \widetilde{\mathbf{z}} \in \mathcal{Z}$
  - This is not possible since $\overline{\mathbf{A}}(\mathbf{z})$ are **not imputable** under H_0^Y

Randomization	Networks	Infections	Test Statistic
$\mathbf{z}_1$	$\overline{\mathbf{A}}(\mathbf{z}_1) = ?$	$\overline{\mathbf{Y}}(\mathbf{z}_1) = \overline{\mathbf{Y}}$	$T(\mathbf{z}_1, ?, \overline{\mathbf{Y}})$
$\mathbf{z}_2$	$\overline{\mathbf{A}}(\mathbf{z}_2) = ?$	$\overline{\mathbf{Y}}(\mathbf{z}_2) = \overline{\mathbf{Y}}$	$T(\mathbf{z}_2, ?, \overline{\mathbf{Y}})$
$\ldots$	$\ldots$	$\ldots$	$\ldots$
$\mathbf{z}_{\lvert\mathcal{Z}\rvert}$	$\overline{\mathbf{A}}(\mathbf{z}_{\lvert\mathcal{Z}\rvert}) = ?$	$\overline{\mathbf{Y}}(\mathbf{z}_{\lvert\mathcal{Z}\rvert}) = \overline{\mathbf{Y}}$	$T(\mathbf{z}_{\lvert\mathcal{Z}\rvert}, ?, \overline{\mathbf{Y}})$

Testing H_0^Y

"How unlikely are the observed data under H_0^Y?"

- Choose a test statistic $T = T(\mathbf{Z}, \overline{\mathbf{A}}, \overline{\mathbf{Y}})$
- Find the distribution of T under $H_0^Y : \overline{\mathbf{Y}}(\mathbf{z}) = \overline{\mathbf{Y}}(\widetilde{\mathbf{z}})$ for all $\mathbf{z}, \widetilde{\mathbf{z}} \in \mathcal{Z}$
  - This is not possible since $\overline{\mathbf{A}}(\mathbf{z})$ are **not imputable** under $H_0^Y \implies H_0^Y$ is **nonsharp**

Randomization	Networks	Infections	Test Statistic
$\mathbf{z}_1$	$\overline{\mathbf{A}}(\mathbf{z}_1) = ?$	$\overline{\mathbf{Y}}(\mathbf{z}_1) = \overline{\mathbf{Y}}$	$T(\mathbf{z}_1, ?, \overline{\mathbf{Y}})$
$\mathbf{z}_2$	$\overline{\mathbf{A}}(\mathbf{z}_2) = ?$	$\overline{\mathbf{Y}}(\mathbf{z}_2) = \overline{\mathbf{Y}}$	$T(\mathbf{z}_2, ?, \overline{\mathbf{Y}})$
$\ldots$	$\ldots$	$\ldots$	$\ldots$
$\mathbf{z}_{\lvert\mathcal{Z}\rvert}$	$\overline{\mathbf{A}}(\mathbf{z}_{\lvert\mathcal{Z}\rvert}) = ?$	$\overline{\mathbf{Y}}(\mathbf{z}_{\lvert\mathcal{Z}\rvert}) = \overline{\mathbf{Y}}$	$T(\mathbf{z}_{\lvert\mathcal{Z}\rvert}, ?, \overline{\mathbf{Y}})$

Testing H_0^Y

- **Problem**: H_0^Y is nonsharp

Testing H_0^Y

- **Problem**: H_0^Y is nonsharp
  $\implies$ we don't know $F_T(t|H_0^Y)$ and thus cannot compute a p-value testing H_0^Y

Testing H_0^Y

- **Problem**: H_0^Y is nonsharp
  $\Longrightarrow$ we don't know $F_T(t|H_0^Y)$ and thus cannot compute a p-value testing H_0^Y
- **Solution**: model the distribution of potential network histories

Testing H_0^Y

- **Problem**: H_0^Y is nonsharp

  $\Longrightarrow$ we don't know $F_T(t|H_0^Y)$ and thus cannot compute a p-value testing H_0^Y
- **Solution**: model the distribution of potential network histories
  - Define $q(\bar{\mathbf{a}}, \mathbf{z}) = \Pr\{\overline{\mathbf{A}}(\mathbf{z}) = \bar{\mathbf{a}}\}$, the **probability mass function** (PMF) of $\overline{\mathbf{A}}(\mathbf{z})$

Testing H_0^Y

- **Problem**: H_0^Y is nonsharp

  $\implies$ we don't know $F_T(t|H_0^Y)$ and thus cannot compute a p-value testing H_0^Y
- **Solution**: model the distribution of potential network histories
  - Define $q(\overline{\mathbf{a}}, \mathbf{z}) = \Pr\{\overline{\mathbf{A}}(\mathbf{z}) = \overline{\mathbf{a}}\}$, the **probability mass function** (PMF) of $\overline{\mathbf{A}}(\mathbf{z})$
  - Then

$$F_T(t|H_0^Y) = \Pr\left(T \leq t|H_0^Y\right) = \sum_{\mathbf{z}\in\mathcal{Z}} \sum_{\overline{\mathbf{a}}\in\mathcal{A}} r(\mathbf{z})q(\overline{\mathbf{a}}, \mathbf{z})\mathbb{1}\{T(\mathbf{z}, \overline{\mathbf{a}}, \overline{\mathbf{Y}}) \leq t\}$$

Testing H_0^Y

- **Problem**: H_0^Y is nonsharp

  $\implies$ we don't know $F_T(t|H_0^Y)$ and thus cannot compute a p-value testing H_0^Y
- **Solution**: model the distribution of potential network histories
  - Define $q(\overline{\mathbf{a}}, \mathbf{z}) = \Pr\{\overline{\mathbf{A}}(\mathbf{z}) = \overline{\mathbf{a}}\}$, the **probability mass function** (PMF) of $\overline{\mathbf{A}}(\mathbf{z})$
  - Then

$$F_T(t|H_0^Y) = \Pr\left(T \leq t|H_0^Y\right) = \sum_{\mathbf{z}\in\mathcal{Z}} \sum_{\overline{\mathbf{a}}\in\mathcal{A}} r(\mathbf{z})q(\overline{\mathbf{a}}, \mathbf{z})\mathbb{1}\{T(\mathbf{z}, \overline{\mathbf{a}}, \overline{\mathbf{Y}}) \leq t\}$$

  - p-value: $\rho^Y = F_T(T|H_0^Y)$

Testing H_0^Y

- **Problem**: H_0^Y is nonsharp
  $\implies$ we don't know $F_T(t|H_0^Y)$ and thus cannot compute a p-value testing H_0^Y
- **Solution**: model the distribution of potential network histories
  - Define $q(\bar{\mathbf{a}}, \mathbf{z}) = \Pr\{\overline{\mathbf{A}}(\mathbf{z}) = \bar{\mathbf{a}}\}$, the **probability mass function** (PMF) of $\overline{\mathbf{A}}(\mathbf{z})$
  - Then

$$F_T(t|H_0^Y) = \Pr\left(T \leq t|H_0^Y\right) = \sum_{\mathbf{z} \in \mathcal{Z}} \sum_{\bar{\mathbf{a}} \in \mathcal{A}} r(\mathbf{z}) q(\bar{\mathbf{a}}, \mathbf{z}) \mathbb{1}\{T(\mathbf{z}, \bar{\mathbf{a}}, \overline{\mathbf{Y}}) \leq t\}$$

  - p-value: $\rho^Y = F_T(T|H_0^Y)$
  - If/when q is unknown, estimate it

Testing H_0^Y

- **Problem**: H_0^Y is nonsharp
  $\implies$ we don't know $F_T(t|H_0^Y)$ and thus cannot compute a p-value testing H_0^Y
- **Solution**: model the distribution of potential network histories
  - Define $q(\overline{\mathbf{a}}, \mathbf{z}) = \Pr\{\overline{\mathbf{A}}(\mathbf{z}) = \overline{\mathbf{a}}\}$, the **probability mass function** (PMF) of $\overline{\mathbf{A}}(\mathbf{z})$
  - Then
    $$F_T(t|H_0^Y) = \Pr\left(T \leq t|H_0^Y\right) = \sum_{\mathbf{z}\in\mathcal{Z}} \sum_{\overline{\mathbf{a}}\in\mathcal{A}} r(\mathbf{z}) q(\overline{\mathbf{a}}, \mathbf{z}) \mathbb{1}\{T(\mathbf{z}, \overline{\mathbf{a}}, \overline{\mathbf{Y}}) \leq t\}$$
  - p-value: $\rho^Y = F_T(T|H_0^Y)$
  - If/when q is unknown, estimate it
    - E.g., using a **Separable Temporal Exponential Random Graph Model**

Testing H_0^Y

- **Problem**: H_0^Y is nonsharp

  $\implies$ we don't know $F_T(t|H_0^Y)$ and thus cannot compute a p-value testing H_0^Y

- **Solution**: model the distribution of potential network histories
  - Define $q(\bar{\mathbf{a}}, \mathbf{z}) = \Pr\{\overline{\mathbf{A}}(\mathbf{z}) = \bar{\mathbf{a}}\}$, the **probability mass function** (PMF) of $\overline{\mathbf{A}}(\mathbf{z})$
  - Then

$$F_T(t|H_0^Y) = \Pr\left(T \leq t|H_0^Y\right) = \sum_{\mathbf{z}\in\mathcal{Z}} \sum_{\bar{\mathbf{a}}\in\mathcal{A}} r(\mathbf{z})q(\bar{\mathbf{a}}, \mathbf{z})\mathbb{1}\{T(\mathbf{z}, \bar{\mathbf{a}}, \overline{\mathbf{Y}}) \leq t\}$$

