Network Causal Inference on Social Media Influence Operations

Harvard Applied Statistics Workshop (Gov 3009)
Oct 31st, 2018

Edward K. Kao, Steven T. Smith

Joint work with Olga Simek, Danelle C. Shah, and Donald B. Rubin

LINCOLN LABORATORY
MASSACHUSETTS INSTITUTE OF TECHNOLOGY

HARVARD UNIVERSITY
DEPARTMENT OF STATISTICS

This material is based upon work supported by the Assistant Secretary of Defense for Research and Engineering under Air Force Contract No. FA8721-05-C-0002 and/or FA8702-15-D-0001. Any opinions, findings, conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the Assistant Secretary of Defense for Research and Engineering.

Outline

- Motivation and introduction
- Network potential outcome causal framework
  - Basic building block: network potential outcomes
  - Theories for design and analysis to address network confounders
- Application on social media influence operations
  - Case study: 2017 French Presidential Election

Motivation for Network Causal Inference

- **How do we quantify the social impact of certain individuals on a network?**

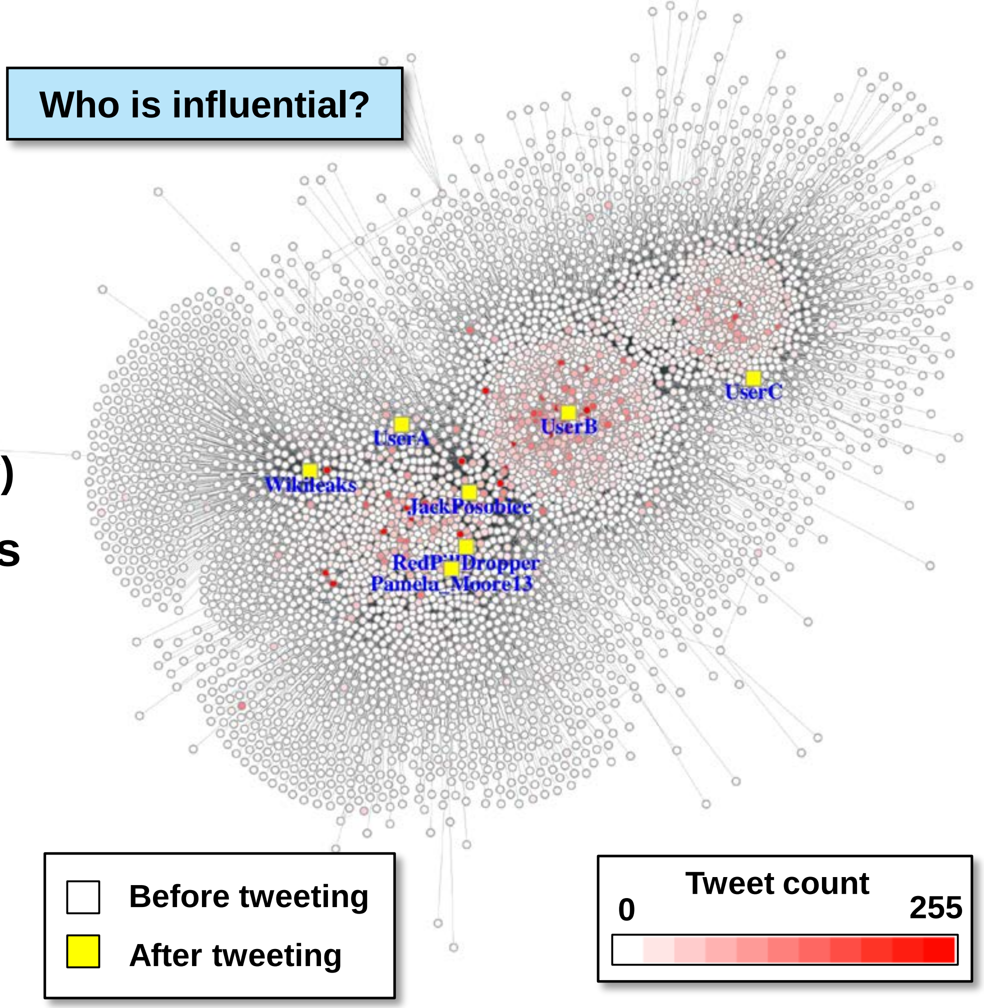
- **Network causal inference provides a framework to quantify impact**
  - **Attributes impact correctly**

    Correlation $\nRightarrow$ Causation

    Disentangle impact from network confounders (e.g. homophily*)
  - **Predictive inference guides optimal “campaign” strategies**

- **Many applications**
  - **Marketing, public health, education, etc.**
  - **Security: influence operations on social media**

#MacronLeaks Spread on Twitter Network



Who is influential?

---

*Shalizi and Thomas, Homophily and Contagion Are Generically Confounded in Observational Social Network Studies *Sociological Methods & Research* 40(2):211–239 (2011)

Publications for This Talk

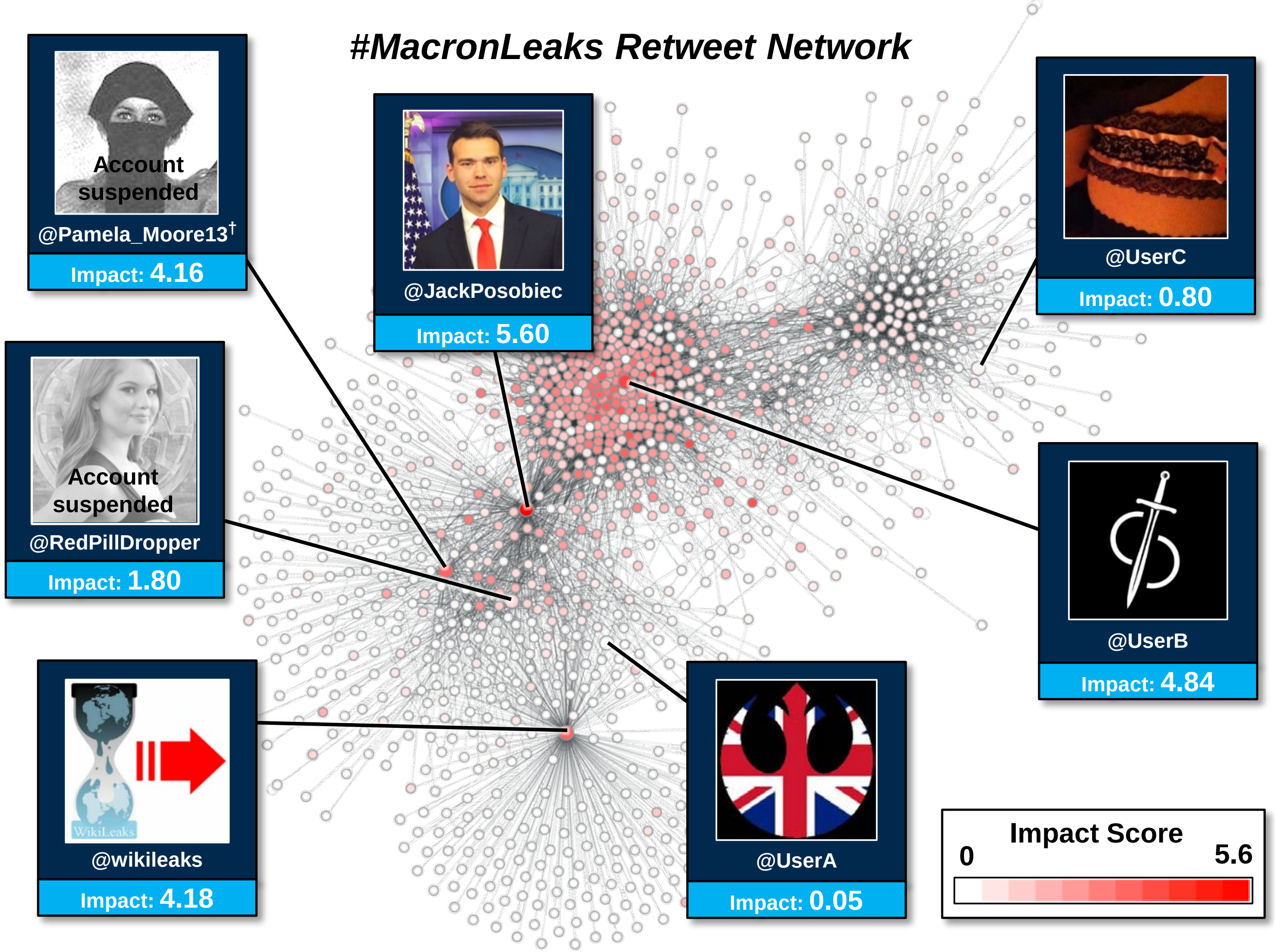
- **Smith, Kao, Simek, Shah, and Rubin, *Influence estimation on social media networks using causal inference*, in *Proc. IEEE SSP* (2018) (patent pending)**
- **Kao, *Causal inference under network interference: A framework for experiments on social networks*. Ph.D. Thesis, Harvard University (2017)**
- **Kao, Airoldi, and Rubin, *Causal inference under network interference: A network potential outcome framework with Bayesian imputation*, in preparation**

Causal Inference Under Network Interference:
Open Area for Methodology Work

We propose a framework for estimating general causal effects under network interference via principled design and estimation

- **Early work (interference as nuisance):**
  - Designs for interference (David & Kempton, 1996, Azais & Bailey, 1993)
- **Hypothesis testing on the presence of effects**
  - Interference between units in randomized experiments (Rosenbaum, 2007, Bowers et. al. 2013)
  - Exact P-values for network interference via artificial experiment (Athey, Eckles, & Imbens, 2015) and conditioning mechanism (Basse, Feller, & Toulis 2018)
- **Estimation of specific causal effects:**
  - Two-staged randomization (Hudgens & Halloran, 2008)
  - Inverse-probability weighting (Aronow & Samii, 2012)
  - Graph cluster randomization (Ugander et al., 2013)
  - Design and estimation under specific structures of network interference (Sussman & Airoldi 2017)
- **Entanglement with social confounders:**
  - Unidentifiability of peer effects among social confounders (Manski, 1993, Shalizi & Thomas 2011)
  - Causal diagram for interference (Ogburn & Vanderweele, 2014)

Causal Impact Estimation on #MacronLeaks Narrative



Screen name	T	RT	F	Earliest time	Pagerank Centrality	Impact*
@JackPosobiec	95	47k	261 k	18:49	2.84	5.60
@RedPillDropper	32	8k	8 k	19:33	2.86	1.80
@UserA	256	59k	1 k	19:34	27.08	0.05
@UserB	260	54k	3 k	20:25	57.05	4.84
@wikileaks	25	63k	5515k	20:32	2.80	4.18
@Pamela_Moore13†	4	4k	54 k	21:14	2.79	4.16
@UserC	1305	51k	< 1 k	22:16	6.36	0.80

*Tweets (T), Retweets (RT), Followers (F), Causal influence estimate**

- **Causal impact score measures contribution to narrative flow on the network, beyond activity-based and topological statistics**
- **High impact accounts corroborated with evidence from the U.S. Congress† and journalistic reports**

---

*Smith et al., Influence estimation on social media networks using causal inference, in *Proc. IEEE SSP* (2018)

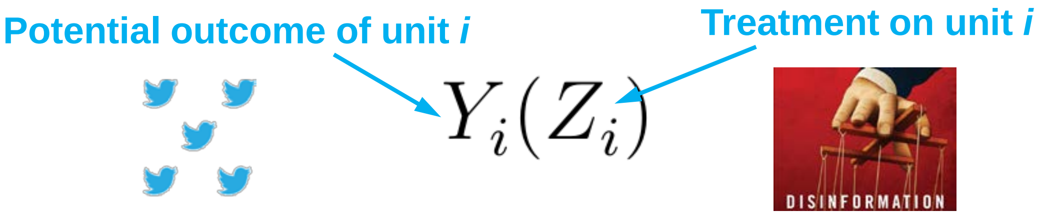
†U.S. HPSCI. Exhibit of the user account handles that Twitter has identified as being tied to Russia's "Internet Research Agency."

