A Decomposition Analysis of Diffusion Over a Large Network

Mahdi E. Kahou and Kyungchul Song
University of British Columbia

A Celebration of Peter Phillips’ Forty Years at Yale
October 19, 2018
Diffusion Over a Network

- Suppose that we observe cross-sectional outcomes over two periods $t = 0, 1$:

  $$(Y_{i,0})_{i \in N}, (Y_{i,1})_{i \in N}.$$  

- A directed network $G = (N, E)$ is recorded. If a link $ij \in E$ represents a causation from cross-sectional unit $i$ in period 0 to cross-sectional unit $j$, $G$ is a causal graph for $(Y_1, Y_0)$,

  $$Y_0 = (Y_{i,0})_{i \in N}, \text{ and } Y_1 = (Y_{i,1})_{i \in N}. $$
Diffusion Over a Network

Period 1

Period 2
Indian Villages and Microfinancing Programs

- From Banerjee et al. (2013)

The first period
Indian Villages and Microfinancing Programs

The next period
Questions

- How can we quantify the spillover effect of $Y_{0,j}$’s on $Y_{1,i}$’s along the graph $G$?
- How can we disentangle the impact of covariates from the true diffusion?
Example: Purchase of iPhones by Students

• Suppose that students’ friendship network is given.

• In the first period, we observe some students purchase iPhones, and in the next period, we observe other students purchase iPhones.

• Even if we see that iPhone purchase by students in the first period tend to be correlated with that of their friends in the next period, this wouldn’t immediately imply diffusion of iPhone purchases over the friendship network, if the network is formed based on the parents’ income of the students, and the parent’s income influence the purchase of iPhones by the students.

• The covariates may have “some” influence on the estimated diffusion, but how much? Can we quantify the influence?
He and Song (2018)

- This paper introduces a measure of diffusion of outcomes over a large recorded graph over two periods as a weighted average treatment effect.
- This paper develops inference on the diffusion parameter and establishes its asymptotic validity.
This Paper

- We explore the notion of a set of covariates being \textit{irrelevant} in explaining the diffusion of state-changes over a large network.
- This paper modifies the measure of diffusion in He and Song (2018) and decomposes it into one explained by the covariates and the other unexplained by the covariates.
- The role of the covariates is measured by the magnitude of the first component in the decomposition of diffusion.
- This paper provides asymptotic inference on this component.
A Brief Literature Review

- **VAR**: long time series for a small set of variables with unknown linear causal relations.

- **SAR**: linear relations. simultaneity. Cross-sectional dependence may not be patterned after the adjacency matrix.

- **Markov Random Fields, Bayesian Networks**: a graph as a model of causal relations: Lauritzen (1996), Pearl (2009), etc. The literature on statistical inference for a large network appears thin.


- **Permutation-Based Inference**: Chung and Romano (2013), Canay, Romano and Shaikh (2014) and Song (2018).

A Dynamic Causal Graph

A Causal Graph of Four Vertices
Conditional Independence and a Causal Graph

- $N(i) = \{j \in N : ij \in E\}$: the in-neighborhood of node $i$.
- $N(A) \equiv \bigcup_{i \in A} N(i)$, and $d(i) = |N(i)|$.
- $\overline{N}(A) \equiv N(A) \cup A$.
- Let $W_1 = (W_{i,1})_{i \in N}$ and $W_0 = (W_{i,0})_{i \in N}$ be two arrays of random vectors.

**Definition:** We say that $(W_1, W_0)$ has graph $G$ as a *dynamic causal graph* conditional on a $\sigma$-field $\mathcal{G}$, if the following conditions hold.

(i) $\{W_{j,0}\}_{j \in N}$ is conditionally independent (across $j$’s) given $\mathcal{G}$.

(ii) For any $A, B, A', B' \subset N$ such that $\overline{N}(A) \cap \overline{N}(A') = \emptyset$, $B \cap \overline{N}(A') = \emptyset$, and $B' \cap \overline{N}(A) = \emptyset$, $(W_{A,1}, W_{B,0})$ and $(W_{A',1}, W_{B',0})$ are conditionally independent given $\mathcal{G}$.

