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

The overall classification of this briefing is UNCLASSIFIED
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 $\neq$ 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

Publications for This Talk


• 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

• Early work (interference as nuisance):

• 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)

We propose a framework for estimating general causal effects under network interference via principled design and estimation.
Causal Impact Estimation on #MacronLeaks Narrative

#MacronLeaks Retweet Network

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

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 $i$

$Y_i(Z_i)$

Treatment on unit $i$

<table>
<thead>
<tr>
<th>Unit</th>
<th>Potential outcomes</th>
<th>Observed outcomes</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$Z_i = 0$</td>
<td>$Z_i = 1$</td>
</tr>
<tr>
<td>1</td>
<td>$Y_1(0)$</td>
<td>$Y_1(1)$</td>
</tr>
<tr>
<td>2</td>
<td>$Y_2(0)$</td>
<td>$Y_2(1)$</td>
</tr>
</tbody>
</table>

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

Regular Causal Inference: Potential Outcome Framework* and Causal Estimand

Regular potential outcomes:

Potential outcome of unit $i$  
\[ Y_i(Z_i) \]

Treatment on unit $i$

<table>
<thead>
<tr>
<th>Unit</th>
<th>Potential outcomes</th>
<th>Observed outcomes</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$Z_i = 0$</td>
<td>$Z_i = 0$</td>
</tr>
<tr>
<td></td>
<td>$Z_i = 1$</td>
<td>$Z_i = 1$</td>
</tr>
<tr>
<td>1</td>
<td>$Y_1(0)$</td>
<td>$Y_1(0)$</td>
</tr>
<tr>
<td></td>
<td>$Y_1(1)$</td>
<td>$?$</td>
</tr>
<tr>
<td>2</td>
<td>$Y_2(0)$</td>
<td>$?$</td>
</tr>
<tr>
<td></td>
<td>$Y_2(1)$</td>
<td>$Y_2(1)$</td>
</tr>
</tbody>
</table>

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:

\[ P(Y_{mis} | X, Z, Y_{obs}) \]

Unit covariates (attributes)

Missing outcomes

Observed outcomes

Treatment vector

Network potential outcomes may be affected by treatments on other units due to social influence on the network.
Network Causal Inference: Network Potential Outcome Framework*

Network potential outcomes:

\[ Y_i(Z, A) \]

Influence network

Treatment vector

Regular potential outcome

Additional notations:

- The set of all network potential outcomes for unit i:
  \[ Y_i = \{ Y_i(Z = z, A = a) \} \text{ for } \forall z, a \]

- The set of all network potential outcomes:
  \[ Y = \{ Y_i \} \text{ for } \forall i \]

- The set of observed and unobserved network potential outcomes are \( Y_{\text{obs}} \) and \( Y_{\text{mis}} \)

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


*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):

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

where $z_{k+}$ is with unit $k$ treated and $z_{k-}$ without

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

Outcome of unit $i$, when $k$ is NOT treated
Causal Estimands for Influence Operations

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

\[ \zeta_k(z) \overset{\text{def}}{=} \frac{1}{N} \sum_{i=1}^{N} \left( Y_i(z_{k+}, A) - Y_i(z_{k-}, A) \right) \]

where \( z_{k+} \) is with unit \( k \) treated and \( z_{k-} \) without

Outcome of unit \( i \), when \( k \) is treated

Outcome of unit \( i \), when \( k \) is NOT treated

Average Effect of Network Manipulation:

\[ \zeta_{A \rightarrow A'}(z) \overset{\text{def}}{=} \frac{1}{N} \sum_{i=1}^{N} \left( Y_i(z, A') - Y_i(z, A) \right) \]

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):

\[ \zeta_k(z) \overset{\text{def}}{=} \frac{1}{N} \sum_{i=1}^{N} \left( Y_i(z_{k+}, A) - Y_i(z_{k-}, A) \right) \]

where \( z_{k+} \) is with unit \( k \) treated and \( z_{k-} \) without

Outcome of unit \( i \), when \( k \) is treated
Outcome of unit \( i \), when \( k \) is NOT treated