Outline

- Motivation and introduction
- Network potential outcome causal framework
  - ➡ Basic building block: network potential outcomes
  - Theories for design and analysis to address network confounders
- Application on social media influence operations
  - Case study: 2017 French Presidential Election

Regular Causal Inference:
Potential Outcome Framework* and Causal Estimand

Regular potential outcomes:



Potential outcome of unit *i*

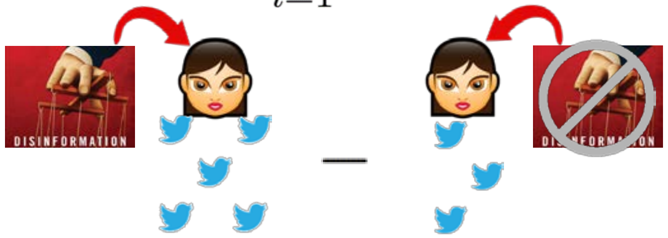
Treatment on unit *i*

$$Y_i(Z_i)$$

Unit	Potential outcomes		Observed outcomes	
	$Z_i = 0$	$Z_i = 1$	$Z_i = 0$	$Z_i = 1$
1	$Y_1(0)$	$Y_1(1)$	$Y_1(0)$	?
2	$Y_2(0)$	$Y_2(1)$	?	$Y_2(1)$

Population average treatment effect:

$$\tau_{ATE} = \frac{1}{N}\sum_{i=1}^{N} Y_i(1) - Y_i(0)$$

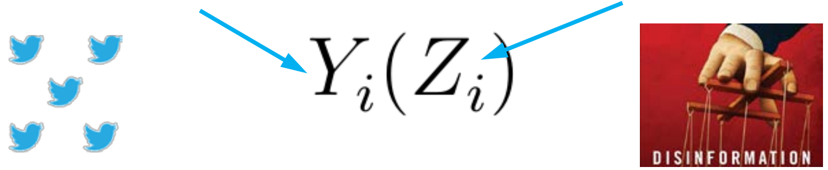


Averaged over the population of N accounts

*Imbens, Guido and Rubin, Donald, "*Causal Inference for Statistics,* Social *and Biomedical Sciences.*" Cambridge University Press (2015).

Regular Causal Inference:
Potential Outcome Framework* and Causal Estimand

Regular potential outcomes:



Potential outcome of unit *i*

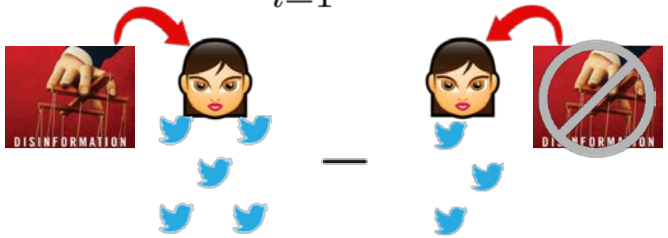
Treatment on unit *i*

$$Y_i(Z_i)$$

Unit	Potential outcomes		Observed outcomes	
	$Z_i = 0$	$Z_i = 1$	$Z_i = 0$	$Z_i = 1$
1	$Y_1(0)$	$Y_1(1)$	$Y_1(0)$	?
2	$Y_2(0)$	$Y_2(1)$	?	$Y_2(1)$

Population average treatment effect:

$$\tau_{ATE} = \frac{1}{N}\sum_{i=1}^{N} Y_i(1) - Y_i(0)$$



Averaged over the population of N accounts

...

1. **Outcome of a unit only depends on its own treatment**
2. **Estimating causal effect is essentially filling in the missing outcomes, by computing:**

Unit covariates (attributes)

Observed outcomes

$$P(\mathbf{Y}_{\text{mis}} | \mathbf{X}, \mathbf{Z}, \mathbf{Y}_{\text{obs}})$$

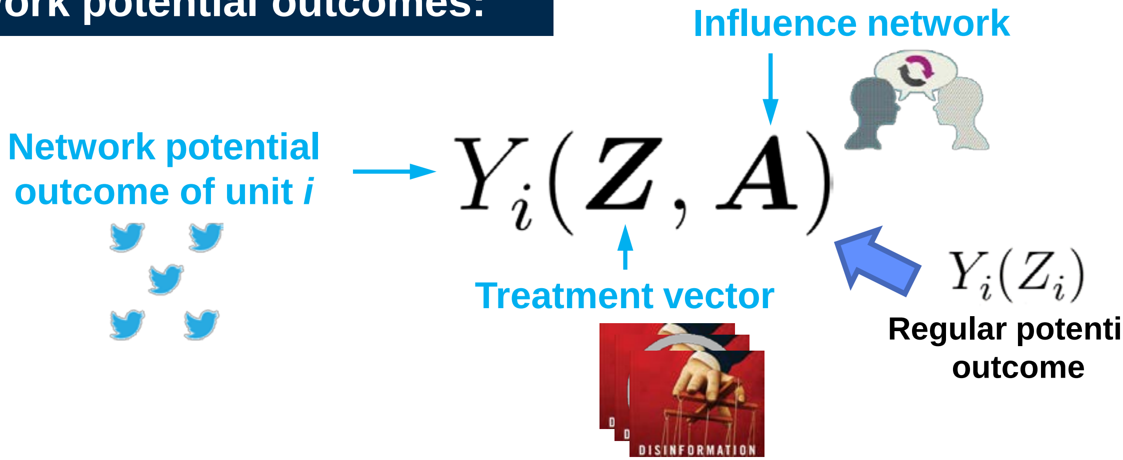
Missing outcomes

Treatment vector

*Imbens, Guido and Rubin, Donald, "*Causal Inference for Statistics,* Social *and Biomedical Sciences.*" Cambridge University Press (2015).

Network Causal Inference: Network Potential Outcome Framework*

Network potential outcomes:



Influence network

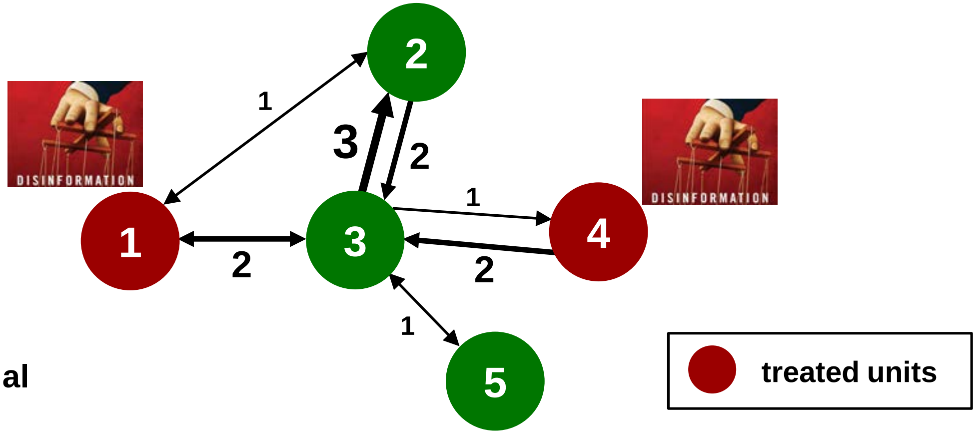
Network potential outcome of unit *i*

$$Y_i(\boldsymbol{Z}, \boldsymbol{A})$$

Treatment vector

$$Y_i(Z_i)$$

Regular potential outcome



Network potential outcomes may be affected by treatments on other units due to social influence on the network

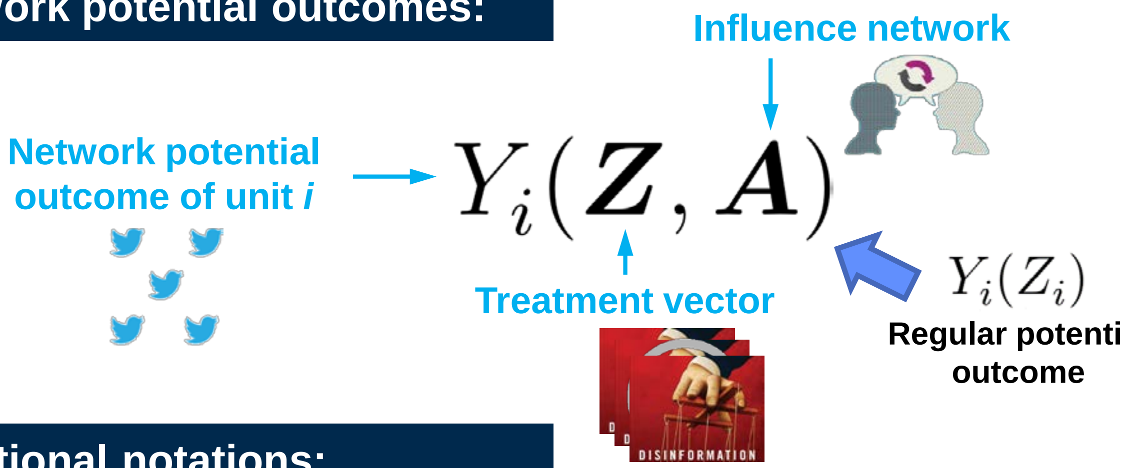
---

*Kao, Causal inference under network interference: A framework for experiments on social networks. Ph.D. Thesis, Harvard University (2017)

*Kao, Airoldi, and Rubin, Causal inference under network interference: A network potential outcome framework with Bayesian imputation, in preparation

Network Causal Inference:
Network Potential Outcome Framework*

Network potential outcomes:



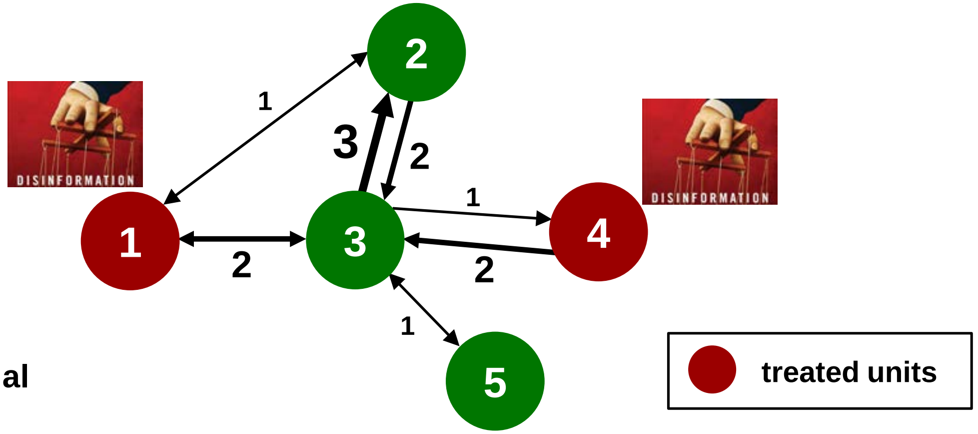
Network potential outcome of unit *i* → $Y_i(\boldsymbol{Z}, \boldsymbol{A})$

Influence network

Treatment vector

$Y_i(Z_i)$

Regular potential outcome



Additional notations:

- **The set of all network potential outcomes for unit i:** $\mathbb{Y}_i = \{Y_i(\boldsymbol{Z} = \boldsymbol{z}, \boldsymbol{A} = \boldsymbol{a})\} \quad \text{for } \forall \boldsymbol{z}, \boldsymbol{a}$
- **The set of all network potential outcomes:** $\mathbb{Y} = \{\mathbb{Y}_i\} \quad \text{for } \forall i$
- **The set of observed and unobserved network potential outcomes are** $\mathbb{Y}_{\text{obs}}$ **and** $\mathbb{Y}_{\text{mis}}$