(cf. Lauritzen (1996), Pearl (2000), Lee and Song (2018), He and Song (2018).)
Causal Modeling of Diffusion

- For each \( i \in N \), let \( \{j_1, ..., j_{d(i)}\} \) be its in-neighbors in \( G \), where \( d(i) = |N(i)| \), i.e., the in-degree of vertex \( i \).

- **Potential outcome** for unit \( i \) in period 1 when its previous period outcome is \( d_i \), and its in-neighbors’ previous period outcomes \( d_{j_1}, ..., d_{j_{d(i)}} \):

\[
Y_{i,1}^* (d_i, d_{j_1}, ..., d_{j_{d(i)}}) = \varphi_i (d_i, d_{j_1}, ..., d_{j_{d(i)}}; u_i),
\]

where \( \varphi_i (\cdot, ..., \cdot; u_i) : \{0, 1\}^{d(i)+1} \to \mathbb{R} \) is a given map, and \( u_i \) unobserved heterogeneity.

- **Observed outcome** \( Y_{i,1} = Y_{i,1}^* (Y_{i,0}, Y_{j_1,0}, ..., Y_{j_{d(i)}+1,0}) \).

**Assumption: (Dynamic Causal Graph):** \( (u, Y_0) \) has \( G \) as a dynamic causal graph conditional on \( X, G \), where \( u = (u_i)_{i \in N} \).
**Diffusion as a Causal Parameter**

- For each $i \in N$ and its in-neighbor $j$, we define $Y^*_{ij}(d)$ to be the same as $Y_{i,1}$ except that we replace $Y_{j,0}$ in the argument by $d$.
- Define the measure of diffusion as

$$D \equiv \sum_{i \in N} \frac{1}{|N(i)|} \sum_{j \in N(i)} \mathbb{E}[Y^*_{ij}(1) - Y^*_{ij}(0)|X, G]w_j,$$

where $w_j \geq 0$ is a weight given to $j$ such that $\sum_{j \in N} w_j = 1$. (cf. He and Song (2018))
- We choose

$$w_j \equiv \frac{\mu_{j,0}(1 - \mu_{j,0})}{\sum_{\ell \in N} \mu_{\ell,0}(1 - \mu_{\ell,0})},$$

where $\mu_{j,0} \equiv \mathbb{E}_G[Y_{j,0}]$. (cf. Crump, Hotz, Imbens and Mitnik (2006))
Diffusion as a Causal Parameter

- $D$ represents a weighted average of the spillover effects over $j \in N$, and is the parameter of interest in this paper.

- $D$ is defined to be conditional on $(X, G)$. The advantage of this is that it enables us to perform inference without specifying the cross-sectional dependence structure of $X_i$’s, and the network formation process.

- When $G$ is formed based on $X$ (e.g. homophily), the cross-sectional dependence of $X$ given $G$ depends on the details of the way the network formation process is modeled.
Measuring Spatio-Temporal Dependence

- Define

\[ A_{i,0} = \frac{1}{|N(i)|} \sum_{j \in N(i)} Y_{j,0}. \]

- Essentially, our spatio-temporal dependence measure is based on the covariance between \( A_{i,0} \) and \( Y_{i,1} \).

- **Measuring the Role of Covariates:** Conditional Covariance and Unconditional Covariance:

\[ C_1 = \mathbb{E}[(A - \mathbb{E}A)(Y - \mathbb{E}Y)] \text{ vs.} \]
\[ C_2 = \mathbb{E}[(A - \mathbb{E}[A|X])(Y - \mathbb{E}[Y|X])|X], \]

- The role of covariates is captured by \( H \equiv C_1 - C_2 \).

- **Problem:** Using the conditional probabilities will require us to model the cross-sectional dependence ordering among the covariates again.
Measuring Spatio-Temporal Dependence

- **“Unconditional Mean”:** Instead of $E[Y]$, we use

$$\frac{1}{n} \sum_{i=1}^{n} E[Y_i \mid X_i].$$

- $X_i \in \mathbb{R}^p$ and $S = \{1, 2, \ldots, p\}$. For each $S \subset S$, we define

$$\mu_{i,0,S}^A = E[A_{i,0} \mid X_S, G],$$

where $X_S = (X_{i,S})_{i \in N}$.