Average Effect of Network Manipulation:

\[ \zeta_{A \rightarrow A'}(z) \overset{\text{def}}{=} \frac{1}{N} \sum_{i=1}^{N} \left( Y_i(z, A') - Y_i(z, A) \right) \]

Outcome of unit \( i \), under network \( A' \)
Outcome of unit \( i \), under network \( A \)

Estimate the missing outcomes, by computing:

\[ P(Y_{mis}|X, Z, A, Y_{obs}) \]

Unit covariates (attributes) Influence network
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

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

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 $X$, when conditioned on, leads to independence between treatment assignment $Z$ and the potential outcomes:

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

This also simplifies the computation of missing outcomes:

$$P(Y_{\text{mis}}|X, Z, Y_{\text{obs}}) = P(Y_{\text{mis}}|X, Y_{\text{obs}})$$
Desired property indicating all confounders are accounted for in $X$

$$P(Y_{mis}|X, Z, A', Y_{obs}) = P(Y_{mis}|X, Y_{obs})$$

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

$$P(Y_{mis} | X, Z, A', Y_{obs}) = P(Y_{mis} | X, Y_{obs})$$

Missing outcomes Treatment vector Observed outcomes

How to obtain this property?

Unconfounded Treatment Condition

$$P(Z | X, A, Y) = P(Z | X, A, Y')$$
for all $Z, X, A, Y$, and $Y'$

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 \( X \)

Unit covariates (attributes) Influence network

\[
P(Y_{mis} | X, Z, A', Y_{obs}) = P(Y_{mis} | X, Y_{obs})
\]

Missing outcomes Treatment vector Observed outcomes

How to obtain this property?

Unconfounded Treatment Condition

\[
P(Z | X, A, Y) = P(Z | X, A, Y')
\]

for all \( Z, X, A, Y, \) and \( Y' \)

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

Unconfounded Network Condition

\[
P(A | X, Y) = P(A | X, Y')
\]

for all \( A, X, Y, \) and \( Y' \)

Meet this condition by including the confounding network model parameters in \( X \):

\[
A \sim H_G(X_G)
\]

\[
X \supseteq \bar{X}_G
\]

e.g. community membership, degree distribution

Disentangle network confounders such as homophily
Overcoming Selection Bias from Treatment and Network Confounders

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

Unit covariates (attributes) Influence network

$P(Y_{mis}|X, Z, A, Y_{obs}) = P(Y_{mis}|X, Y_{obs})$

Missing outcomes Treatment vector Observed outcomes

How to obtain this property?

Unconfounded Treatment Condition

$P(Z|X, A, Y) = P(Z|X, A, Y')$

for all $Z, X, A, Y,$ and $Y'$

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

Unconfounded Network Condition

$P(A|X, Y) = P(A|X, Y')$

for all $A, X, Y,$ and $Y'$

Meet this condition by including the confounding network model parameters in $X$:

$A \sim H_G(X_G)$

$X \supseteq \tilde{X}_G$

e.g. community membership, degree distribution

The confounders $X$ should be accounted for via both balancing and estimation adjustment

Disentangle network confounders such as homophily
Toy Example: Why is Unconfounded Influence Network Important?

Simulated experiment: estimate social impact

Treatments are completely randomized

Influence network

$$Z = \begin{bmatrix} 1 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \end{bmatrix}, \quad A = \begin{bmatrix} 0 & 1 & 2 & 0 & 0 \\ 1 & 0 & 2 & 0 & 0 \\ 2 & 3 & 0 & 1 & 1 \\ 0 & 0 & 2 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \end{bmatrix}$$
Toy Example: Why is Unconfounded Influence Network Important?

Simulated experiment: estimate social impact

Treatments are completely randomized

Outcome model:

\[ Y(Z, A) = \tau Z + \gamma A^T Z + \beta X + \mu + \epsilon \]

Network confounder: activity level

Influence network

\[ Z = \begin{bmatrix} 1 & 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \end{bmatrix} \]

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

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

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

Treatment effects

Primary effect

Social effect

Covariate effect

Constant effect

Random effect

\[ \tau = 30 \]

\[ \gamma = 5 \]

\[ \beta = 10 \]

\[ \mu = 0 \]
Toy Example: Why is Unconfounded Influence Network Important?