Network potential outcomes may be affected by treatments on other units due to social influence on the network

---

*Kao, Causal inference under network interference: A framework for experiments on social networks. Ph.D. Thesis, Harvard University (2017)

*Kao, Airoldi, and Rubin, Causal inference under network interference: A network potential outcome framework with Bayesian imputation, in preparation

Causal Estimands for Influence Operations

Average Effect of Treatment on One Individual (Individual Impact):

$$\zeta_k(\boldsymbol{z}) \overset{\text{def}}{=} \frac{1}{N}\sum_{i=1}^{N}\Big(Y_i(\boldsymbol{z}_{k+}, \boldsymbol{A}) - Y_i(\boldsymbol{z}_{k-}, \boldsymbol{A})\Big)$$

Outcome of unit *i*, when *k* is treated

Outcome of unit *i*, when *k* is NOT treated

where z_{k+} is with unit *k* treated

and z_{k-} without

Causal Estimands for Influence Operations

Average Effect of Treatment on One Individual (Individual Impact):

$$\zeta_k(\boldsymbol{z}) \overset{\text{def}}{=} \frac{1}{N}\sum_{i=1}^{N}\Big(Y_i(\boldsymbol{z}_{k+}, \boldsymbol{A}) - Y_i(\boldsymbol{z}_{k-}, \boldsymbol{A})\Big)$$

Outcome of unit *i*, when *k* is treated

Outcome of unit *i*, when *k* is NOT treated

where z_{k+} is with unit *k* treated

and z_{k-} without

Average Effect of Network Manipulation:

$$\zeta_{\boldsymbol{A}\to\boldsymbol{A}'}(\boldsymbol{z}) \overset{\text{def}}{=} \frac{1}{N}\sum_{i=1}^{N}\Big(Y_i(\boldsymbol{z}, \boldsymbol{A}') - Y_i(\boldsymbol{z}, \boldsymbol{A})\Big)$$

Outcome of unit *i*, under network *A'*

Outcome of unit *i*, under network *A*

Define the causal estimands to quantify individual impact and the effect of a specific influence network manipulation

Causal Estimands for Influence Operations

Average Effect of Treatment on One Individual (Individual Impact):

$$\zeta_k(\boldsymbol{z}) \stackrel{\text{def}}{=} \frac{1}{N}\sum_{i=1}^{N}\left(Y_i(\boldsymbol{z}_{k+}, \boldsymbol{A}) - Y_i(\boldsymbol{z}_{k-}, \boldsymbol{A})\right)$$

Outcome of unit *i*, when *k* is treated

Outcome of unit *i*, when *k* is NOT treated

where z_{k+} is with unit *k* treated

and z_{k-} without

Average Effect of Network Manipulation:

$$\zeta_{\boldsymbol{A}\to\boldsymbol{A}'}(\boldsymbol{z}) \stackrel{\text{def}}{=} \frac{1}{N}\sum_{i=1}^{N}\left(Y_i(\boldsymbol{z}, \boldsymbol{A}') - Y_i(\boldsymbol{z}, \boldsymbol{A})\right)$$

Outcome of unit *i*, under network *A'*

Outcome of unit *i*, under network *A*

Estimate the missing outcomes, by computing:

Unit covariates (attributes)

Influence network

$$P(\mathbb{Y}_{mis}|\boldsymbol{X}, \boldsymbol{Z}, \boldsymbol{A}, \mathbb{Y}_{obs})$$

Missing outcomes

Treatment vector

Observed outcomes

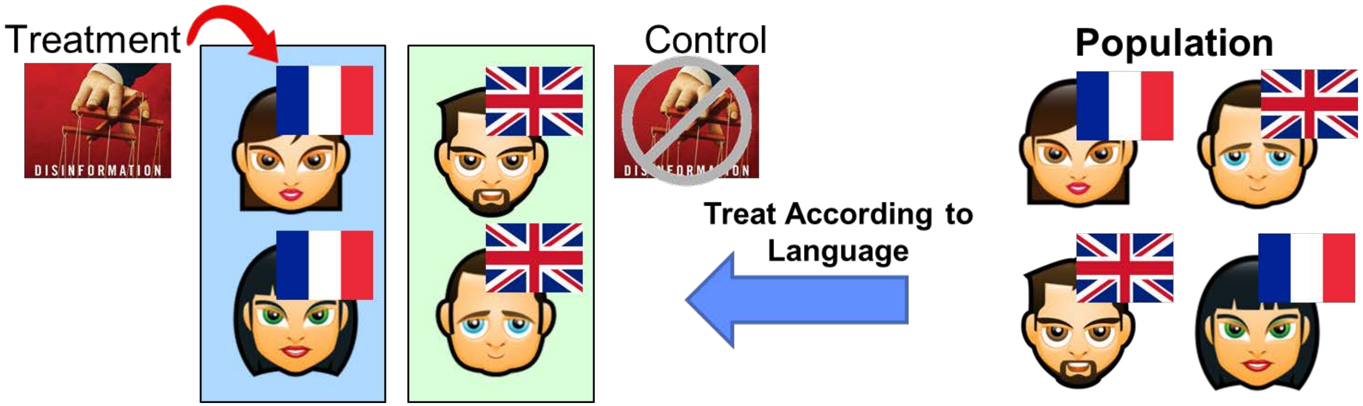
Define the causal estimands to quantify individual impact and the effect of a specific influence network manipulation

Outline

- Motivation and introduction
- Network potential outcome causal framework
  - Basic building block: network potential outcomes
  - Theories for design and analysis to address network confounders
- Application on social media influence operations
  - Case study: 2017 French Presidential Election

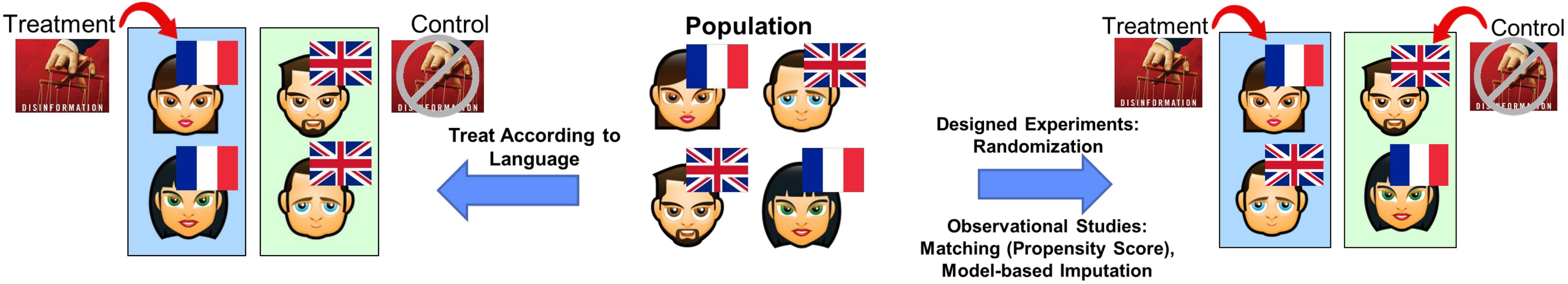
Regular Causal Inference:

Overcoming Selection Bias



Problem: Language effect on outcomes leads to biased causal estimate, if language is a confounder.

Regular Causal Inference:
Overcoming Selection Bias



Treat According to Language

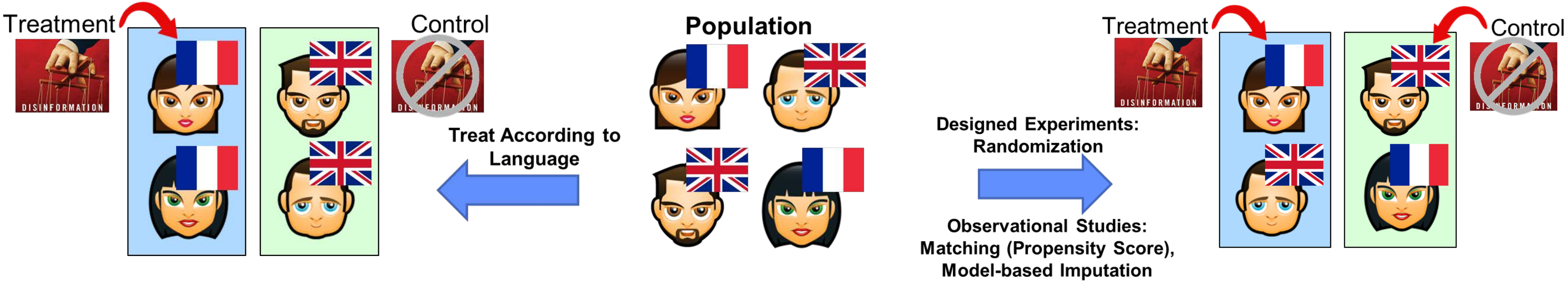
Designed Experiments: Randomization

Observational Studies: Matching (Propensity Score), Model-based Imputation

Problem: Language effect on outcomes leads to biased causal estimate, if language is a confounder.

Solution: Balancing language between treatment and control groups or adjusting for it in the estimation.

Regular Causal Inference:
Overcoming Selection Bias



Treatment

Control

Treat According to Language

Population

Designed Experiments: Randomization

Observational Studies: Matching (Propensity Score), Model-based Imputation

Treatment

Control

Problem: Language effect on outcomes leads to biased causal estimate, if language is a confounder.

Solution: Balancing language between treatment and control groups or adjusting for it in the estimation.

What are the confounding covariates that need to be accounted for?

They are the unit covariates $\boldsymbol{X}$, when conditioned on, leads to independence between treatment assignment $\boldsymbol{Z}$ and the potential outcomes:

$$P(\boldsymbol{Z}|\boldsymbol{X},\boldsymbol{Y}) = P(\boldsymbol{Z}|\boldsymbol{X},\boldsymbol{Y}') \quad \textit{for all } \boldsymbol{Z},\boldsymbol{X},\boldsymbol{Y}, \textit{and } \boldsymbol{Y}'$$

This also simplifies the computation of missing outcomes: $P(\boldsymbol{Y}_{\text{mis}}|\boldsymbol{X},\boldsymbol{Z},\boldsymbol{Y}_{\text{obs}}) = P(\boldsymbol{Y}_{\text{mis}}|\boldsymbol{X},\boldsymbol{Y}_{\text{obs}})$

Overcoming Selection Bias from Treatment and Network Confounders

Desired property indicating all confounders are accounted for in *X*

Unit covariates (attributes)

Influence network

$$P(\mathbb{Y}_{mis}|\boldsymbol{X}, \boldsymbol{Z}, \boldsymbol{A}, \mathbb{Y}_{obs}) = P(\mathbb{Y}_{mis}|\boldsymbol{X}, \mathbb{Y}_{obs})$$

Missing outcomes

Treatment vector

Observed outcomes

How to obtain this property?

Overcoming Selection Bias from Treatment and Network Confounders

Desired property indicating all confounders are accounted for in *X*

Unit covariates (attributes)

Influence network

$$P(\mathbb{Y}_{mis}|\boldsymbol{X},\boldsymbol{Z},\boldsymbol{A},\mathbb{Y}_{obs}) = P(\mathbb{Y}_{mis}|\boldsymbol{X},\mathbb{Y}_{obs})$$

Missing outcomes

Treatment vector

Observed outcomes

How to obtain this property?