  - p-value: $\rho^Y = F_T(T|H_0^Y)$
  - If/when q is unknown, estimate it
    - E.g., using a **Separable Temporal Exponential Random Graph Model**
  - If/when the double sum $\sum_{\mathbf{z}\in\mathcal{Z}} \sum_{\bar{\mathbf{a}}\in\mathcal{A}}$ cannot be computed exactly, approximate it using a Monte-Carlo method

Testing H_0^Y

- **Problem**: H_0^Y is nonsharp
  $\implies$ we don't know $F_T(t|H_0^Y)$ and thus cannot compute a p-value testing H_0^Y
- **Solution**: model the distribution of potential network histories
  - Define $q(\overline{\mathbf{a}}, \mathbf{z}) = \Pr\{\overline{\mathbf{A}}(\mathbf{z}) = \overline{\mathbf{a}}\}$, the **probability mass function** (PMF) of $\overline{\mathbf{A}}(\mathbf{z})$
  - Then

$$F_T(t|H_0^Y) = \Pr\left(T \leq t|H_0^Y\right) = \sum_{\mathbf{z} \in \mathcal{Z}} \sum_{\overline{\mathbf{a}} \in \mathcal{A}} r(\mathbf{z}) q(\overline{\mathbf{a}}, \mathbf{z}) \mathbb{1}\{T(\mathbf{z}, \overline{\mathbf{a}}, \overline{\mathbf{Y}}) \leq t\}$$

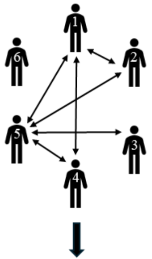
  - p-value: $\rho^Y = F_T(T|H_0^Y)$
  - If/when q is unknown, estimate it
    - E.g., using a **Separable Temporal Exponential Random Graph Model**
  - If/when the double sum $\sum_{\mathbf{z} \in \mathcal{Z}} \sum_{\overline{\mathbf{a}} \in \mathcal{A}}$ cannot be computed exactly, approximate it using a Monte-Carlo method
- **Proposition**: This procedure will **asymptotically** control the type I error rate

Exponential Family Random Graph Models (ERGMs)

Goal: model a network ($\mathbf{A}$) given covariates ($\mathbf{X}$)

Exponential Family Random Graph Models (ERGMs)

Goal: model a network ($\mathbf{A}$) given covariates ($\mathbf{X}$)



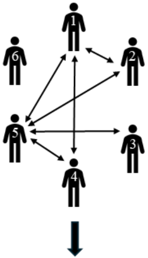
	1	2	3	4	5	6
1	0	1	0	1	1	0
2	1	0	0	0	1	0
3	0	0	0	0	1	0
4	1	0	0	0	1	0
5	1	1	1	1	0	0
6	0	0	0	0	0	0

Exponential Family Random Graph Models (ERGMs)

Goal: model a network (**A**) given covariates (**X**)

An **ERGM** assumes:

$$\Pr_\theta(\mathbf{A} = \mathbf{a} | \mathbf{X} = \mathbf{x}) = \frac{\exp\{\theta \mathbf{g}(\mathbf{a}, \mathbf{x})\}}{\kappa_{\mathbf{g}}(\theta, \mathcal{A}, \mathbf{x})}$$



	1	2	3	4	5	6
1	0	1	0	1	1	0
2	1	0	0	0	1	0
3	0	0	0	0	1	0
4	1	0	0	0	1	0
5	1	1	1	1	0	0
6	0	0	0	0	0	0

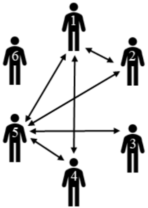
Exponential Family Random Graph Models (ERGMs)

Goal: model a network ($\mathbf{A}$) given covariates ($\mathbf{X}$)

An **ERGM** assumes:

$$\Pr_\theta(\mathbf{A} = \mathbf{a} | \mathbf{X} = \mathbf{x}) = \frac{\exp\{\theta \mathbf{g}(\mathbf{a}, \mathbf{x})\}}{\kappa_{\mathbf{g}}(\theta, \mathcal{A}, \mathbf{x})}$$

- normalizing constant: $\kappa_{\mathbf{g}}(\theta, \mathcal{A}, \mathbf{x}) = \sum_{\mathbf{a} \in \mathcal{A}} \exp\{\theta \mathbf{g}(\mathbf{a}, \mathbf{x})\}$



	1	2	3	4	5	6
1	0	1	0	1	1	0
2	1	0	0	0	1	0
3	0	0	0	0	1	0
4	1	0	0	0	1	0
5	1	1	1	1	0	0
6	0	0	0	0	0	0

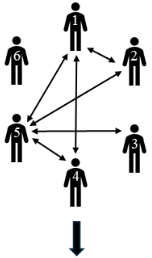
Exponential Family Random Graph Models (ERGMs)

Goal: model a network ($\mathbf{A}$) given covariates ($\mathbf{X}$)

An **ERGM** assumes:

$$\Pr_\theta(\mathbf{A} = \mathbf{a} | \mathbf{X} = \mathbf{x}) = \frac{\exp\{\theta \mathbf{g}(\mathbf{a}, \mathbf{x})\}}{\kappa_{\mathbf{g}}(\theta, \mathcal{A}, \mathbf{x})}$$

- normalizing constant: $\kappa_{\mathbf{g}}(\theta, \mathcal{A}, \mathbf{x}) = \sum_{\mathbf{a} \in \mathcal{A}} \exp\{\theta \mathbf{g}(\mathbf{a}, \mathbf{x})\}$
- $\mathcal{A}$: space of possible networks



	1	2	3	4	5	6
1	0	1	0	1	1	0
2	1	0	0	0	1	0
3	0	0	0	0	1	0
4	1	0	0	0	1	0
5	1	1	1	1	0	0
6	0	0	0	0	0	0

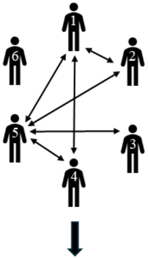
Exponential Family Random Graph Models (ERGMs)

Goal: model a network (**A**) given covariates (**X**)

An **ERGM** assumes:

$$\Pr_\theta(\mathbf{A} = \mathbf{a} | \mathbf{X} = \mathbf{x}) = \frac{\exp\{\theta \mathbf{g}(\mathbf{a}, \mathbf{x})\}}{\kappa_{\mathbf{g}}(\theta, \mathcal{A}, \mathbf{x})}$$

- normalizing constant: $\kappa_{\mathbf{g}}(\theta, \mathcal{A}, \mathbf{x}) = \sum_{\mathbf{a} \in \mathcal{A}} \exp\{\theta \mathbf{g}(\mathbf{a}, \mathbf{x})\}$
- $\mathcal{A}$: space of possible networks
- θ: model coefficients



	1	2	3	4	5	6
1	0	1	0	1	1	0
2	1	0	0	0	1	0
3	0	0	0	0	1	0
4	1	0	0	0	1	0
5	1	1	1	1	0	0
6	0	0	0	0	0	0

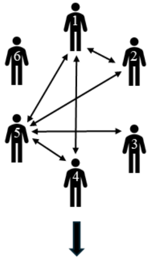
Exponential Family Random Graph Models (ERGMs)

Goal: model a network (**A**) given covariates (**X**)

An **ERGM** assumes:

$$\Pr_\theta(\mathbf{A} = \mathbf{a} | \mathbf{X} = \mathbf{x}) = \frac{\exp\{\theta \mathbf{g}(\mathbf{a}, \mathbf{x})\}}{\kappa_{\mathbf{g}}(\theta, \mathcal{A}, \mathbf{x})}$$

- normalizing constant: $\kappa_{\mathbf{g}}(\theta, \mathcal{A}, \mathbf{x}) = \sum_{\mathbf{a} \in \mathcal{A}} \exp\{\theta \mathbf{g}(\mathbf{a}, \mathbf{x})\}$
- $\mathcal{A}$: space of possible networks
- θ: model coefficients
- $\mathbf{g}(\mathbf{a}, \mathbf{x})$: sufficient statistics



	1	2	3	4	5	6
1	0	1	0	1	1	0
2	1	0	0	0	1	0
3	0	0	0	0	1	0
4	1	0	0	0	1	0
5	1	1	1	1	0	0
6	0	0	0	0	0	0

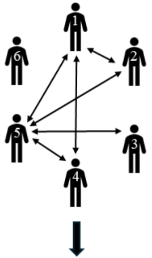
Exponential Family Random Graph Models (ERGMs)

Goal: model a network (**A**) given covariates (**X**)

An **ERGM** assumes:

$$\Pr_{\theta}(\mathbf{A} = \mathbf{a} | \mathbf{X} = \mathbf{x}) = \frac{\exp\{\theta \mathbf{g}(\mathbf{a}, \mathbf{x})\}}{\kappa_{\mathbf{g}}(\theta, \mathcal{A}, \mathbf{x})}$$

- normalizing constant: $\kappa_{\mathbf{g}}(\theta, \mathcal{A}, \mathbf{x}) = \sum_{\mathbf{a} \in \mathcal{A}} \exp\{\theta \mathbf{g}(\mathbf{a}, \mathbf{x})\}$
- $\mathcal{A}$: space of possible networks
- θ: model coefficients
- $\mathbf{g}(\mathbf{a}, \mathbf{x})$: sufficient statistics, for example:
  - $g_{\text{edges}}(\mathbf{a}, \mathbf{x}) = \sum_{i,j} a_{ij}$



	1	2	3	4	5	6
1	0	1	0	1	1	0
2	1	0	0	0	1	0
3	0	0	0	0	1	0
4	1	0	0	0	1	0
5	1	1	1	1	0	0
6	0	0	0	0	0	0