Simulated experiment: estimate social impact

Treatments are completely randomized

More active units have high degrees in the network

Outcome model:

\[ Y(Z, A) = \tau Z + \gamma A^T Z + \beta X + \mu + \epsilon \]

Network confounder: activity level

Influence network

\[ Z = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \end{bmatrix} \]

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

Treatment effects

- Primary effect \( \tau = 30 \)
- Social effect \( \gamma = 5 \)
- Covariate effect \( \beta = 10 \)
- Constant effect \( \mu = 0 \)
- Random effect

Simulated experiment: estimate social impact

Network confounder: activity level
Toy Example: Why is Unconfounded Influence Network Important?

Simulated experiment: estimate social impact

Treatments are completely randomized

Influence network

Network confounder: activity level

More active units have high degrees in the network

Outcome model:

$Y(Z, A) = \tau Z + \gamma A^T Z + \beta X + \mu + \epsilon$

Treatment effects

- Primary effect
- Social effect
- Covariate effect
- Constant effect
- Random effect

Individual baseline

- $\tau = 30$
- $\gamma = 5$
- $\beta = 10$
- $\mu = 0$

Social effect coefficient

<table>
<thead>
<tr>
<th>True Value</th>
<th>Estimation with $X$ (90% PI*)</th>
<th>Estimation w/o $X$ (90% PI*)</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\gamma$</td>
<td>5</td>
<td>(4.22, 5.97)</td>
</tr>
</tbody>
</table>

* 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
• 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**

<table>
<thead>
<tr>
<th>Cyber</th>
<th>Social Media</th>
<th>Traditional Media</th>
</tr>
</thead>
<tbody>
<tr>
<td><img src="https://example.com" alt="Image of a screen shot" /></td>
<td><img src="https://example.com" alt="Image of tweets" /></td>
<td><img src="https://example.com" alt="Image of a sign saying 'Abstention'" /></td>
</tr>
</tbody>
</table>

**Hacks and Leaks**

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

**False amplification via bots and inauthentic “sock puppet” accounts**

**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.

---

*Chekinov and Bogdanov, On the nature and content of new generation warfare, Military Thought*
• On May 5, 2017 thousands of internal *En Marche!* documents were leaked online
• #MacronLeaks IO campaign involved sources, coordinated amplifiers, and bots
Causal Impact Estimation on #MacronLeaks Narrative

- **#MacronLeaks Retweet Network**

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

- “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

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

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

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

Available at https://arxiv.org/abs/1804.04109

Smith et al., Influence estimation on social media networks using causal inference, in Proc. IEEE SSP, to appear (patent pending)
Network Causal Inference for Impact Estimation*

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

Network potential outcome model:
\[ Y_i \sim \text{Poisson}(\lambda_i) \]
\[ \log \lambda_i = \tau Z_i + \Sigma \Pi \gamma_j s_i + \beta^T x_i + \mu + \epsilon_i \]

- 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

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:
\[ z_k = \text{Average} \left[ Y_i(z_{k^+}, A) - Y_i(z_{k^-}, A) \right] \]

- 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:

\[ Y_i \sim \text{Poisson}(\lambda_i) \]
\[ \log \lambda_i = \tau Z_i + \sum \pi_j s_i + \beta^T x_i + \mu + \epsilon_i \]

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

\[ z_k = \text{Average} \left[ Y_i(z_{k^+}, A) - Y_i(z_{k^-}, A) \right] \]

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

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

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 \( Y_i \mid Z_i, A_i \)

\[
E[Y_i] = g^{-1}(\tau Z_i) + \tau \gamma_1 s_{1,i} + \tau \gamma_1 \gamma_2 s_{2,i} + \ldots + \tau \gamma_1 \ldots \gamma_L s_{L,i} + \beta^T x_i + \mu + \epsilon_i
\]

- **Primary treatment effect**
- **1st-hop peer effect**
- **2nd-hop peer effect**
- **Lth-hop peer effect**
- **Covariate effect**
- **Random effect**
- Mean effect

\( A_{l,j,i} \sim \text{Pois}(\lambda_{j,i}) \)