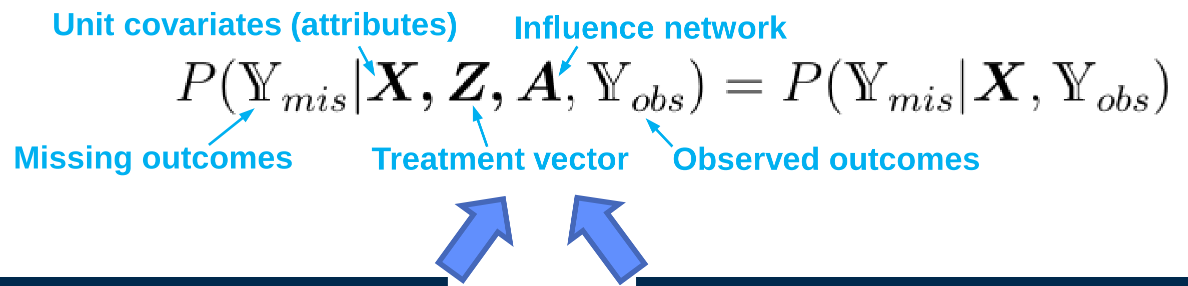
Unconfounded Treatment Condition

$$P(\boldsymbol{Z}|\boldsymbol{X},\boldsymbol{A},\mathbb{Y}) = P(\boldsymbol{Z}|\boldsymbol{X},\boldsymbol{A},\mathbb{Y}')$$
$$\textit{for all } \boldsymbol{Z},\boldsymbol{X},\boldsymbol{A},\mathbb{Y},\textit{ and } \mathbb{Y}'$$

Met if treatment is completely random or determined by the covariates and network

Overcoming Selection Bias from Treatment and Network Confounders

Desired property indicating all confounders are accounted for in *X*



$$P(\mathbb{Y}_{mis}|\boldsymbol{X},\boldsymbol{Z},\boldsymbol{A},\mathbb{Y}_{obs}) = P(\mathbb{Y}_{mis}|\boldsymbol{X},\mathbb{Y}_{obs})$$

How to obtain this property?

Unconfounded Treatment Condition

$$P(\boldsymbol{Z}|\boldsymbol{X},\boldsymbol{A},\mathbb{Y}) = P(\boldsymbol{Z}|\boldsymbol{X},\boldsymbol{A},\mathbb{Y}')$$
$$for\ all\ \boldsymbol{Z},\boldsymbol{X},\boldsymbol{A},\mathbb{Y},\ and\ \mathbb{Y}'$$

Met if treatment is completely random or determined by the covariates and network

Unconfounded Network Condition

$$P(\boldsymbol{A}|\boldsymbol{X},\mathbb{Y}) = P(\boldsymbol{A}|\boldsymbol{X},\mathbb{Y}')$$
$$for\ all\ \boldsymbol{A},\boldsymbol{X},\mathbb{Y},\ and\ \mathbb{Y}'$$

Disentangle network confounders such as homophily

Meet this condition by including the confounding network model parameters in $\boldsymbol{X}$:

$$\boldsymbol{A} \sim H_G(\boldsymbol{X}_G)$$
$$\boldsymbol{X} \supseteq \widetilde{\boldsymbol{X}}_G$$

e.g. community membership, degree distribution

Overcoming Selection Bias from Treatment and Network Confounders

Desired property indicating all confounders are accounted for in *X*

Unit covariates (attributes) Influence network

$$P(\mathbb{Y}_{mis}|\boldsymbol{X}, \boldsymbol{Z}, \boldsymbol{A}, \mathbb{Y}_{obs}) = P(\mathbb{Y}_{mis}|\boldsymbol{X}, \mathbb{Y}_{obs})$$

Missing outcomes Treatment vector Observed outcomes

How to obtain this property?

Unconfounded Treatment Condition

$$P(\boldsymbol{Z}|\boldsymbol{X}, \boldsymbol{A}, \mathbb{Y}) = P(\boldsymbol{Z}|\boldsymbol{X}, \boldsymbol{A}, \mathbb{Y}')$$
$$for\ all\ \boldsymbol{Z}, \boldsymbol{X}, \boldsymbol{A}, \mathbb{Y}, and\ \mathbb{Y}'$$

Met if treatment is completely random or determined by the covariates and network

Unconfounded Network Condition

$$P(\boldsymbol{A}|\boldsymbol{X}, \mathbb{Y}) = P(\boldsymbol{A}|\boldsymbol{X}, \mathbb{Y}')$$
$$for\ all\ \boldsymbol{A}, \boldsymbol{X}, \mathbb{Y}, and\ \mathbb{Y}'$$

Disentangle network confounders such as homophily

Meet this condition by including the confounding network model parameters in $\boldsymbol{X}$:

$$\boldsymbol{A} \sim H_G(\boldsymbol{X}_G)$$
$$\boldsymbol{X} \supseteq \widetilde{\boldsymbol{X}}_G$$

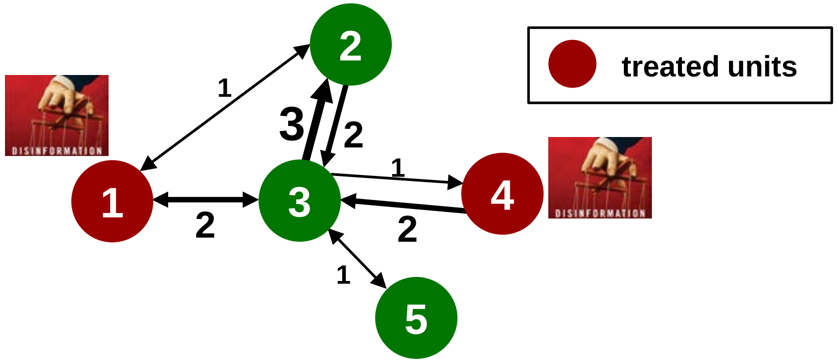
e.g. community membership, degree distribution

The confounders $\boldsymbol{X}$ should be accounted for via both balancing and estimation adjustment

Toy Example: Why is Unconfounded Influence Network Important?

Simulated experiment: estimate social impact

Treatments are completely randomized



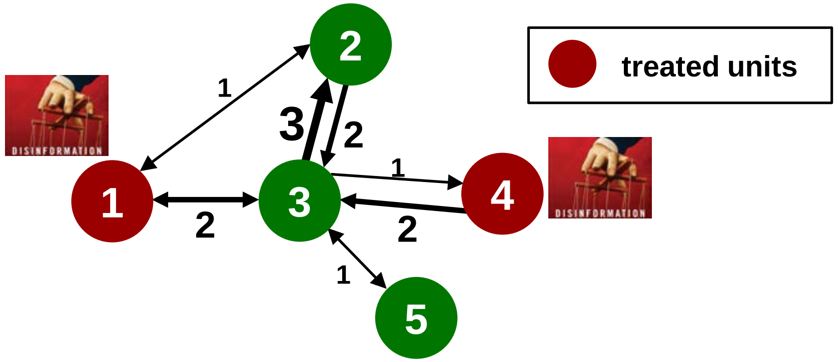
Influence network

$$Z = \begin{bmatrix} 1 \\ 0 \\ 0 \\ 1 \\ 0 \end{bmatrix} \qquad A = \begin{bmatrix} 0 & 1 & 2 & 0 & 0 \\ 1 & 0 & 2 & 0 & 0 \\ 2 & 3 & 0 & 1 & 1 \\ 0 & 0 & 2 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \end{bmatrix}$$

Toy Example: Why is Unconfounded Influence Network Important?

Simulated experiment: estimate social impact

Treatments are completely randomized



$$Z = \begin{bmatrix} 1 \\ 0 \\ 0 \\ 1 \\ 0 \end{bmatrix}$$

Influence network

$$A = \begin{bmatrix} 0 & 1 & 2 & 0 & 0 \\ 1 & 0 & 2 & 0 & 0 \\ 2 & 3 & 0 & 1 & 1 \\ 0 & 0 & 2 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \end{bmatrix}$$

Outcome model:

$$Y(Z,A) = \overbrace{\underset{\text{Primary effect}}{\tau Z} + \underset{\text{Social effect}}{\gamma A^T Z}}^{\text{Treatment effects}} + \overbrace{\underset{\text{Covariate effect}}{\beta X} + \underset{\text{Constant effect}}{\mu} + \underset{\text{Random effect}}{\epsilon}}^{\text{Individual baseline}}$$

Network confounder: activity level

$$X = \begin{bmatrix} 0.4 \\ 0.7 \\ 1.4 \\ 0.3 \\ 0.2 \end{bmatrix}$$

$$Y_{obs} = \begin{bmatrix} 34.02 \\ 11.91 \\ 33.79 \\ 32.93 \\ 2.07 \end{bmatrix}$$

$\tau = 30$

$\gamma = 5$

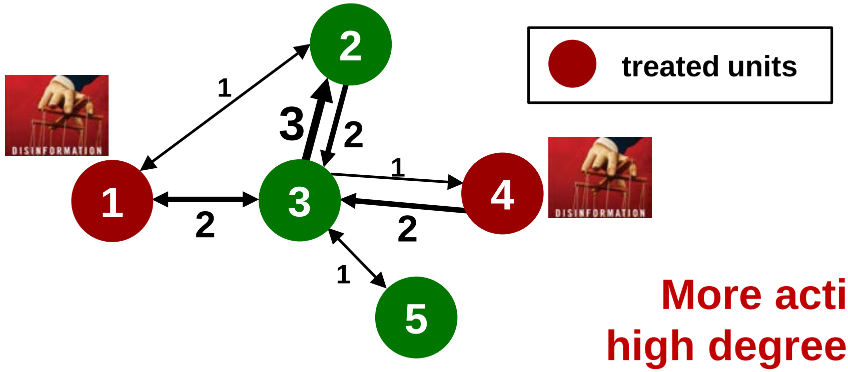
$\beta = 10$

$\mu = 0$

Toy Example: Why is Unconfounded Influence Network Important?

Simulated experiment: estimate social impact

Treatments are completely randomized



Outcome model:

$$Y(Z,A) = \overbrace{\underset{\text{Primary effect}}{\tau Z} + \underset{\text{Social effect}}{\gamma A^T Z}}^{\text{Treatment effects}} + \overbrace{\underset{\text{Covariate effect}}{\beta X} + \underset{\text{Constant effect}}{\mu} + \underset{\text{Random effect}}{\epsilon}}^{\text{Individual baseline}}$$

Network confounder: activity level

$$X = \begin{bmatrix} 0.4 \\ 0.7 \\ \boxed{1.4} \\ 0.3 \\ 0.2 \end{bmatrix}$$

$$Y_{obs} = \begin{bmatrix} 34.02 \\ 11.91 \\ 33.79 \\ 32.93 \\ 2.07 \end{bmatrix}$$

$\tau = 30$

$\gamma = 5$

$\beta = 10$

$\mu = 0$

More active units have high degrees in the network

$$Z = \begin{bmatrix} 1 \\ 0 \\ 0 \\ 1 \\ 0 \end{bmatrix}$$

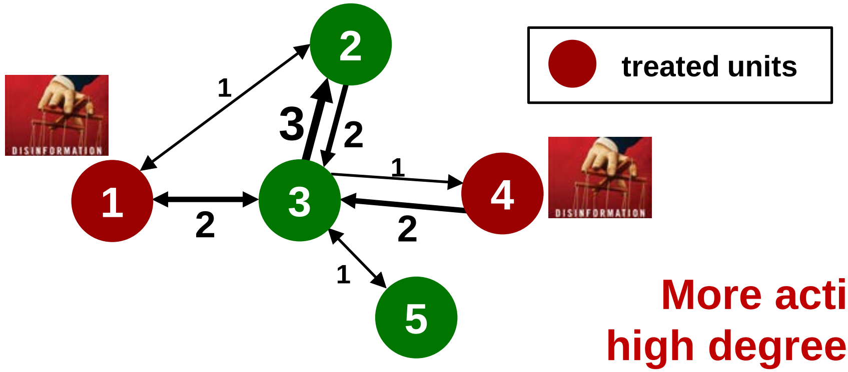
Influence network

$$A = \begin{bmatrix} 0 & 1 & 2 & 0 & 0 \\ 1 & 0 & 2 & 0 & 0 \\ 2 & 3 & 0 & 1 & 1 \\ 0 & 0 & 2 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \end{bmatrix}$$

Toy Example: Why is Unconfounded Influence Network Important?