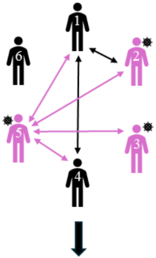
Exponential Family Random Graph Models (ERGMs)

Goal: model a network ($\mathbf{A}$) given covariates ($\mathbf{X}$)

An **ERGM** assumes:

$$\Pr_\theta(\mathbf{A} = \mathbf{a} | \mathbf{X} = \mathbf{x}) = \frac{\exp\{\theta \mathbf{g}(\mathbf{a}, \mathbf{x})\}}{\kappa_{\mathbf{g}}(\theta, \mathcal{A}, \mathbf{x})}$$

- normalizing constant:
  $\kappa_{\mathbf{g}}(\theta, \mathcal{A}, \mathbf{x}) = \sum_{\mathbf{a} \in \mathcal{A}} \exp\{\theta \mathbf{g}(\mathbf{a}, \mathbf{x})\}$
- $\mathcal{A}$: space of possible networks
- θ: model coefficients
- $\mathbf{g}(\mathbf{a}, \mathbf{x})$: sufficient statistics, for example:
  - $g_{\text{edges}}(\mathbf{a}, \mathbf{x}) = \sum_{i,j} a_{ij}$
  - $g_Z(\mathbf{a}, \mathbf{x}) = \sum_{i,j} a_{ij}(z_i + z_j)$



	1	2	3	4	5	6
1	0	1	0	1	1	0
2	1	0	0	0	1	0
3	0	0	0	0	1	0
4	1	0	0	0	1	0
5	1	1	1	1	0	0
6	0	0	0	0	0	0

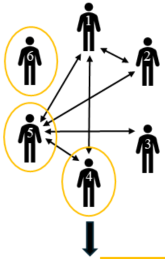
Exponential Family Random Graph Models (ERGMs)

Goal: model a network (**A**) given covariates (**X**)

An **ERGM** assumes:

$$\Pr_\theta(\mathbf{A} = \mathbf{a} | \mathbf{X} = \mathbf{x}) = \frac{\exp\{\theta \mathbf{g}(\mathbf{a}, \mathbf{x})\}}{\kappa_{\mathbf{g}}(\theta, \mathcal{A}, \mathbf{x})}$$

- normalizing constant:
  $\kappa_{\mathbf{g}}(\theta, \mathcal{A}, \mathbf{x}) = \sum_{\mathbf{a} \in \mathcal{A}} \exp\{\theta \mathbf{g}(\mathbf{a}, \mathbf{x})\}$
- $\mathcal{A}$: space of possible networks
- θ: model coefficients
- $\mathbf{g}(\mathbf{a}, \mathbf{x})$: sufficient statistics, for example:
  - $g_{\text{edges}}(\mathbf{a}, \mathbf{x}) = \sum_{i,j} a_{ij}$
  - $g_Z(\mathbf{a}, \mathbf{x}) = \sum_{i,j} a_{ij}(z_i + z_j)$
  - $g_Y(\mathbf{a}, \mathbf{x}) = \sum_{i,j} a_{ij}(y_i + y_j)$



	1	2	3	4	5	6
1	0	1	0	1	1	0
2	1	0	0	0	1	0
3	0	0	0	0	1	0
4	1	0	0	0	1	0
5	1	1	1	1	0	0
6	0	0	0	0	0	0

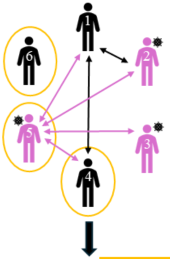
Exponential Family Random Graph Models (ERGMs)

Goal: model a network ($\mathbf{A}$) given covariates ($\mathbf{X}$)

An **ERGM** assumes:

$$\Pr_\theta(\mathbf{A}=\mathbf{a}|\mathbf{X}=\mathbf{x}) = \frac{\exp\{\theta \mathbf{g}(\mathbf{a},\mathbf{x})\}}{\kappa_{\mathbf{g}}(\theta,\mathcal{A},\mathbf{x})}$$

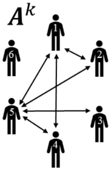
- normalizing constant:
  $\kappa_{\mathbf{g}}(\theta,\mathcal{A},\mathbf{x}) = \sum_{\mathbf{a}\in\mathcal{A}} \exp\{\theta \mathbf{g}(\mathbf{a},\mathbf{x})\}$
- $\mathcal{A}$: space of possible networks
- θ: model coefficients
- $\mathbf{g}(\mathbf{a},\mathbf{x})$: sufficient statistics, for example:
  - $g_{\text{edges}}(\mathbf{a},\mathbf{x}) = \sum_{i,j} a_{ij}$
  - $g_Z(\mathbf{a},\mathbf{x}) = \sum_{i,j} a_{ij}(z_i + z_j)$
  - $g_Y(\mathbf{a},\mathbf{x}) = \sum_{i,j} a_{ij}(y_i + y_j)$
  - $g_{ZY}(\mathbf{a},\mathbf{x}) = \sum_{i,j} a_{ij}(z_i y_i + z_j y_j)$



	1	2	3	4	5	6
1	0	1	0	1	1	0
2	1	0	0	0	1	0
3	0	0	0	0	1	0
4	1	0	0	0	1	0
5	1	1	1	1	0	0
6	0	0	0	0	0	0

Separable Temporal Exponential Family Random Graph Models

A **STERGM** assumes that, at each time step $k = 1, \ldots, \tau - 1$:

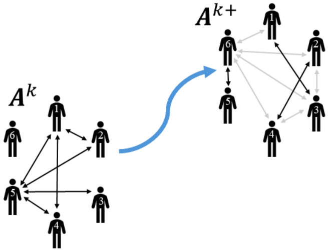


Separable Temporal Exponential Family Random Graph Models

A **STERGM** assumes that, at each time step $k = 1, \ldots, \tau - 1$:

- New edges form according to a **formation** ERGM

$$\Pr_{\theta^+}(\mathbf{A}^{k+} = \mathbf{a}^{k+} | \mathbf{A}^k = \mathbf{a}^k, \mathbf{X}^{k+1} = \mathbf{x}^{k+1}) = \frac{\exp\left(\theta^+ \mathbf{g}^+(\mathbf{a}^{k+}, \mathbf{a}^k, \mathbf{x}^{k+1})\right)}{\kappa_{\mathbf{g}^+}\left(\theta^+, \mathcal{A}^+(\mathbf{a}^k), \mathbf{a}^k, \mathbf{x}^{k+1}\right)},$$

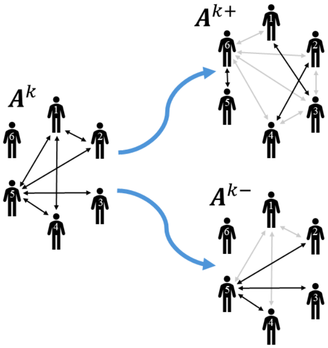


Separable Temporal Exponential Family Random Graph Models

A **STERGM** assumes that, at each time step $k = 1, \ldots, \tau - 1$:

- New edges form according to a **formation** ERGM
- Old edges persist according to a **persistence** ERGM

$$\Pr_{\theta^-}(\mathbf{A}^{k-} = \mathbf{a}^{k-} | \mathbf{A}^k = \mathbf{a}^k, \mathbf{X}^{k+1} = \mathbf{x}^{k+1}) = \frac{\exp\left(\theta^- \mathbf{g}^-(\mathbf{a}^{k-}, \mathbf{a}^k, \mathbf{x}^{k+1})\right)}{\kappa_{\mathbf{g}^-}(\theta^-, \mathcal{A}^-(\mathbf{a}^k), \mathbf{a}^k, \mathbf{x}^{k+1})},$$

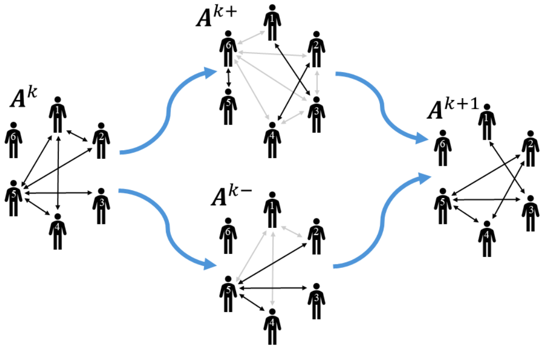


Separable Temporal Exponential Family Random Graph Models

A **STERGM** assumes that, at each time step $k = 1, \ldots, \tau - 1$:

- New edges form according to a **formation** ERGM
- Old edges persist according to a **persistence** ERGM

$$\mathbf{A}^{k+1} = \underbrace{\mathbf{A}^{k}}_{\text{previous network}} \cup \underbrace{(\mathbf{A}^{k+} - \mathbf{A}^{k})}_{\text{new edges formed}} - \underbrace{(\mathbf{A}^{k} - \mathbf{A}^{k-})}_{\text{old edges not persisting}}$$



Simulation Study

Goal: use simulated data to investigate the performance of the testing procedure

Simulation Study

Goal: use simulated data to investigate the performance of the testing procedure

- $n \in \{24, 48\}$ students
- $\tau \in \{5, 10\}$ weeks
- 50:50 cluster randomization with 12 equal-sized clusters