\( s_{1,i} = \sum_{j \in N_i^1} Z_j A_{1,j,i} \sim \text{Pois}(\kappa_{1,i}) \quad \text{where} \quad \kappa_{1,i} = \sum_{j \in N_i^1} Z_j \lambda_{j,i} \)

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

\( \epsilon_i \sim N(0, \sigma^2 = c) \)

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

<table>
<thead>
<tr>
<th>Outcome Distribution</th>
<th>Link Function</th>
<th>Effects Property</th>
</tr>
</thead>
<tbody>
<tr>
<td>Normal</td>
<td>Identity</td>
<td>Additive</td>
</tr>
<tr>
<td>Binomial</td>
<td>Logistic</td>
<td>Additive, slow start and diminishing return</td>
</tr>
<tr>
<td>Poisson</td>
<td>Log</td>
<td>Multiplicative</td>
</tr>
</tbody>
</table>
Causal Estimands for Primary Treatment Effects

Unit-Level Effect With Fixed Neighborhood Assignment:

\[ \xi_i(z_i) \overset{\text{def}}{=} Y_i(Z_i = 1, Z_i = z_i, A) - Y_i(Z_i = 0, Z_i = z_i, A) \]

Unit-Level Effect Averaged Over All Neighborhood Assignments:

\[ \xi_{\text{ave}} \overset{\text{def}}{=} \frac{1}{2(N-1)} \sum_{z \in \mathcal{Z}_i} \xi_i(z_i) \]

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:**

\[
\delta_{\text{fix}}(z) \overset{\text{def}}{=} \frac{1}{N} \sum_{i=1:N} Y_i(Z_i = z_i, Z_{\bar{i}} = z_{\bar{i}}, A) - Y_i(Z_i = z_i, Z_{\bar{i}} = 0, A)
\]

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

\[
\zeta_i(z) \overset{\text{def}}{=} \frac{1}{N} \sum_{j=1}^{N} (Y_j(Z = z_{i+}, A) - Y_j(Z = z_{i-}, A))
\]

where \(Z_{i+}\) is the fixed treatment with unit \(i\) treated and \(Z_{i-}\) without

**Average Effect of Network Manipulation:**

\[
\zeta_{A\rightarrow A'}(z) \overset{\text{def}}{=} \frac{1}{N} \sum_{j=1}^{N} Y_j(Z = z, A') - Y_j(Z = z, A)
\]

Define the causal estimands according to the question of interest
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 \((Z, A)\) does not enter the posterior distribution of the missing potential outcomes:

\[
P(Y_{mis} | X, Z, A, Y_{obs}) = P(Y_{mis} | X, Y_{obs})
\]

This simplification allows us to compute the posterior distribution of \(Y_{mis}\) by accounting for the critical unit covariates \(X\).
Unconfounded Assignment Assumption under Network Interference

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

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

The assignment is unconfounded if the treatment is completely random or determined by the covariates and the influence network.
Unconfounded Influence Network Assumption under Network Interference

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

$$P(\mathcal{A}|\mathbf{X}, Y) = P(\mathcal{A}|\mathbf{X}, Y') \quad \text{for all } \mathcal{A}, \mathbf{X}, Y, \text{ and } Y'$$

We will see how this assumption can be met with a parametric network model.
Theorem: Unconfounded Influence Network by Conditioning on Network Parameters

The unconfounded influence network assumption:

\[ P(A|X, Y) = P(A|X, Y') \quad \text{for all } A, X, Y, \text{ and } Y' \]

is met if:

1. The distribution of the influence network \( A \) can be characterized by a model \( H_G \) with nodal parameters \( X_G \) and population parameters \( \Theta_G \):

\[ A \sim H_G(X_G, \Theta_G) \]

2. The influence network \( A \) correlates with the potential outcomes \( Y \) only through a subset of the nodal parameters \( \tilde{X}_G \subseteq X_G \) and population parameters \( \tilde{\Theta}_G \subseteq \Theta_G \)

3. The unit covariates \( X \) contain these network parameters \( \tilde{X}_G, \tilde{\Theta}_G \)

The confounding covariates \( X \) should be accounted for in both the design and analysis phase of the experiment.