Simulated experiment: estimate social impact

Treatments are completely randomized



Outcome model:

$$Y(Z,A) = \overbrace{\underset{\text{Primary effect}}{\tau Z} + \underset{\text{Social effect}}{\gamma A^T Z}}^{\text{Treatment effects}} + \overbrace{\underset{\text{Covariate effect}}{\beta X} + \underset{\text{Constant effect}}{\mu} + \underset{\text{Random effect}}{\epsilon}}^{\text{Individual baseline}}$$

Network confounder: activity level

$$X = \begin{bmatrix} 0.4 \\ 0.7 \\ \boxed{1.4} \\ 0.3 \\ 0.2 \end{bmatrix} \qquad Y_{obs} = \begin{bmatrix} 34.02 \\ 11.91 \\ 33.79 \\ 32.93 \\ 2.07 \end{bmatrix}$$

$\tau = 30$

$\gamma = 5$

$\beta = 10$

$\mu = 0$

More active units have high degrees in the network

Influence network

$$Z = \begin{bmatrix} 1 \\ 0 \\ 0 \\ 1 \\ 0 \end{bmatrix} \qquad A = \begin{bmatrix} 0 & 1 & 2 & 0 & 0 \\ 1 & 0 & 2 & 0 & 0 \\ 2 & 3 & 0 & 1 & 1 \\ 0 & 0 & 2 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \end{bmatrix}$$

Social effect coefficient	True Value	Estimation with X (90% PI*)	Estimation w/o X (90% PI*)
γ	5	(4.22, 5.97)	(7.59, 7.95)

* 90% posterior interval obtained through Bayesian regression with weakly informative prior

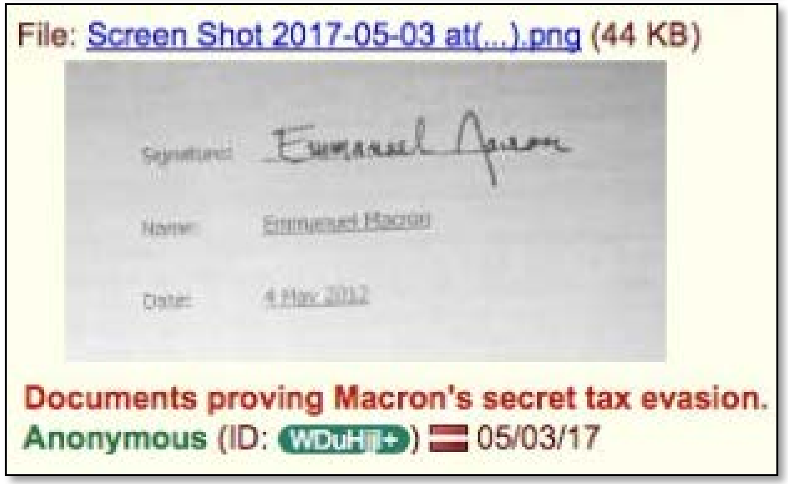
Not conditioning on X breaks the unconfounded influence network condition and leads to biased social effect causal estimate

Outline

- Motivation and introduction
- Network potential outcome causal framework
  - Basic building block: network potential outcomes
  - Theories for design and analysis to address network confounders
- **Application on social media influence operations**
  - ➡ **Case study: 2017 French Presidential Election**

IO Narratives in 2017 French Presidential Election

Cyber



File: Screen Shot 2017-05-03 at(...).png (44 KB)

Signature: Emmanuel Macron

Name: Emmanuel Macron

Date: 4 May 2012

Documents proving Macron's secret tax evasion.

Anonymous (ID: WDuH[illegible]+) 05/03/17

Hacks and Leaks

*“Encrypted data flowing in public communication channels will be among the coveted targets for cyber-attacks”**

Social Media

False amplification via bots and inauthentic “sock puppet” accounts

Traditional Media

RT EN FRANÇAIS

«Sans moi le 7 mai», l'abstentionnisme gagne Twitter

24 avr. 2017, 13:56

Legitimization of fringe narratives; Information manipulation through bias, slant, distortion, omission

During the 2017 French elections IO campaigns were waged on multiple fronts, exploiting global information technologies and networks.

---

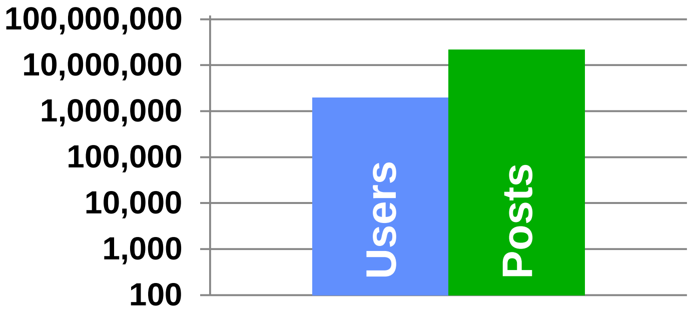
*Чекинов и Богданов, О характере и содержании войны нового поколения, *Военная Мысль* 10:13–24 (2013)
[Chekinov and Bogdanov, On the nature and content of new generation warfare, *Military Thought*]

Detection of IO Narratives on Social Media

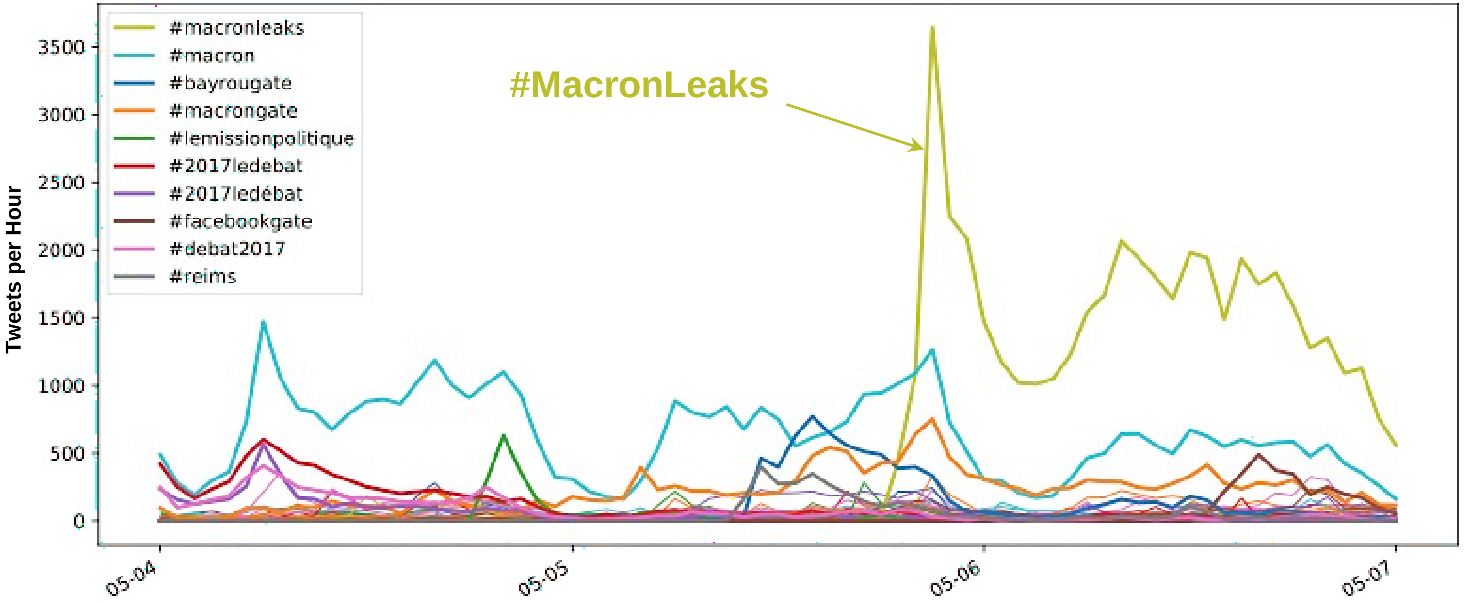
2017 French Elections

- **On May 5, 2017 thousands of internal *En Marche!* documents were leaked online**
- **#MacronLeaks IO campaign involved sources, coordinated amplifiers, and bots**

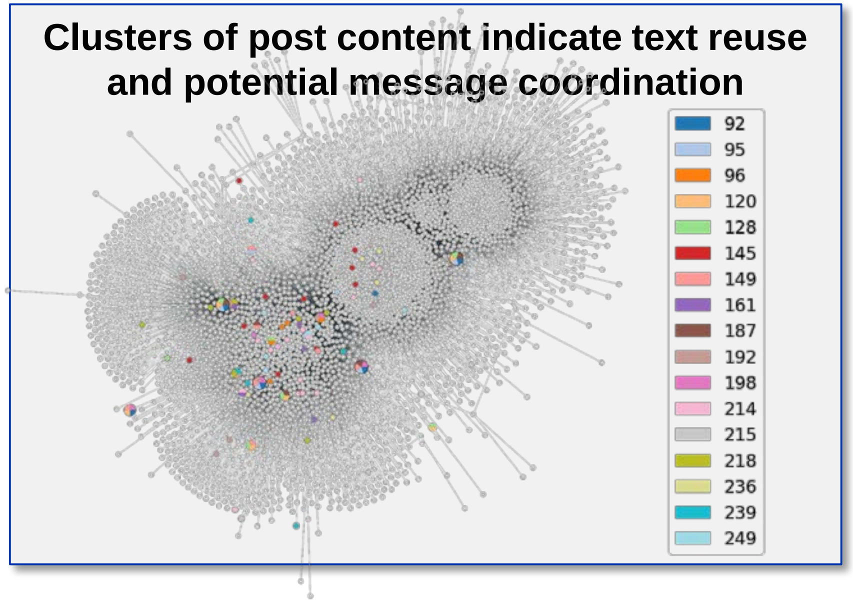
370 GB social media collected April–May 2017