Simulation Study

Goal: use simulated data to investigate the performance of the testing procedure

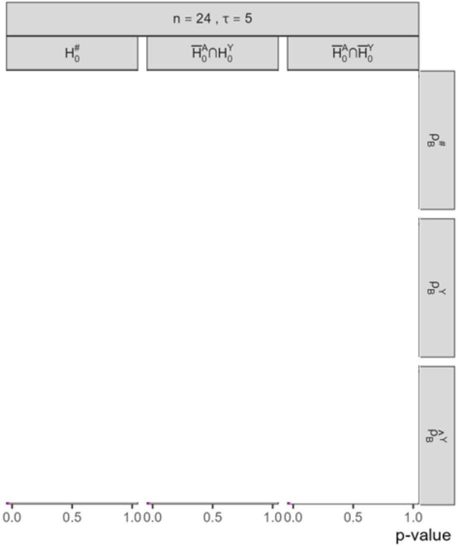
- $n \in \{24, 48\}$ students
- $\tau \in \{5, 10\}$ weeks
- 50:50 cluster randomization with 12 equal-sized clusters
- Three scenarios:
  1. $H_0^\sharp$: no effect of intervention on networks or infection
  2. $\overline{H}_0^A \cap H_0^Y$: no effect of intervention on infection
  3. $\overline{H}_0^A \cap \overline{H}_0^Y$: effect of intervention on both networks and infection

Simulation Study

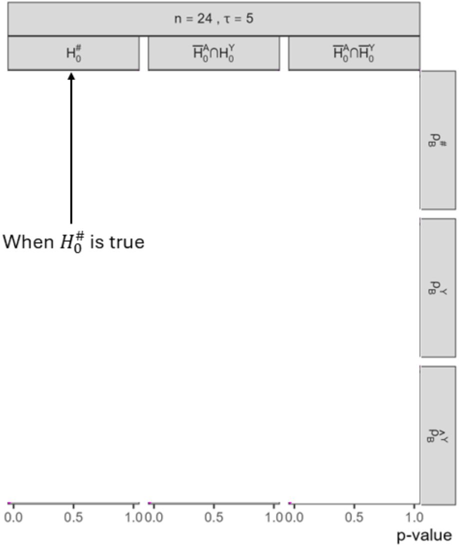
Goal: use simulated data to investigate the performance of the testing procedure

- $n \in \{24, 48\}$ students
- $\tau \in \{5, 10\}$ weeks
- 50:50 cluster randomization with 12 equal-sized clusters
- Three scenarios:
  1. $H_0^\sharp$: no effect of intervention on networks or infection
  2. $\overline{H}_0^A \cap H_0^Y$: no effect of intervention on infection
  3. $\overline{H}_0^A \cap \overline{H}_0^Y$: effect of intervention on both networks and infection
- Three testing procedures:
  1. $\rho_B^\sharp$: testing $H_0^\sharp$
  2. ρ_B^Y: testing H_0^Y using known q
  3. $\widehat{\rho}_B^Y$: testing H_0^Y using estimated q

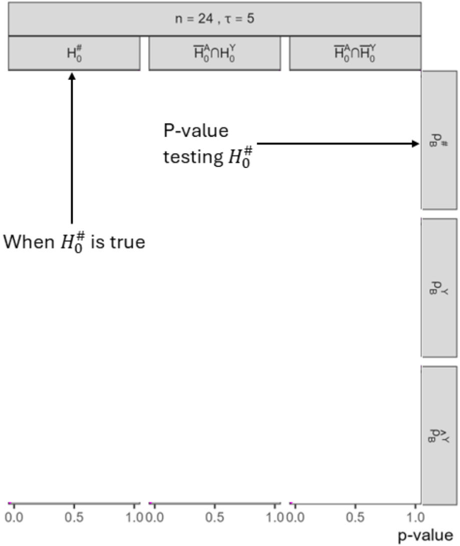
Simulation Results: Empirical CDF of P-Values



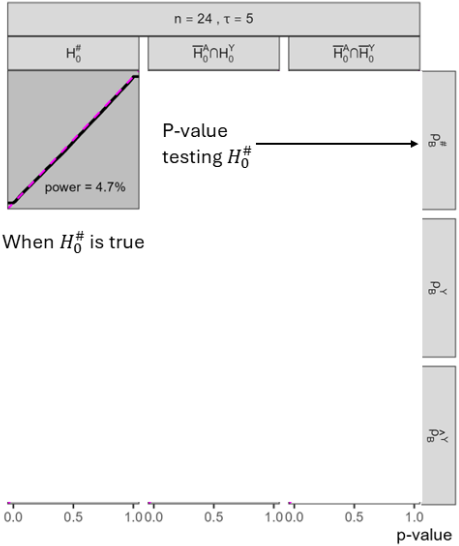
Simulation Results: Empirical CDF of P-Values



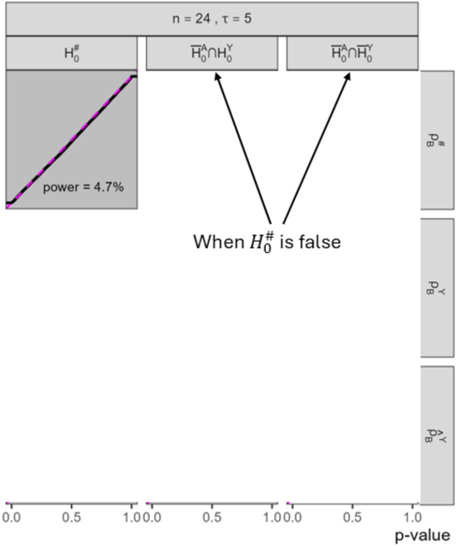
Simulation Results: Empirical CDF of P-Values



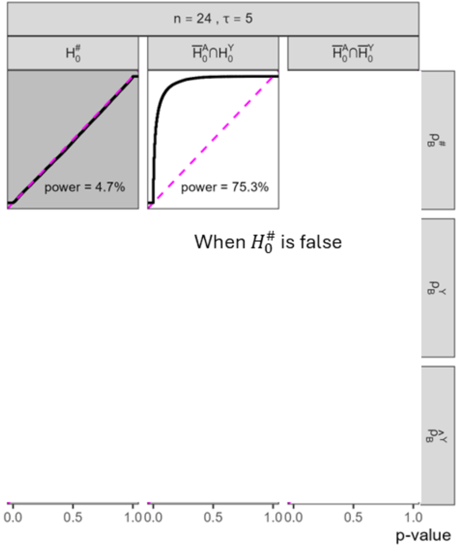
Simulation Results: Empirical CDF of P-Values



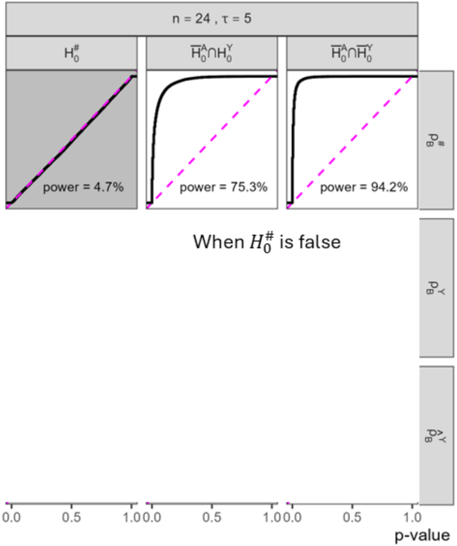
Simulation Results: Empirical CDF of P-Values



Simulation Results: Empirical CDF of P-Values

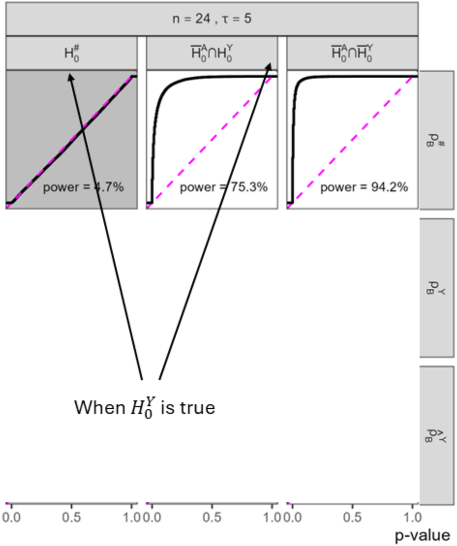


Simulation Results: Empirical CDF of P-Values

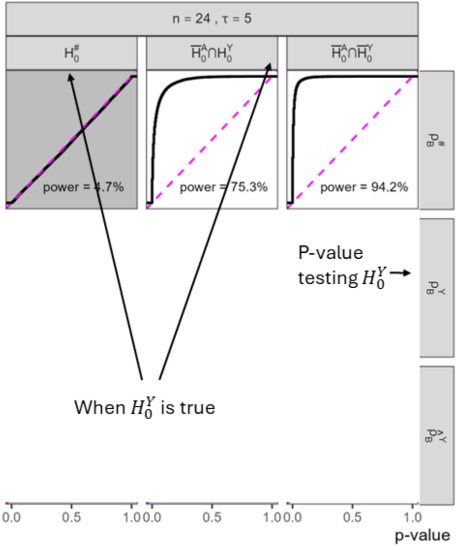


When $H_0^{\#}$ is false

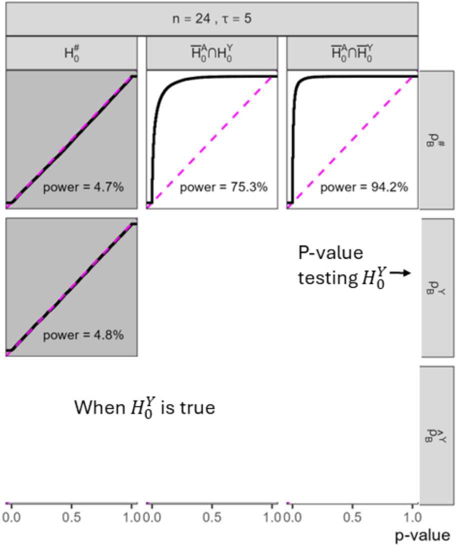
Simulation Results: Empirical CDF of P-Values