- When $S = S$, we simply write

$$\mu_{i,0}^A = \mu_{i,0,S}^A.$$

- We also define $\mu_{i,0,-S}^A$ where we use $X_{-S}$ instead of $X_S$. 
Let $R(i)$ be i.i.d. following a uniform distribution on $\{1, 2, \ldots, n\}$.

We define

$$C_S \equiv \begin{cases} 
\frac{\text{Cov}_F(A_{R(i),0} - \mu_{A_{R(i),0},-S}, Y_{R(i),1})}{\nu^2} & \text{if } S \neq S, \text{ and} \\
\frac{\text{Cov}_F(A_{R(i),0}, Y_{R(i),1})}{\nu^2} & \text{if } S = S,
\end{cases}$$

where

$$\nu^2 = \frac{1}{n} \sum_{j \in N} \text{Var}_F(Y_{j,0}).$$
Decomposition of Spatio-Temporal Dependence

- Note that

\[ C_\emptyset = \frac{1}{nv^2} \sum_{i \in N} E_{TF} [(A_{i,0} - \mu_{i,0}^A)(Y_{i,1} - E_{TF}[Y_1])] . \]

Therefore, \( C_\emptyset \) is a spatio-temporal dependence measure of \((Y_0, Y_1)\) that can be estimated without omitting any covariates.

- For each \( S \subset S \), we decompose:

\[ C_S = C_\emptyset + H_S, \]  \hspace{1cm} (3)

where

\[ H_S \equiv \begin{cases} 
\frac{\text{Cov}_{TF}(\mu_{R(i),0}^A - \mu_{R(i),0}^A, -S, Y_{R(i),1})}{v^2} & \text{if } S \neq S, \text{ and} \\
\frac{\text{Cov}_{TF}(\mu_{R(i),0}^A, Y_{R(i),1})}{v^2} & \text{if } S = S 
\end{cases} \]

- \( H_S \) captures the role of the contribution of covariates \( X_{i,S} \) to the measured spatio-temporal dependence \( C_S \).
Identifying $D$ as $C_S$

Assumption ($S$-unconfoundedness): For each $i \in N$ and $j \in \bar{N}(i)$, $(Y_{ij}^*(1), Y_{ij}^*(0), X_{i,S})$ is conditionally independent of $Y_{j,0}$ given $(X_{-S}, G)$

- Under $S$-unconfoundedness, the covariate $X_{i,S}$ is irrelevant in identifying the causal effect $D$.

Lemma: Suppose that $S$-unconfoundedness holds for some $S \subset S$. Then, for all $S' \subset S$,

$$D = C_{S'}.$$ 

Furthermore, in this case, we have

$$H_{S'} = 0.$$
Spurious Diffusion and $H_S$

- $H_S$ captures bias caused to $D$ as measured by $C_S$ when covariates $X_{i,S}$ are omitted.
- If $H_S \neq 0$, the $S$-unconfoundedness is violated, and using $C_S$ as a measure of diffusion $D$ is dubious.
Cross-Sectional Dependence of $X_i$ and Spurious Diffusion

- Cross-sectional dependence of $X_i$ can cause bias to the measured diffusion even when there is no diffusion in truth.

**Lemma:** Suppose that $E[Y_{j,0}|X,G] = E[Y_{j,0}|X_j,G]$, and

$$\phi_i(d_i, d_{j_1}, ..., d_{j_{d(i)}}; u_i) = \tilde{\phi}_i(d_i; u_i),$$

for some map $\tilde{\phi}_i$. Suppose further that $u_i$ is conditionally independent of $X_{-i}$ given $(X_i,G)$, where $X_{-i} = (X_j)_{j \in N:i \neq j}$.