Twitter hashtag usage in days leading up to election



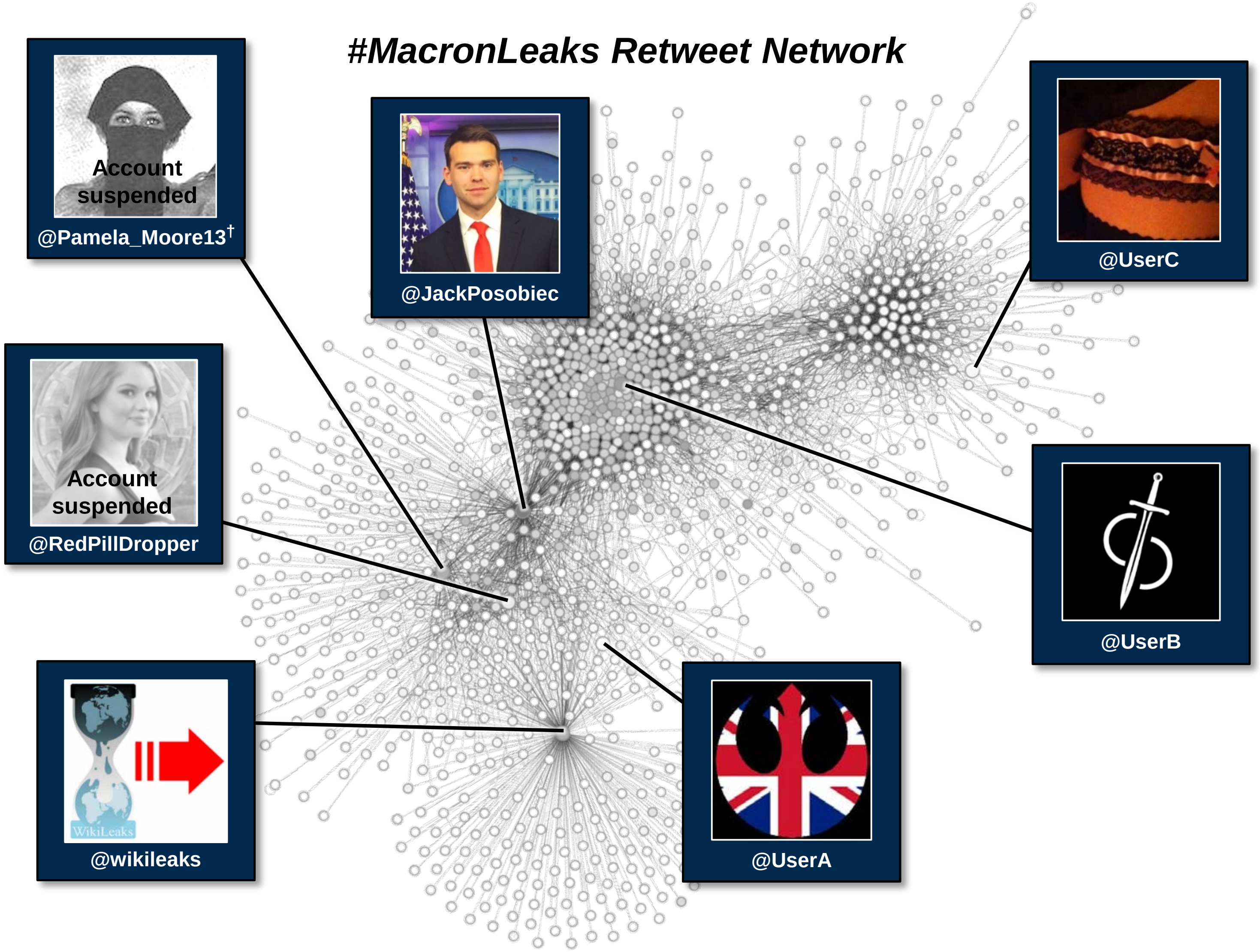
#MacronLeaks



Clusters of post content indicate text reuse and potential message coordination

Causal Impact Estimation on #MacronLeaks Narrative

#MacronLeaks Retweet Network

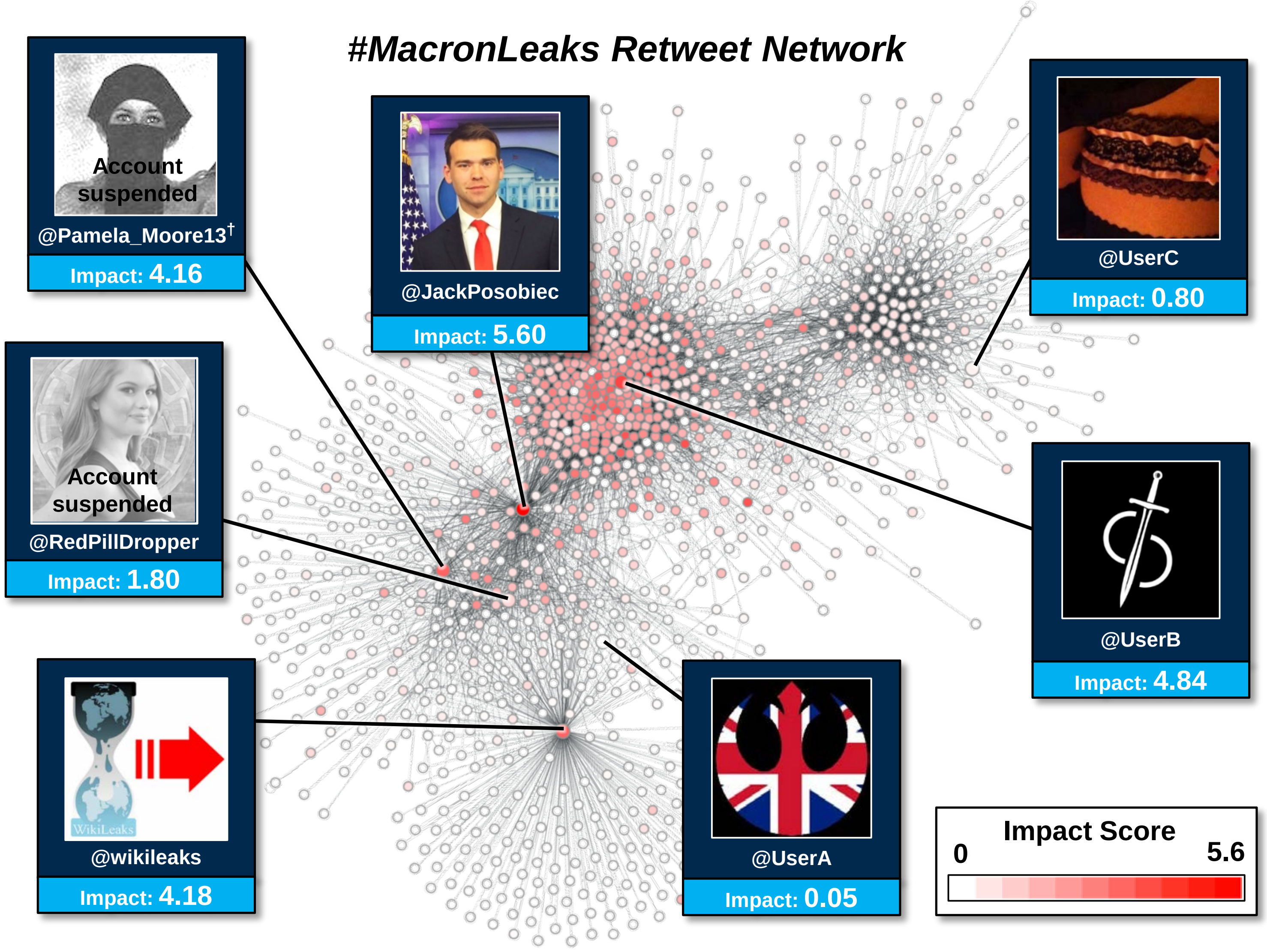


Screen name	T	RT	F	Earliest time	Pagerank Centrality
@JackPosobiec	95	47k	261 k	18:49	2.84
@RedPillDropper	32	8k	8 k	19:33	2.86
@UserA	256	59k	1 k	19:34	27.08
@UserB	260	54k	3 k	20:25	57.05
@wikileaks	25	63k	5515k	20:32	2.80
@Pamela_Moore13†	4	4k	54 k	21:14	2.79
@UserC	1305	51k	< 1 k	22:16	6.36

Tweets (T), Retweets (RT), Followers (F)

- **“Impact” is often quantified by count-based statistics or network metrics (e.g., retweets or centrality)**
- **These measures do not fully capture the extent to which the network exposure of a narrative/rumor can be attributed to any particular individual(s)**

Causal Impact Estimation on #MacronLeaks Narrative



#MacronLeaks Retweet Network

Screen name	T	RT	F	Earliest time	Pagerank Centrality	Impact*
@JackPosobiec	95	47k	261 k	18:49	2.84	5.60
@RedPillDropper	32	8k	8 k	19:33	2.86	1.80
@UserA	256	59k	1 k	19:34	27.08	0.05
@UserB	260	54k	3 k	20:25	57.05	4.84
@wikileaks	25	63k	5515k	20:32	2.80	4.18
@Pamela_Moore13†	4	4k	54 k	21:14	2.79	4.16
@UserC	1305	51k	< 1 k	22:16	6.36	0.80

*Tweets (T), Retweets (RT), Followers (F), Causal influence estimate**

- **Causal impact score measures contribution to narrative flow on the network, beyond activity-based and topological statistics**
- **High impact accounts corroborated with evidence from the U.S. Congress† and journalistic reports**

---

*Smith et al., Influence estimation on social media networks using causal inference, in *Proc. IEEE SSP* (2018)

†U.S. HPSCI. Exhibit of the user account handles that Twitter has identified as being tied to Russia's "Internet Research Agency."

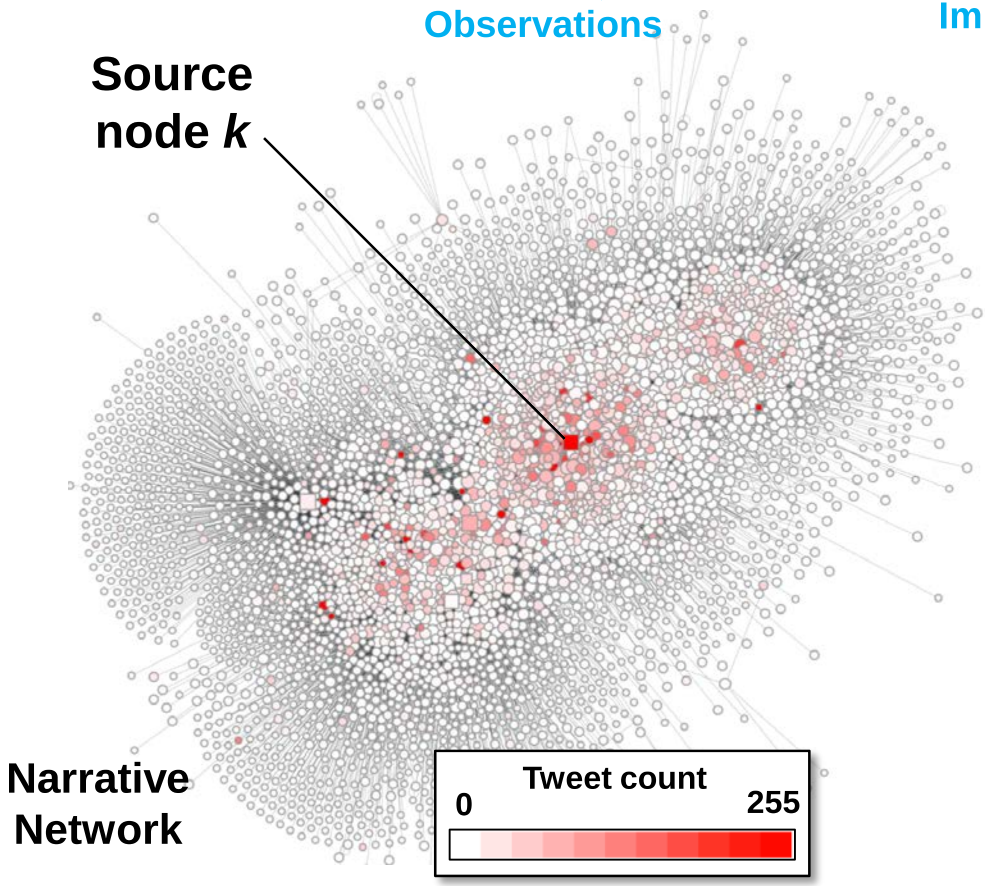
Network Causal Inference for Impact Estimation*

Impact Estimand:

$$\zeta_k = \text{Average}[\; Y_i\,(z_{k+}, \text{A}) - Y_i\,(z_{k-}, \text{A})\;]$$

Observations

Imputations



- **Outcomes are the individual activities on the narrative (e.g. tweet counts)**
- **Explicitly measures each account’s contribution to the outcomes**

*Smith et al., Influence estimation on social media networks using causal inference, in *Proc. IEEE SSP*, to appear (patent pending)
Available at https://arxiv.org/abs/1804.04109