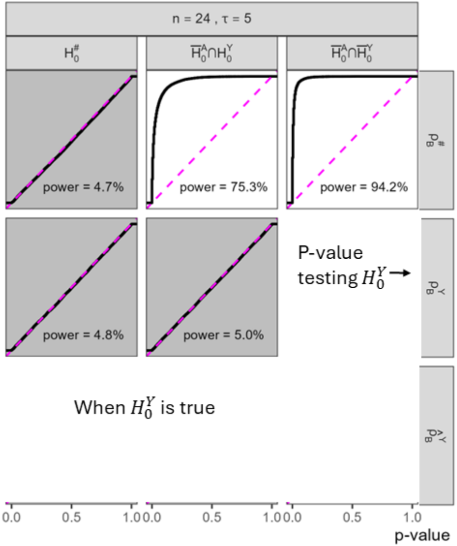
Simulation Results: Empirical CDF of P-Values



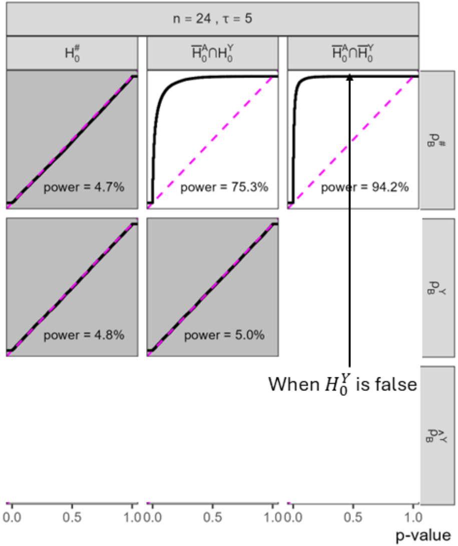
Simulation Results: Empirical CDF of P-Values



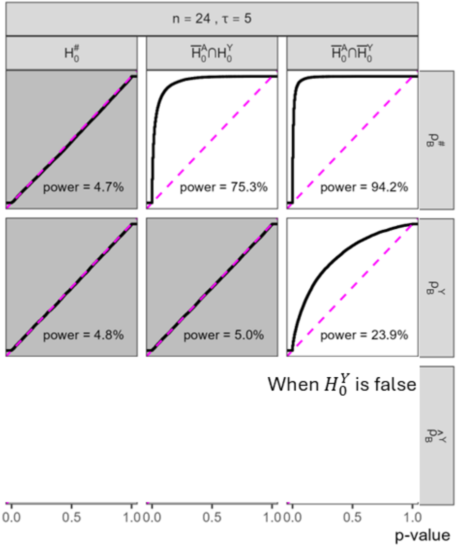
Simulation Results: Empirical CDF of P-Values



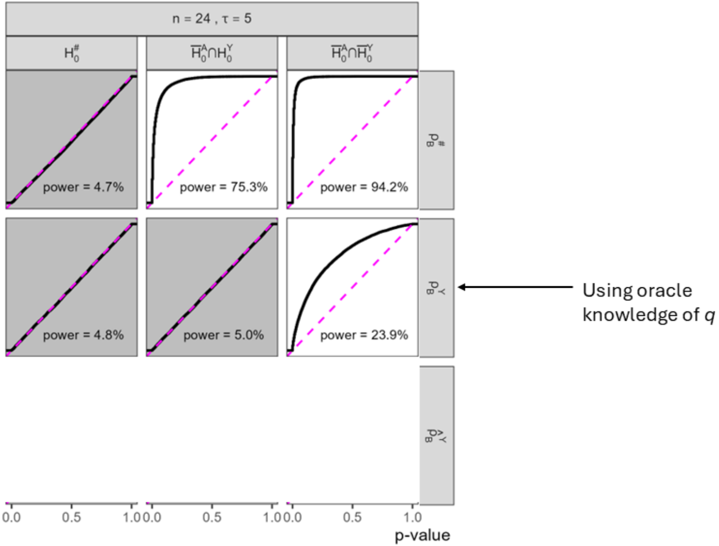
Simulation Results: Empirical CDF of P-Values



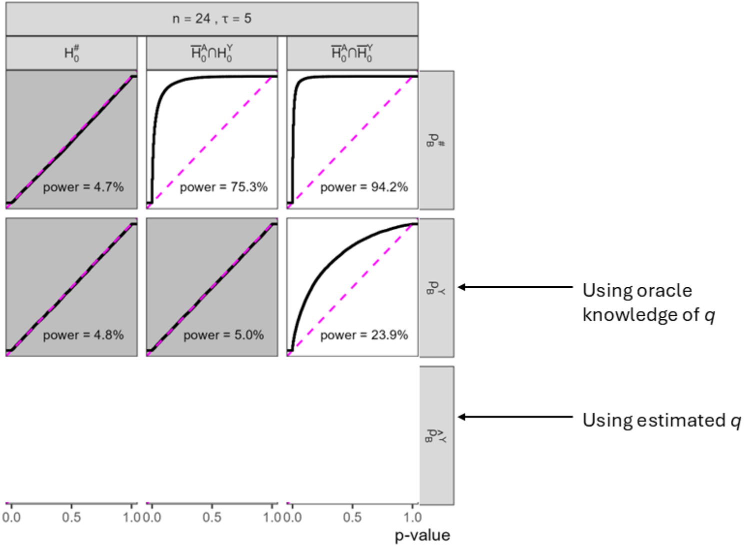
Simulation Results: Empirical CDF of P-Values



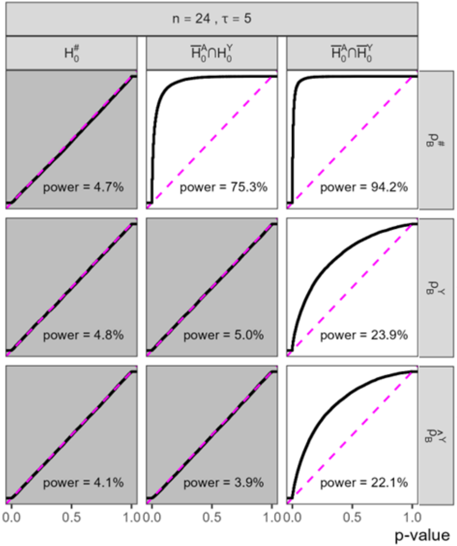
Simulation Results: Empirical CDF of P-Values



Simulation Results: Empirical CDF of P-Values



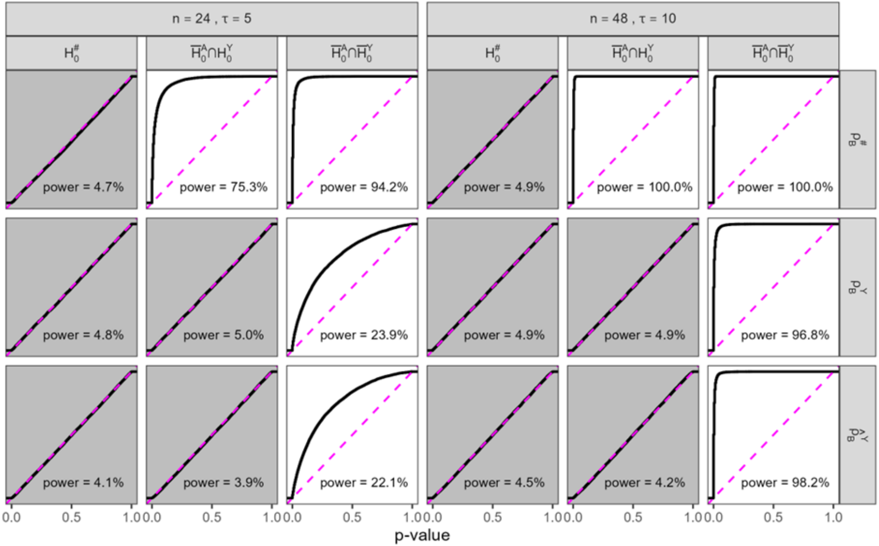
Simulation Results: Empirical CDF of P-Values



Using oracle knowledge of q

Using estimated q

Simulation Results: Empirical CDF of P-Values



Separability of STERGMs

1. $\mathbf{A}^{k+} \perp\!\!\!\perp \mathbf{A}^{k-} | \mathbf{A}^{k}$, i.e., the formation and persistence processes are conditionally independent give the network at time k

Separability of STERGMs

1. $\mathbf{A}^{k+} \perp\!\!\!\perp \mathbf{A}^{k-} | \mathbf{A}^{k}$, i.e., the formation and persistence processes are conditionally independent give the network at time k

$$\begin{tikzcd}
 & \mathbf{A}^{1+} \arrow[dr] & & \mathbf{A}^{2+} \arrow[dr] & & \mathbf{A}^{3+} \arrow[dr] & \\
\mathbf{A}^{1} \arrow[ur] \arrow[rr] \arrow[dr] & & \mathbf{A}^{2} \arrow[ur] \arrow[rr] \arrow[dr] & & \mathbf{A}^{3} \arrow[ur] \arrow[rr] \arrow[dr] & & \mathbf{A}^{4} \\
 & \mathbf{A}^{1-} \arrow[ur] & & \mathbf{A}^{2-} \arrow[ur] & & \mathbf{A}^{3-} \arrow[ur] &
\end{tikzcd}$$

Separability of STERGMs

1. $\mathbf{A}^{k+} \perp\!\!\!\perp \mathbf{A}^{k-} | \mathbf{A}^{k}$, i.e., the formation and persistence processes are conditionally independent give the network at time k
2. the parameter space for $\theta = (\theta^{+}, \theta^{-})$ is the product of the parameter spaces for θ^{+} and θ^{-}