Then for all $S \subset S$, $E[H_S|G] \neq 0$ implies that $\{X_j\}$ is conditionally cross-sectionally dependent across $j$'s given $G$. 
Estimation and Inference on $\hat{H}_S$

**Assumption:** For all $i, j \in N$, and for all $x$,

$$E_{\mathcal{F}}[Y_{i,0}|X_i = x, G] = E_{\mathcal{F}}[Y_{j,0}|X_j = x, G].$$

(4)

- We first estimate $\mu_{i,0,S}^A$ and obtain $\hat{\mu}_{i,0,S}^A$.
- We construct

$$\hat{H}_S = \frac{1}{n\hat{\vartheta}^2} \sum_{i \in N} (\hat{\mu}_{i,0}^A - \hat{\mu}_{i,0,-S}^A)(Y_{i,1} - \bar{Y}_1),$$

where

$$\hat{\vartheta}^2 = \frac{1}{n} \sum_{j \in N} (Y_{j,0} - \hat{\mu}_{j,0})^2.$$
Asymptotic Linear Representation

**Assumption (Nondegeneracy):** There exist small $c > 0$ and large $M > 0$ such that the following is satisfied for all $n \geq 1$:

(i) $\sigma^2 > c$ and $\text{Var}(\frac{1}{\sqrt{n}} \sum_{i=1}^{n} q_i) > c$.

(ii) Both $\frac{1}{n} \sum_{i \in N} |N(i)|$ and $\frac{1}{n} \sum_{i \in N} |N_O(i)|$ are bounded away from zero by $c$.

**Assumption:** Graph $G$ has the maximum degree bounded by $M$ for all $n \geq 1$.

**Assumption:** $\max_{i \in N} |\hat{\mu}_{i,1} - \mu_{i,1}| = O_P(\omega_n)$ for some $\omega_n \to 0$ such that $\sqrt{n} \omega_n^2 \to 0$ as $n \to \infty$. 
Asymptotic Linear Representation

Define

\[ q_{i,S} = \Delta_{i,S}^e (\varepsilon_{i,1} - \tilde{\varepsilon}_1) + \Delta_{i,S}^\mu \mathbb{E}_F [Y_{i,1} - \bar{Y}_1] - (\varepsilon_{i,0} - \mathbb{E}_F [\varepsilon_{i,0}^2]) H_S, \]

where \( \varepsilon_{i,1} = Y_{i,1} - \mu_{i,1} \), and \( \tilde{\varepsilon}_1 = \frac{1}{n} \sum_{i \in N} \varepsilon_{i,1} \), and

\[ \Delta_{i,S}^\mu = \mu_{i,0} - \mu_{i,0,-s} - \frac{1}{n} \sum_{\ell \in N} (\mu_{\ell,0} - \mu_{\ell,0,-s}), \]

\[ \Delta_{i,S}^e = e_{i,0} - e_{i,0,-s} - \frac{1}{n} \sum_{\ell \in N} (e_{\ell,0} - e_{\ell,0,-s}), \]

and

\[ e_{j,0,-s} = \frac{\sum_{k \in N : k \neq j} \varepsilon_{k,0} 1\{ X_{k,-s} = X_{j,-s} \}}{\sum_{k \in N : k \neq j} 1\{ X_{k,-s} = X_{j,-s} \}} , \text{ and } e_{i,0,-s} = \frac{1}{|N(i)|} \sum_{j \in N(i)} e_{j,0}. \]
Asymptotic Linear Representation

**Lemma:** For each $S \subset \$,

$$\sqrt{n}(\hat{H}_{S} - H_{S}) = \frac{1}{v^2 \sqrt{n}} \sum_{i \in \mathcal{N}} q_{i,S} + o_{P}(1).$$

- For the construction of the variance estimator, we let $G^* = (N, E^*)$ be the undirected graph such that $ij \in E^*$ if and only if $\overline{N}(i) \cap \overline{N}(j) \neq \emptyset$.

- Then it is not hard to see that under the dynamic causal graph assumption, $q_{i,S}$’s have $G^*$ as a conditional dependency graph given $(X, G)$.

- Let $\hat{q}_{i,S}$ be a consistent estimator of $q_{i,S}$. We construct

$$\hat{\sigma}^2_{\hat{S}} = \frac{1}{n \hat{\sigma}^4} \sum_{i,j \in \mathcal{N}: \overline{N}(i) \cap \overline{N}(j) \neq \emptyset} \hat{q}_{i,S} \hat{q}_{j,S},$$

(5)
Asymptotic Confidence Intervals

- The asymptotic confidence intervals for $H_S$ are constructed by

$$C_{1-\alpha, S, \infty} = \left[ \hat{H}_S - \frac{c_{1-\alpha/2}\hat{\sigma}_S}{\sqrt{n}}, \hat{H}_S + \frac{c_{1-\alpha/2}\hat{\sigma}_S}{\sqrt{n}} \right].$$

where $c_{1-\alpha/2}$ denotes the $1 - \alpha/2$ percentile of $N(0, 1)$.