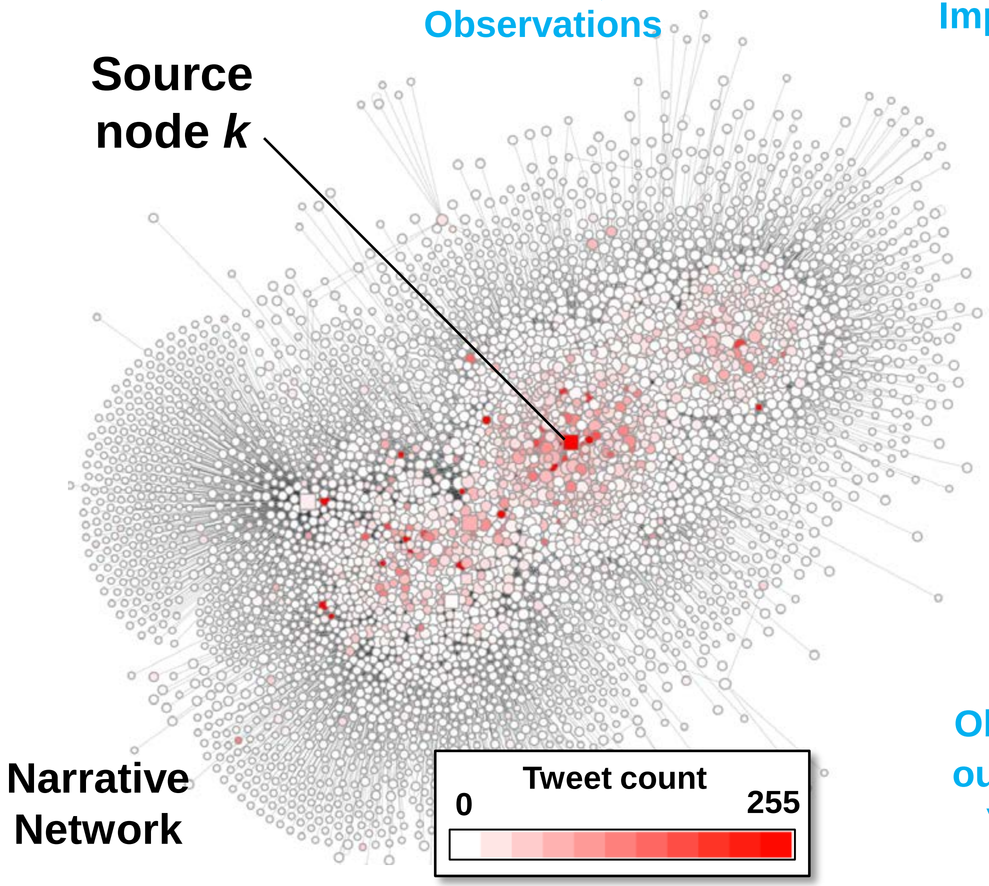
Network Causal Inference for Impact Estimation*

Impact Estimand:

$$\zeta_k = \text{Average}[\ Y_i\,(z_{k+}, A) - Y_i\,(z_{k-}, A)\ ]$$

Observations

Imputations



- Outcomes are the individual activities on the narrative (e.g. tweet counts)
- Explicitly measures each account's contribution to the outcomes
- Outcome model accounts for narrative propagation on the network

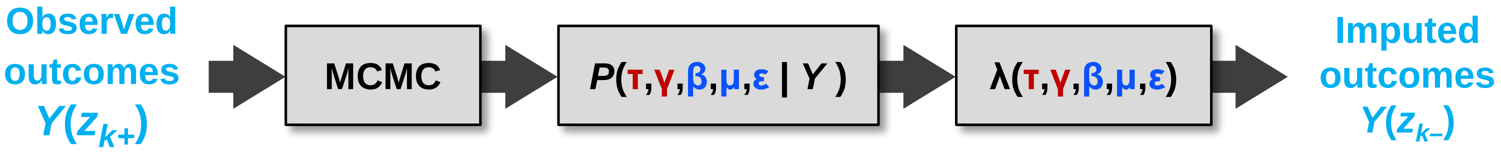
Network potential outcome model:

$$Y_i \sim \text{Poisson}(\lambda_i)$$

$$\log \lambda_i = \tau Z_i + \Sigma\Pi\tau\gamma_j\, s_i + \beta^T x_i + \mu + \varepsilon_i$$

Exposure to source

Individual baseline



Observed outcomes $Y(z_{k+})$

MCMC

$P(\tau,\gamma,\beta,\mu,\varepsilon \mid Y)$

$\lambda(\tau,\gamma,\beta,\mu,\varepsilon)$

Imputed outcomes $Y(z_{k-})$

*Smith et al., Influence estimation on social media networks using causal inference, in *Proc. IEEE SSP*, to appear (patent pending)
Available at https://arxiv.org/abs/1804.04109

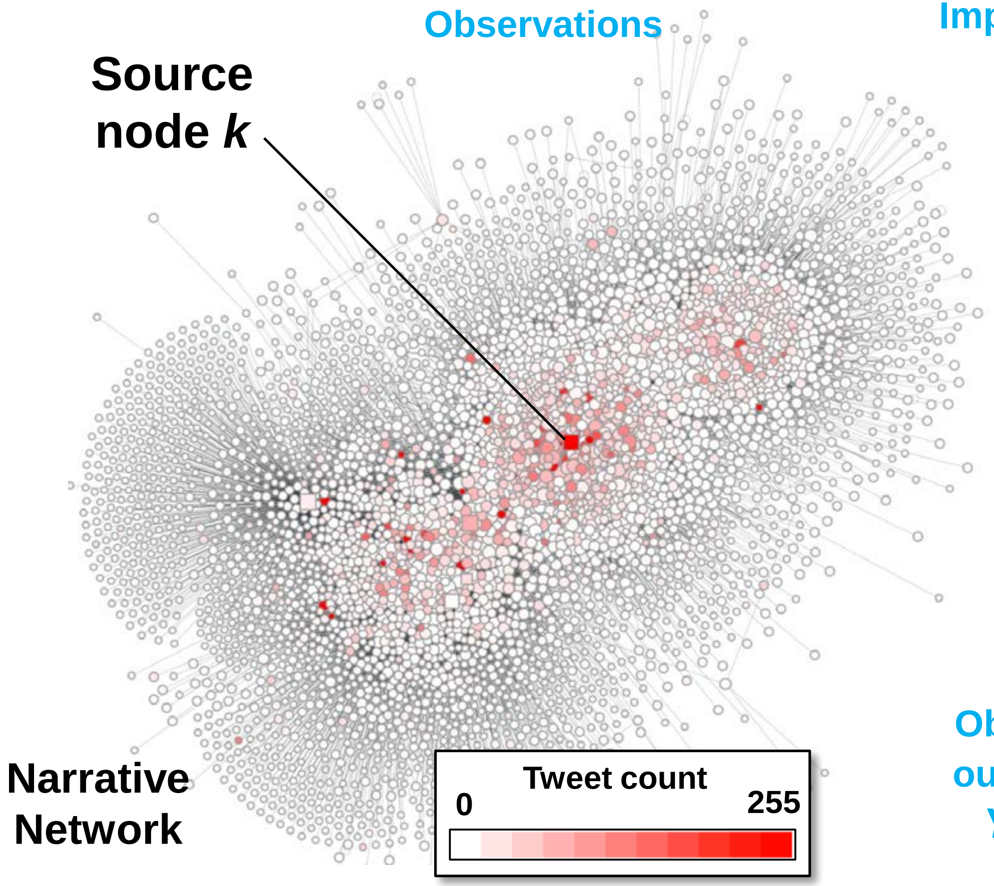
Network Causal Inference for Impact Estimation*

Impact Estimand:

$$\zeta_k = \text{Average}[\ Y_i\,(z_{k+}, A) - Y_i\,(z_{k-}, A)\ ]$$

Observations

Imputations



- Outcomes are the individual activities on the narrative (e.g. tweet counts)
- Explicitly measures each account’s contribution to the outcomes
- Outcome model accounts for narrative propagation on the network
- Causal framework disentangles confounders (e.g. homophily) from social influence

Network potential outcome model:

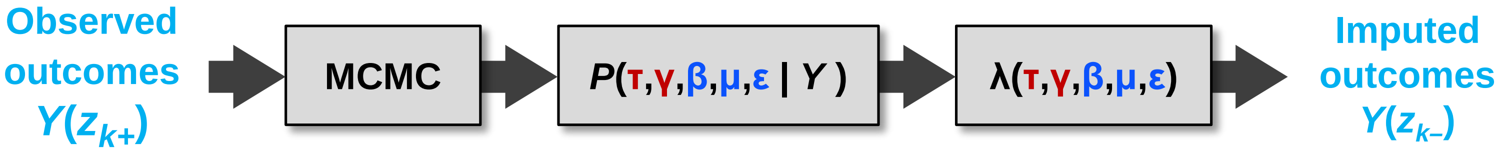
$$Y_i \sim \text{Poisson}(\lambda_i)$$

$$\log \lambda_i = \tau Z_i + \Sigma\Pi\tau\gamma_j s_i + \beta^T x_i + \mu + \varepsilon_i$$

Accounts for confounders (e.g. node degrees and community membership)

Exposure to source

Individual baseline



Observed outcomes $Y(z_{k+})$

MCMC

$P(\tau,\gamma,\beta,\mu,\varepsilon \mid Y)$

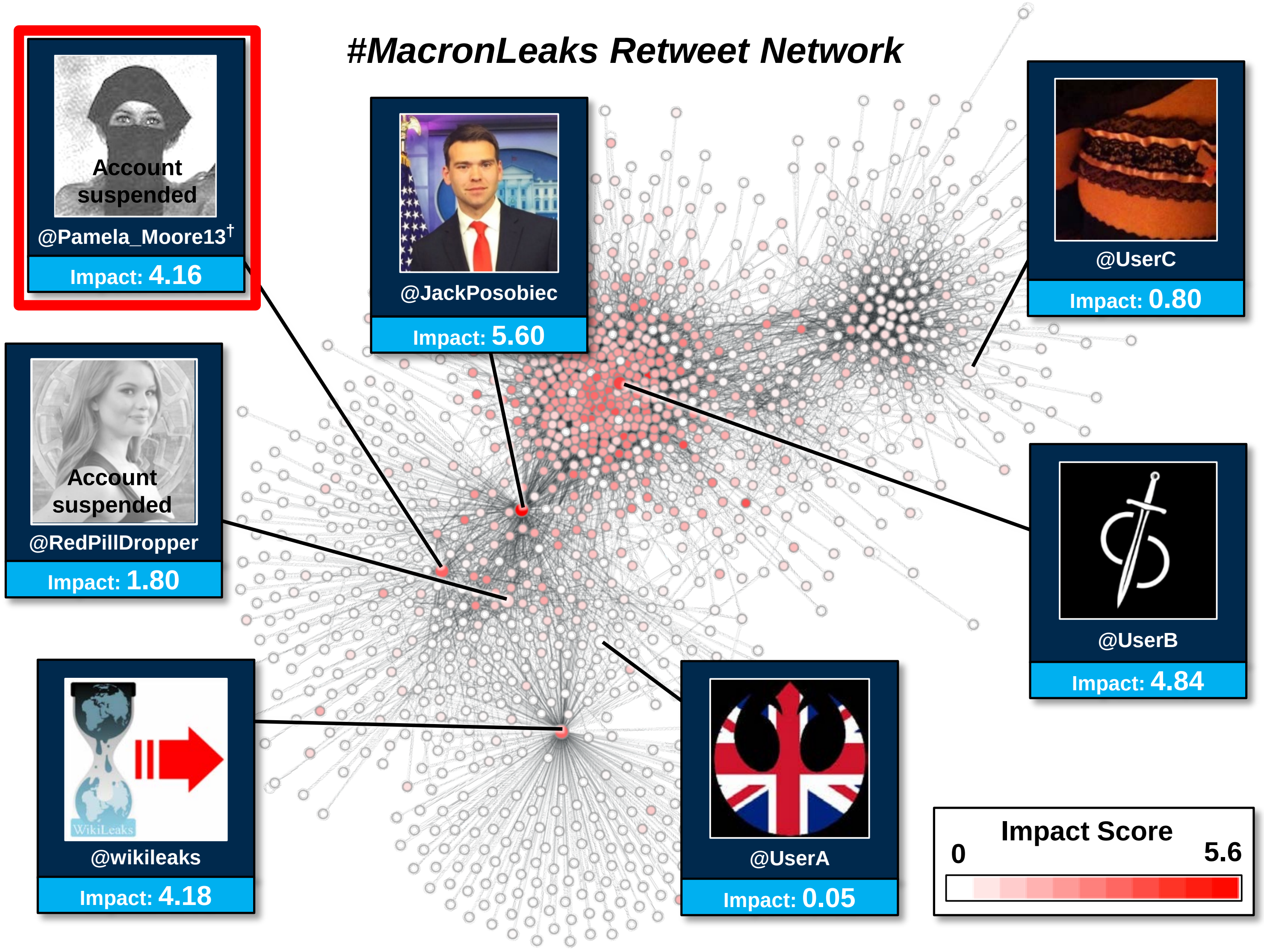
$\lambda(\tau,\gamma,\beta,\mu,\varepsilon)$