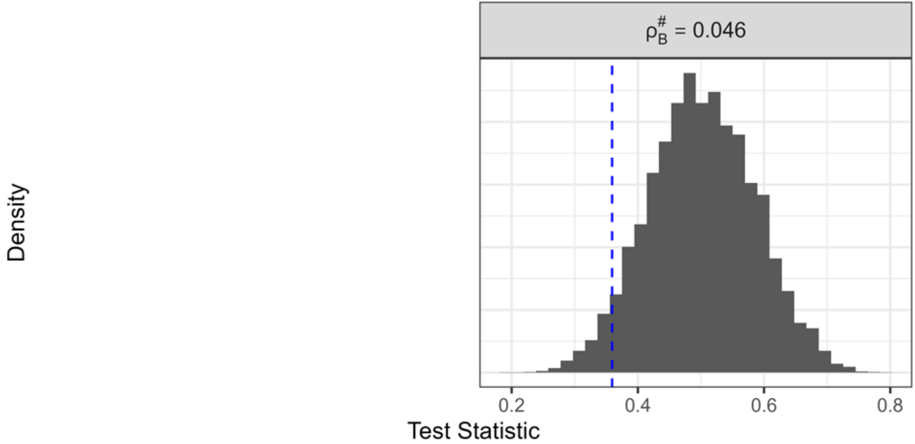
$$
\begin{tikzcd}
 & \mathbf{A}^{1+} \arrow[dr] & & \mathbf{A}^{2+} \arrow[dr] & & \mathbf{A}^{3+} \arrow[dr] & \\
\mathbf{A}^{1} \arrow[ur] \arrow[rr] \arrow[dr] & & \mathbf{A}^{2} \arrow[ur] \arrow[rr] \arrow[dr] & & \mathbf{A}^{3} \arrow[ur] \arrow[rr] \arrow[dr] & & \mathbf{A}^{4} \\
 & \mathbf{A}^{1-} \arrow[ur] & & \mathbf{A}^{2-} \arrow[ur] & & \mathbf{A}^{3-} \arrow[ur] &
\end{tikzcd}
$$

eX-FLU: Hypothesis Test Results

- Test statistic (proportion of possible transmission events attributable to students in the intervention group): $T = 0.359$.

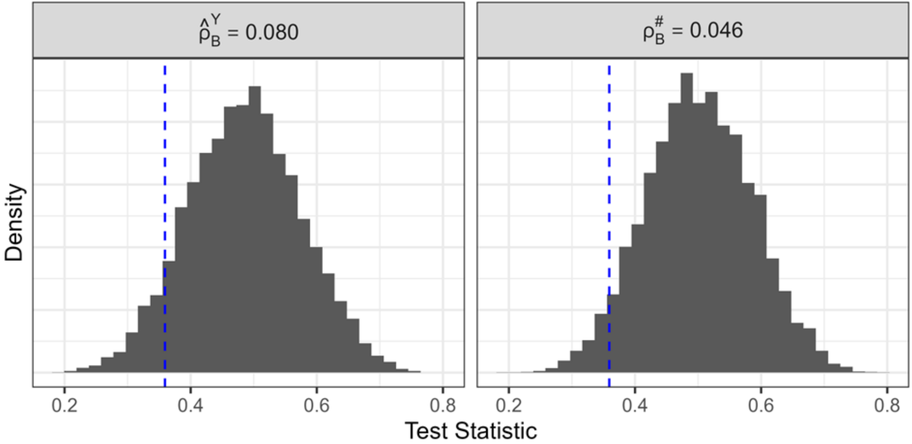
eX-FLU: Hypothesis Test Results

- Test statistic (proportion of possible transmission events attributable to students in the intervention group): $T = 0.359$.



eX-FLU: Hypothesis Test Results

- Test statistic (proportion of possible transmission events attributable to students in the intervention group): $T = 0.359$.



Discussion

Summary:

- Novel testing procedure for testing H_0^Y
- Theoretical and empirical demonstration of type I error control
- Moderate evidence against H_0^Y in the eX-FLU trial

Discussion

Summary:

- Novel testing procedure for testing H_0^Y
- Theoretical and empirical demonstration of type I error control
- Moderate evidence against H_0^Y in the eX-FLU trial

Not included in this talk:

- What diagnostics can we check?
- How sensitive is the procedure to STERGM specification?
- What else can we learn from the STERGM?

Discussion

Summary:

- Novel testing procedure for testing H_0^Y
- Theoretical and empirical demonstration of type I error control
- Moderate evidence against H_0^Y in the eX-FLU trial

Not included in this talk:

- What diagnostics can we check?
- How sensitive is the procedure to STERGM specification?
- What else can we learn from the STERGM?

Future directions:

- Account for measurement error in self-reported networks
- Estimate causal parameter(s)

Discussion

Summary:

- Novel testing procedure for testing H_0^Y
- Theoretical and empirical demonstration of type I error control
- Moderate evidence against H_0^Y in the eX-FLU trial

Not included in this talk:

- What diagnostics can we check?
- How sensitive is the procedure to STERGM specification?
- What else can we learn from the STERGM?

Future directions:

- Account for measurement error in self-reported networks
- Estimate causal parameter(s)

Project status:

- Under revision with *PNAS*

Questions?

contact: richabr@wfu.edu

References I

Allison E. Aiello, Amanda M. Simanek, Marisa C. Eisenberg, Alison R. Walsh, Brian Davis, Erik Volz, Caroline Cheng, Jeanette J. Rainey, Amra Uzicanin, Hongjiang Gao, Nathaniel Osgood, Dylan Knowles, Kevin Stanley, Kara Tarter, and Arnold S. Monto. Design and methods of a social network isolation study for reducing respiratory infection transmission: The eX-FLU cluster randomized trial. *Epidemics*, 15:38–55, June 2016. ISSN 17554365. doi: 10.1016/j.epidem.2016.01.001. URL https://linkinghub.elsevier.com/retrieve/pii/S1755436516000025.

Shaina J. Alexandria, Michael G. Hudgens, and Allison E. Aiello. Assessing Intervention Effects in a Randomized Trial Within a Social Network. *Biometrics*, 79(2):1409–1419, June 2023. ISSN 0006-341X, 1541-0420. doi: 10.1111/biom.13606. URL https://academic.oup.com/biometrics/article/79/2/1409-1419/7513977.

R. A. Fisher. The design of experiments. *Nature*, 137(3459):252–254, February 1936. ISSN 0028-0836, 1476-4687. doi: 10.1038/137252a0.

Michael G Hudgens and M. Elizabeth Halloran. Toward causal inference with interference. *Journal of the American Statistical Association*, 103(482):832–842, 2008. ISSN 0162-1459, 1537-274X. doi: 10.1198/016214508000000292.

David M. Ritzwoller, Joseph P. Romano, and Azeem M. Shaikh. Randomization inference: Theory and applications, 2024.

Yao Zhang and Qingyuan Zhao. What is a randomization test? *Journal of the American Statistical Association*, 118(544):2928–2942, October 2023. ISSN 0162-1459, 1537-274X. doi: 10.1080/01621459.2023.2199814.

P.N. Zivich, M.C. Eisenberg, A.S. Monto, A. Uzicanin, R. S. Baric, T. P. Sheahan, J. J. Rainey, H. Gao, and A. E. Aiello. Transmission of viral pathogens in a social network of university students: the eX-FLU study. *Epidemiology and Infection*, 148:e267, 2020. ISSN 0950-2688, 1469-4409. doi: 10.1017/S0950268820001806. URL https://www.cambridge.org/core/product/identifier/S0950268820001806/type/journal_article.

Supplementary Slides

Choice of Test Statistic

Test statistics: "proportion of possible transmission events attributable to students in the intervention group:"

$$T^{**}(\mathbf{Z}, \overline{\mathbf{A}}, \overline{\mathbf{Y}}) = \frac{\sum_{k=2}^{\tau} \sum_{i=1}^{n} \sum_{j \neq i} Z_i E_{ijk}^{**}}{\sum_{k=2}^{\tau} \sum_{i=1}^{n} \sum_{j \neq i} E_{ijk}^{**}}.$$

	From Infected to Infected	**From Infected**
Contact at $k-1$	$E_{ijk}^{11} = Y_i^{k-1} A_{ij}^{k-1} Y_j^k$	$E_{ijk}^{12} = Y_i^{k-1} A_{ij}^{k-1}$
Contact at k	$E_{ijk}^{21} = Y_i^{k-1} A_{ij}^{k} Y_j^k$	$E_{ijk}^{22} = Y_i^{k-1} A_{ij}^{k}$
Contact at $k-1$ or k	$E_{ijk}^{31} = Y_i^{k-1} (A_{ij}^{k-1} \vee A_{ij}^{k}) Y_j^k$	$E_{ijk}^{32} = Y_i^{k-1} (A_{ij}^{k-1} \vee A_{ij}^{k})$
Contact at $k-1$ and k	$E_{ijk}^{41} = Y_i^{k-1} (A_{ij}^{k-1} A_{ij}^{k}) Y_j^k$	$E_{ijk}^{42} = Y_i^{k-1} (A_{ij}^{k-1} A_{ij}^{k})$

Table: Definitions of a possible transmission event E_{ijk} from student i to student j at time k. Y_i^k is an indicator for student i being infected at week k, A_{ij}^k is an indicator for students i and j being in contact at week k, and $a \vee b$ denotes the maximum of a and b.