**Theorem**

*For each $S \subseteq S,*

$$\liminf_{n \to \infty} P \{ H_S \in C_{1-\alpha, S, \infty} \} \geq 1 - \alpha.$$
Monte Carlo Simulations

- We generate $X_i = [X_{i,1}, X_{i,2}]$, binary covariates, permitting $X_{i,1}$ to be cross-sectionally dependent, and $X_{i,2}$ i.i.d.

- As for outcomes, we set

$$Y_{j,0} = 1 \left\{ \frac{\exp(c_0 + X_j'\gamma_0)}{1 + \exp(c_0 + X_j'\gamma_0)} \geq u_{j,0} \right\},$$

where $u_{j,0}$’s are drawn from the uniform distribution on $[0, 1]$ and we set $\gamma_{0,1} = 1$ and $\gamma_{0,2} = 1$.

- We define

$$Y_{i,1} = 1\{c_0 + \delta_0 \bar{Y}_{i,0} + X_i'\beta_0 - u_{i,1} > 0\},$$

where $u_{i,1}$’s are drawn i.i.d. from $N(0, 1)$, and

$$\bar{Y}_{i,0} = \frac{1}{d(i)} \sum_{j \in N(i)} Y_{j,0}.$$
Graph Summary Statistics

The Degree Characteristics of the Graphs

<table>
<thead>
<tr>
<th></th>
<th>( m = 1 )</th>
<th>( m = 3 )</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>max. deg</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( n = 500 )</td>
<td>38</td>
<td>64</td>
</tr>
<tr>
<td>( n = 1000 )</td>
<td>34</td>
<td>83</td>
</tr>
<tr>
<td><strong>ave. deg</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( n = 500 )</td>
<td>1.964</td>
<td>5.796</td>
</tr>
<tr>
<td>( n = 1000 )</td>
<td>1.983</td>
<td>5.898</td>
</tr>
</tbody>
</table>
Results

The Empirical Coverage Probabilities (95% Nominal Level)

<table>
<thead>
<tr>
<th></th>
<th>$m = 1$</th>
<th></th>
<th>$m = 3$</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>$n$</td>
<td>$\delta_0 = 0$</td>
<td>$\delta_0 = 1$</td>
<td>$\delta_0 = 0$</td>
<td>$\delta_0 = 1$</td>
</tr>
<tr>
<td>500</td>
<td>0.9480</td>
<td>0.8830</td>
<td>0.9500</td>
<td>0.9290</td>
</tr>
<tr>
<td>1000</td>
<td>0.9430</td>
<td>0.9070</td>
<td>0.9540</td>
<td>0.9550</td>
</tr>
</tbody>
</table>

The Mean-Length of Confidence Intervals

<table>
<thead>
<tr>
<th></th>
<th>$m = 1$</th>
<th></th>
<th>$m = 3$</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>$n$</td>
<td>$\delta_0 = 0$</td>
<td>$\delta_0 = 1$</td>
<td>$\delta_0 = 0$</td>
<td>$\delta_0 = 1$</td>
</tr>
<tr>
<td>500</td>
<td>0.0715</td>
<td>0.0879</td>
<td>0.0310</td>
<td>0.0326</td>
</tr>
<tr>
<td>1000</td>
<td>0.0462</td>
<td>0.0639</td>
<td>0.0191</td>
<td>0.0202</td>
</tr>
</tbody>
</table>
Remaining Works

- It is expected that permutation-based confidence intervals, if we can construct them, will perform better than asymptotic inference. (cf. Song (2018))

- To our best knowledge, it is not possible to construct permutation-based confidence intervals on $H_S$ without knowing the dependence ordering of covariates $X_i$’s.

- However, we can construct permutation testing of $S$-unconfoundedness condition, because in this case, we can impose the null restriction in deriving asymptotic validity.