Imputed outcomes $Y(z_{k-})$

*Smith et al., Influence estimation on social media networks using causal inference, in *Proc. IEEE SSP*, to appear (patent pending)
Available at https://arxiv.org/abs/1804.04109

Causal Impact Estimation on #MacronLeaks Narrative

#MacronLeaks Retweet Network



Screen name	T	RT	F	Earliest time	Pagerank Centrality	Impact*
@JackPosobiec	95	47k	261 k	18:49	2.84	5.60
@RedPillDropper	32	8k	8 k	19:33	2.86	1.80
@UserA	256	59k	1 k	19:34	27.08	0.05
@UserB	260	54k	3 k	20:25	57.05	4.84
@wikileaks	25	63k	5515k	20:32	2.80	4.18
@Pamela_Moore13[†]	4	4k	54 k	21:14	2.79	4.16
@UserC	1305	51k	< 1 k	22:16	6.36	0.80

*Tweets (T), Retweets (RT), Followers (F), Causal influence estimate**

- Causal impact score measures contribution to narrative flow on the network, beyond activity-based and topological statistics
- High impact accounts corroborated with evidence from the U.S. Congress[†] and journalistic reports

---

[†] U.S. HPSCI. Exhibit of user accounts that Twitter has identified as being tied to Russia's "Internet Research Agency." (Nov 2017)

*Smith et al., Influence estimation on social media networks using causal inference, in *Proc. IEEE SSP* (2018)

Summary and Future Work

- Presented a network causal inference framework to quantify social impact
- Applied to finding key influencers in social media influence operations
  - Demonstrated on the 2017 French Presidential Election
  - On-going work:
    - Detect and characterize more complex narratives
    - Recommend intervention via predictive inference and network control
- Open questions:
  - How to effectively balance confounders across many treatment exposure groups?
  - How to best impute missing network potential outcomes and mitigate model mis-specifications?
  - Other applications for network causal inference?

Backups

Modeling the Potential Outcomes With Network Propagation GLM (Net-Prop GLM)

Generalized linear model (GLM) with the appropriate link function g() and distribution for the potential outcomes $Y_i(Z_{\mathcal{N}_i}, A_{\mathcal{N}_i})$

$$E[Y_i] = g^{-1}(\overset{\text{primary treatment effect}}{\boxed{\tau Z_i}} + \underset{\text{1}^{\text{st}}\text{-hop peer effect}}{\boxed{\tau\gamma_1 s_{1,i}}} + \overset{\text{2}^{\text{nd}}\text{-hop peer effect}}{\boxed{\tau\gamma_1\gamma_2 s_{2,i}}} + \dots + \underset{\text{L}^{\text{th}}\text{-hop peer effect}}{\boxed{\tau\gamma_1\dots\gamma_L s_{L,i}}} + \overset{\text{covariate effect}}{\boxed{\boldsymbol{\beta^T x_i}}} + \underset{\text{mean effect}}{\boxed{\mu}} + \overset{\text{random effect}}{\boxed{\epsilon_i}})$$

$$A_{l,ji} \overset{\text{iid}}{\sim} \text{Pois}(\lambda_{ji})$$

$$\epsilon_i \sim \mathbb{N}(0, \sigma^2 = c)$$

$$s_{1,i} = \sum_{j \in \mathcal{N}_i^1} Z_j A_{1,ji} \sim \text{Pois}(\kappa_{1,i}) \quad \text{where} \quad \kappa_{1,i} = \sum_{j \in \mathcal{N}_i^1} Z_j \lambda_{ji}$$

$$s_{l,i} = \sum_{j \in \mathcal{N}_i^1} s_{l-1,j} A_{l,ji} \sim \text{Pois}(\kappa_{l,i}) \quad \text{where} \quad \kappa_{l,i} = \sum_{j \in \mathcal{N}_i^1} s_{l-1,j} \lambda_{ji}$$

parameter estimation using MCMC with Bayesian regressions and M-H steps

Outcome Distribution	Link Function	Effects Property
Normal	Identity	Additive
Binomial	Logistic	Additive, slow start and diminishing return
Poisson	Log	Multiplicative

Causal Estimands for Primary Treatment Effects

Unit-Level Effect With Fixed Neighborhood Assignment:

$$\xi_i(\boldsymbol{z}_{\bar{\imath}}) \overset{\text{def}}{=} Y_i(Z_i = 1, \boldsymbol{Z}_{\bar{\imath}} = \boldsymbol{z}_{\bar{\imath}}, \boldsymbol{A}) - Y_i(Z_i = 0, \boldsymbol{Z}_{\bar{\imath}} = \boldsymbol{z}_{\bar{\imath}}, \boldsymbol{A})$$

Unit-Level Effect Averaged Over All Neighborhood Assignments:

$$\xi_i^{\text{ave}} \overset{\text{def}}{=} \frac{1}{2^{(N-1)}} \sum_{\boldsymbol{z} \in \mathcal{Z}_{\bar{\imath}}} \xi_i(\boldsymbol{z}_{\bar{\imath}})$$

Average Population Effect Over All Neighborhood Assignments:

$$\xi^{\text{ave}} \overset{\text{def}}{=} \frac{1}{N} \sum_{i=1:N} \xi_i^{\text{ave}}$$

Network potential outcomes are the basic building blocks for simple to more complicated causal quantities

Causal Estimands for Fixed Treatment Assignment

Average Peer Effect:

$$\delta^{\text{fix}}(\boldsymbol{z}) \overset{\text{def}}{=} \frac{1}{N} \sum_{i=1:N} Y_i(Z_i = z_i, \boldsymbol{Z}_{\bar{\imath}} = \boldsymbol{z}_{\bar{\imath}}, \boldsymbol{A}) - Y_i(Z_i = z_i, \boldsymbol{Z}_{\bar{\imath}} = \boldsymbol{0}, \boldsymbol{A})$$

Average Effect of Treatment on One Individual (Individual Impact):

$$\zeta_i(\boldsymbol{z}) \overset{\text{def}}{=} \frac{1}{N} \sum_{j=1}^{N} \Big(Y_j(\boldsymbol{Z} = \boldsymbol{z}_{i+}, \boldsymbol{A}) - Y_j(\boldsymbol{Z} = \boldsymbol{z}_{i-}, \boldsymbol{A}) \Big)$$

where $\boldsymbol{z}_{i+}$ is the fixed treatment with unit *i* treated and $\boldsymbol{z}_{i-}$ without

Average Effect of Network Manipulation:

$$\zeta_{\boldsymbol{A} \to \boldsymbol{A}'}(\boldsymbol{z}) \overset{\text{def}}{=} \frac{1}{N} \sum_{j=1}^{N} Y_j(\boldsymbol{Z} = \boldsymbol{z}, \boldsymbol{A}') - Y_j(\boldsymbol{Z} = \boldsymbol{z}, \boldsymbol{A})$$

Define the causal estimands according to the question of interest

Simplification Through Unconfoundedness Assumptions

Theorem: Simplified Imputation Under Network Interference

If the unconfounded treatment assignment assumption and the unconfounded influence network assumption are both met, the network treatment mechanism $(\boldsymbol{Z}, \boldsymbol{A})$ does not enter the posterior distribution of the missing potential outcomes:

$$P(\mathbb{Y}_{mis}|\boldsymbol{X}, \boldsymbol{Z}, \boldsymbol{A}, \mathbb{Y}_{obs}) = P(\mathbb{Y}_{mis}|\boldsymbol{X}, \mathbb{Y}_{obs})$$

This simplification allows us to compute the posterior distribution of $\mathbb{Y}_{mis}$ by accounting for the critical unit covariates $\boldsymbol{X}$

Unconfounded Treatment Assignment Assumption

Unconfounded Assignment Assumption under Network Interference

Conditional on the relevant unit covariates $\boldsymbol{X}$ and the influence network $\boldsymbol{A}$, the treatment assignment $\boldsymbol{Z}$ does not depend on the potential outcomes:

$$P(\boldsymbol{Z}|\boldsymbol{X}, \boldsymbol{A}, \mathbb{Y}) = P(\boldsymbol{Z}|\boldsymbol{X}, \boldsymbol{A}, \mathbb{Y}') \quad \mathit{for\ all}\ \boldsymbol{Z}, \boldsymbol{X}, \boldsymbol{A}, \mathbb{Y}, \mathit{and}\ \mathbb{Y}'$$

The assignment is unconfounded if the treatment is completely random or determined by the covariates and the influence network

Unconfounded Influence Network Assumption

Unconfounded Influence Network Assumption under Network Interference

Conditional on the relevant unit covariates $\boldsymbol{X}$, the influence network $\boldsymbol{A}$ does not depend on the potential outcomes:

$$P(\boldsymbol{A}|\boldsymbol{X}, \mathbb{Y}) = P(\boldsymbol{A}|\boldsymbol{X}, \mathbb{Y}') \quad \textit{for all } \boldsymbol{A}, \boldsymbol{X}, \mathbb{Y}, \textit{and } \mathbb{Y}'$$

We will see how this assumption can be met with a parametric network model

Principled Methodology for Network Causal Inference

Theorem: Unconfounded Influence Network by Conditioning on Network Parameters

The unconfounded influence network assumption:

$$P(\boldsymbol{A}|\boldsymbol{X},\mathbb{Y}) = P(\boldsymbol{A}|\boldsymbol{X},\mathbb{Y}') \quad \textit{for all } \boldsymbol{A}, \boldsymbol{X}, \mathbb{Y}, \textit{and } \mathbb{Y}'$$

is met if:

1. The distribution of the influence network $\boldsymbol{A}$ can be characterized by a model H_G with nodal parameters X_G and population parameters $\boldsymbol{\Theta}_G$:

$$\boldsymbol{A} \sim H_G(\boldsymbol{X}_G, \boldsymbol{\Theta}_G)$$

2. The influence network $\boldsymbol{A}$ correlates with the potential outcomes $\mathbb{Y}$ only through a subset of the nodal parameters $\widetilde{\boldsymbol{X}}_G \in \boldsymbol{X}_G$ and population parameters $\widetilde{\boldsymbol{\Theta}}_G \in \boldsymbol{\Theta}_G$
3. The unit covariates $\boldsymbol{X}$ contain these network parameters $\widetilde{\boldsymbol{X}}_G$, $\widetilde{\boldsymbol{\Theta}}_G$

The confounding covariates $\boldsymbol{X}$ should be accounted for in both the design and analysis phase of the experiment