Testing H_0^Y

Testing Procedure:

1. Estimate the unknown parameter θ in the PMF $q(\bar{\mathbf{a}}, \mathbf{z}; \theta)$ of $\overline{\mathbf{A}}(\mathbf{z})$

Testing H_0^Y

Testing Procedure:

1. Estimate the unknown parameter θ in the PMF $q(\bar{\mathbf{a}}, \mathbf{z}; \theta)$ of $\overline{\mathbf{A}}(\mathbf{z})$
2. Approximate $F_T(t|H_0^Y)$ using B Monte-Carlo replicates:
   for $b = 1$ to B **do:**

Testing H_0^Y

Testing Procedure:

1. Estimate the unknown parameter θ in the PMF $q(\bar{\mathbf{a}}, \mathbf{z}; \theta)$ of $\overline{\mathbf{A}}(\mathbf{z})$
2. Approximate $F_T(t|H_0^Y)$ using B Monte-Carlo replicates:
   for $b = 1$ to B **do:**
   1. generate randomization vector $\mathbf{z}_b$ from $r(\mathbf{z})$

Testing H_0^Y

Testing Procedure:

1. Estimate the unknown parameter θ in the PMF $q(\overline{\mathbf{a}}, \mathbf{z}; \theta)$ of $\overline{\mathbf{A}}(\mathbf{z})$
2. Approximate $F_T(t|H_0^Y)$ using B Monte-Carlo replicates:
   for $b = 1$ to B **do:**
   1. generate randomization vector $\mathbf{z}_b$ from $r(\mathbf{z})$
   2. generate networks $\overline{\mathbf{a}}_b$ from $q(\overline{\mathbf{a}}, \mathbf{z}_b; \widehat{\theta})$

Testing H_0^Y

Testing Procedure:

1. Estimate the unknown parameter θ in the PMF $q(\bar{\mathbf{a}}, \mathbf{z}; \theta)$ of $\overline{\mathbf{A}}(\mathbf{z})$
2. Approximate $F_T(t|H_0^Y)$ using B Monte-Carlo replicates:
   for $b = 1$ to B **do:**
   1. generate randomization vector $\mathbf{z}_b$ from $r(\mathbf{z})$
   2. generate networks $\bar{\mathbf{a}}_b$ from $q(\bar{\mathbf{a}}, \mathbf{z}_b; \widehat{\theta})$
   3. compute the test statistic $T_b = T(\mathbf{z}_b, \bar{\mathbf{a}}_b, \mathbf{Y})$

Testing H_0^Y

Testing Procedure:

1. Estimate the unknown parameter θ in the PMF $q(\bar{\mathbf{a}}, \mathbf{z}; \theta)$ of $\overline{\mathbf{A}}(\mathbf{z})$
2. Approximate $F_T(t|H_0^Y)$ using B Monte-Carlo replicates:
   for $b = 1$ to B **do:**
   1. generate randomization vector $\mathbf{z}_b$ from $r(\mathbf{z})$
   2. generate networks $\bar{\mathbf{a}}_b$ from $q(\bar{\mathbf{a}}, \mathbf{z}_b; \widehat{\theta})$
   3. compute the test statistic $T_b = T(\mathbf{z}_b, \bar{\mathbf{a}}_b, \mathbf{Y})$
3. compute the plug-in p-value testing H_0^Y:

$$\widehat{\rho}_B^Y = B^{-1} \sum_{b=1}^{B} \mathbb{1}(T_b \leq T)$$

Type I Error Control

- **Proposition 2.1**: Let T_N be a test statistic with CDF F_N under H_0.
  1. Under H_0, $F_N(T_N)$ stochastically dominates a Uniform$(0,1)$ distribution for any N.
  2. If the test statistic T_N has a continuous limiting distribution, then $F_N(T_N) \rightarrow^d$ Uniform$(0,1)$.
- **Corollary 2.2**:
  1. Under $H_0^\sharp$, the sharp null p-value $\rho_N^\sharp$ stochastically dominates a Uniform$(0,1)$ distribution for any N.
  2. Under H_0^Y, the oracle p-value ρ_N^Y stochastically dominates a Uniform$(0,1)$ distribution for any N.
  3. If the test statistic T_N has a continuous limiting distribution, then $\rho_N^\sharp \rightarrow^d$ Uniform$(0,1)$ under $H_0^\sharp$ and $\rho_N^Y \rightarrow^d$ Uniform$(0,1)$ under H_0^Y.

Type I Error Control

- **Proposition 2.3**: Let $q(\mathbf{a}, \mathbf{z}, \theta) \equiv \Pr\{\mathbf{A}(\mathbf{z}) = \mathbf{a}; \theta\}$ denote the PMF of the distribution of stochastic potential networks $\mathbf{A}(\mathbf{z})$ at parameter value θ, let $\widehat{\theta}_N$ denote the estimator of θ, and let $F_N(\cdot; \theta)$ denote the CDF of the test statistic T_N at θ, with limiting CDF $F(\cdot; \theta)$. Let θ_0 denote the true value of θ. Assume the following:

  (A1) $\widehat{\theta}_N \to^p \theta_0$

  (A2) $F(t; \theta_0)$ is continuous in t on $\mathbb{R}$

  (A3) there exists a $\delta_0 > 0$ such that

  $$\sup_{\theta \in B_{\delta_0}(\theta_0)} \sup_{t \in \mathbb{R}} |F_N(t; \theta) - F(t; \theta)| \to 0$$

  (A4) $F(t; \theta)$ is continuous in θ at θ_0 uniformly in t, i.e.,

  $$\lim_{\theta \to \theta_0} \sup_{t \in \mathbb{R}} |F(t; \theta) - F(t; \theta_0)| = 0$$

  Then the plug-in p-value $\widehat{p}_N^Y$ converges in distribution to Uniform$(0, 1)$.

Type I Error Control

- **Proposition 2.4**: Let $\rho_N = F_N(T_N)$ for test statistic T_N and CDF F_N (not necessarily the true CDF of T_N). Let $T_N^* = h_N(T_N)$ for a sequence of deterministic, strictly increasing functions h_N. Define $F_N^*(t) = F_N\{h_N^{-1}(t)\}$, the (not necessarily true) CDF of the transformed test statistic, and let $\rho_N^* = F_N^*(T_N^*)$. Then $\rho_N^* = \rho_N$.
- **Corollary 2.5**: Let h_N be a sequence of deterministic, strictly increasing functions.
  1. If the hypotheses of Proposition 2.1 are met for a test statistic T_N, then the results also hold for $T_N^* = h_N(T_N)$.
  2. If the hypotheses of Proposition 2.3 are met for $T_N, F_N(\cdot;\theta)$, then the results also hold for $T_N^* = h_N(T_N), F_N^*(\cdot;\theta) = F_N\{h_N^{-1}(\cdot);\theta\}$.

Simulation Setup

- $n \in \{24, 48\}$ students were equally divided into two residence halls, each with six clusters of equal size
- Within each residence hall group, the clusters were randomized using a 1:1 allocation to either the intervention group or the control group
- Five pairs of students were randomly selected to be roommates, meaning they had contact with each other each week
- Baseline ($k = 1$) face-to-face contacts were simulated between each pair of students with probability 0.5
- Baseline infection statuses were simulated for each student with probability 0.5
- Social networks were simulated over the remaining $\tau \in \{5, 10\}$ weeks according to a STERGM with both formation and persistence models including edge count, intervention assignment Z, infection status Y, $Z \times Y$ interaction, and an offset to force constant edges between roommates
- Infection statuses were simulated for each student i at each week $k \in \{2, \ldots, \tau\}$ with probability $\Pr\left(Y_i^k = 1 | \overline{\mathbf{A}}^{k-1}, \overline{\mathbf{Y}}^{k-1}\right) = h_i(\mathbf{S}^{k-1})$ depending on the vector $\mathbf{S}^{k-1} = \mathbf{A}^{k-1}\mathbf{Y}^{k-1}$ of counts of infected contacts for each student at week $k - 1$.

Simulation Setup

Three scenarios:

Null Hypothesis	**Formation Model Parameters** θ^+	**Persistence Model Parameters** θ^-	**Infection Probability Function** h_i
$H_0^\sharp$	$(-0.5, 0, 0, 0)$	$(-0.5, 0, 0, 0)$	$h_i(\mathbf{s}) = 0.5$
$\overline{H_0^A} \cap H_0^Y$	$(-0.2, 0, 0, -1)$	$(-0.2, 0, 0, -1)$	$h_i(\mathbf{s}) = 0.5$
$\overline{H_0^A} \cap \overline{H_0^Y}$	$(-0.2, 0, 0, -1)$	$(-0.2, 0, 0, -1)$	$h_i(\mathbf{s}) = \frac{n}{2}\text{expit}(s_i - \bar{s}) / \sum_{j=1}^n \text{expit}(s_j - \bar{s})$

Table: The null hypothesis refers to the null hypothesis that is true under the data generating process, formation model parameters are $\theta^+ = (\theta^+_{\text{edges}}, \theta^+_Z, \theta^+_Y, \theta^+_{ZY})$, persistence model parameters are $\theta^- = (\theta^-_{\text{edges}}, \theta^-_Z, \theta^-_Y, \theta^-_{ZY})$, the infection probability function $h_i(\mathbf{s})$ gives the probability of student i being infected at week k given the vector $\mathbf{S}^{k-1} = \mathbf{s} = (s_1, \ldots, s_n)$ of infected contact counts for each student at week $k-1$, and $\bar{s} = n^{-1}\sum_{i=1}^n s_i$.

Three p-values:

- $\rho_B^\sharp$: testing the sharp null
- ρ_B^Y: testing H_0^Y using known q
- $\widehat{\rho}_B^Y$: testing H_0^Y using estimated q

ERGM Change Statistic Model Formulation

- **Change statistic**: $\delta_{\mathbf{g}}(\mathbf{a}, \mathbf{x})_{ij} = \mathbf{g}(\mathbf{a}_{ij}^{+}, \mathbf{x}) - \mathbf{g}(\mathbf{a}_{ij}^{-}, \mathbf{x})$ is the change in network statistic that would occur if a_{ij} were changed from 0 to 1
  - where $\mathbf{a}_{ij}^{+}$ and $\mathbf{a}_{ij}^{-}$ represent the network $\mathbf{a}$ with dyad a_{ij} set to 1 or 0, respectively
- Then the equivalent ERGM specification is

$$\text{logit}\{\Pr_{\theta}(A_{ij} = 1 | \mathbf{A}_{ij}^{C} = \mathbf{a}_{ij}^{C}, \mathbf{X} = \mathbf{x})\} = \theta^{k} \delta_{\mathbf{g}}(\mathbf{a}, \mathbf{x})_{ij}$$

  - where $\mathbf{A}_{ij}^{C}$ represents all dyads in $\mathbf{A}$ except A_{ij}
- **Interpretation of** θ: the change in conditional log-odds of the network associated with a one-unit increase in the corresponding component of $\mathbf{g}(\mathbf{a}, \mathbf{x})$ resulting from switching a particular dyad A_{ij} from 0 to 1 and leaving the rest of the network fixed at $\mathbf{A}_{ij}^{C}$

Dyadic Independence EGRMs

- **Dyadic independence term**: a component g of $\mathbf{g}$ in an ERGM for which the corresponding change statistic $\delta_g(\mathbf{a}, \mathbf{x})_{ij}$ can be calculated for any i, j without knowing $\mathbf{a}$
    - For example, if $g(\mathbf{a}, \mathbf{x}) = \sum_{i,j} a_{ij}(Z_i + Z_j)$ counts the number of edges touching treated nodes, then $\delta_g(\mathbf{a}, \mathbf{x})_{ij} = z_i + z_j$ doesn't depend on $\mathbf{a}$
- **Dyadic independence ERGM**: an ERGM with only dyadic independence terms
    - replace $\delta_g(\mathbf{a}, \mathbf{x})_{ij}$ with $\delta_g(\mathbf{x})_{ij}$ and write the model as

$$\text{logit}\{\Pr\big(A_{ij} = 1 | \mathbf{X} = \mathbf{x}\big)\} = \theta\delta_g(\mathbf{x})_{ij}$$

- **Interpretation of θ**: the change in log-odds of the network associated with a one-unit increase in the corresponding component of $\mathbf{g}(\mathbf{a}, \mathbf{x})$ resulting from switching a particular dyad A_{ij} from 0 to 1

eX-FLU: STERGM Specification

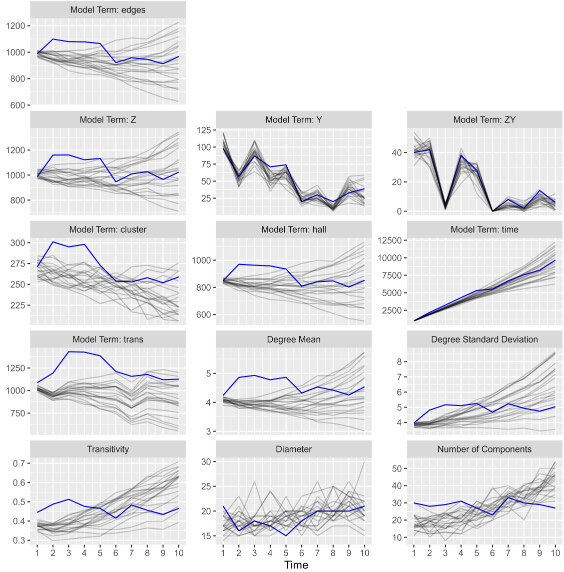
$$\mathbf{g}^{+}(\mathbf{a}^{k+}, \mathbf{a}^{k}, \mathbf{x}^{k+1}) =$$

$$\begin{bmatrix} g^{+}_{\text{edges}}(\mathbf{a}^{k+}, \mathbf{x}^{k+1}) \\ g^{+}_{Z}(\mathbf{a}^{k+}, \mathbf{x}^{k+1}) \\ g^{+}_{Y}(\mathbf{a}^{k+}, \mathbf{x}^{k+1}) \\ g^{+}_{ZY}(\mathbf{a}^{k+}, \mathbf{x}^{k+1}) \\ g^{+}_{\text{room}}(\mathbf{a}^{k+}, \mathbf{x}^{k+1}) \\ g^{+}_{\text{cluster}}(\mathbf{a}^{k+}, \mathbf{x}^{k+1}) \\ g^{+}_{\text{hall}}(\mathbf{a}^{k+}, \mathbf{x}^{k+1}) \\ g^{+}_{\text{time}}(\mathbf{a}^{k+}, \mathbf{x}^{k+1}) \\ g^{+}_{\text{trans}}(\mathbf{a}^{k}, \mathbf{x}^{k+1}) \end{bmatrix} = \begin{bmatrix} \sum_{i,j} a_{ij}^{k+} \\ \sum_{i,j} a_{ij}^{k+}(z_i + z_j) \\ \sum_{i,j} a_{ij}^{k+}(y_i^{k+1} + y_j^{k+1}) \\ \sum_{i,j} a_{ij}^{k+}(z_i y_i^{k+1} + z_j y_j^{k+1}) \sum_{i,j} a_{ij}^{k+} \mathbb{1}(i,j \text{ roommates}) \\ \sum_{i,j} a_{ij}^{k+} \mathbb{1}(i,j \text{ cluster-mates}) \\ \sum_{i,j} a_{ij}^{k+} \mathbb{1}(i,j \text{ hall-mates}) \\ \sum_{i,j} a_{ij}^{k+} \\ \sum_{i,j} \log(1 + \sum_{l} a_{il}^{k} a_{jl}^{k}) \end{bmatrix}$$

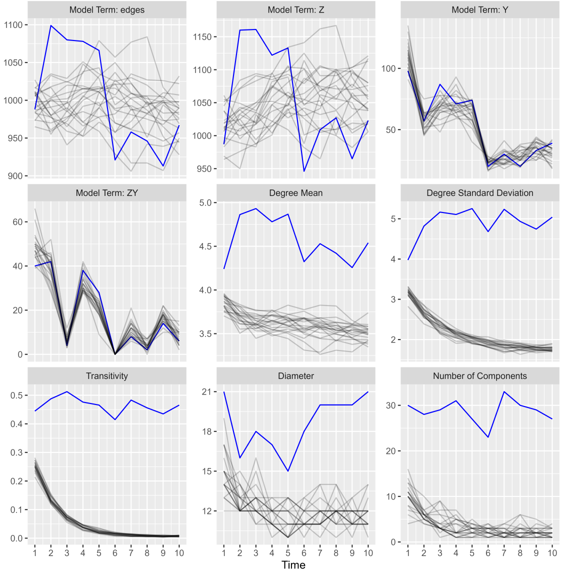
eX-FLU: STERGM Results

Model	Parameter	Estimate (95% Confidence Interval)
Formation	$\theta^{+}_{\text{edges}}$	-7.85 (-8.00, -7.71)
	θ^{+}_{Z}	0.09 (0.02, 0.15)
	θ^{+}_{Y}	0.71 (0.50, 0.93)
	θ^{+}_{ZY}	-0.42 (-0.79, -0.05)
	θ^{+}_{room}	∞ (offset)
	$\theta^{+}_{\text{cluster}}$	2.02 (1.90, 2.14)
	θ^{+}_{hall}	1.71 (1.57, 1.84)
	θ^{+}_{time}	-0.09 (-0.11, -0.07)
	$\theta^{+}_{\text{trans}}$	2.88 (2.82, 2.94)
Persistence	$\theta^{-}_{\text{edges}}$	0.16 (-0.02, 0.33)
	θ^{-}_{Z}	0.03 (-0.03, 0.10)
	θ^{-}_{Y}	-0.06 (-0.32, 0.21)
	θ^{-}_{ZY}	-0.11 (-0.55, 0.33)
	θ^{-}_{room}	∞ (offset)
	$\theta^{-}_{\text{cluster}}$	0.26 (0.13, 0.39)
	θ^{-}_{hall}	0.15 (0.00, 0.31)
	θ^{-}_{time}	0.04 (0.02, 0.06)
	$\theta^{-}_{\text{trans}}$	0.42 (0.35, 0.48)

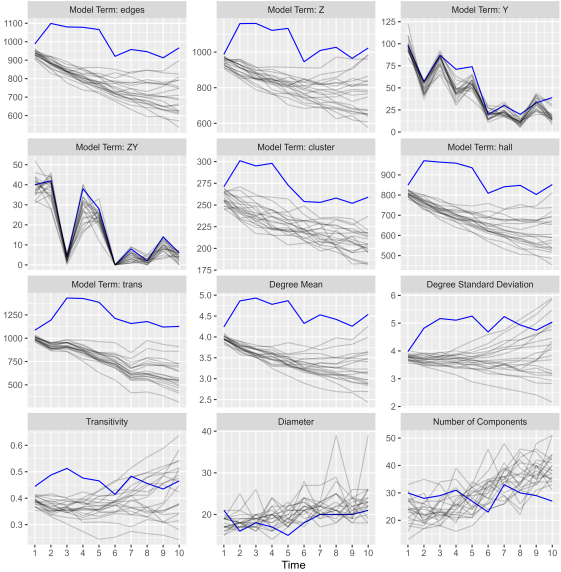
eX-FLU Diagnostics: STERGM



eX-FLU Diagnostics: Alternate STERGM 1



eX-FLU Diagnostics: Alternate STERGM 2



eX-FLU Sensitivity Analysis