Just How Hopeless Is Trying to Learn about Contagion from Social Network Data?

Cosma Shalizi

Statistics Department, Carnegie Mellon University

Santa Fe Institute

31 July 2010, SciFoo

My real interest: how much of the mechanism of a complex system can we reconstruct from observations?

My real interest: how much of the mechanism of a complex system can we reconstruct from observations?
As Yet Another Ex-Physicist, social networks are just large coupled dynamical assemblage. . .

My real interest: how much of the mechanism of a complex system can we reconstruct from observations?
As Yet Another Ex-Physicist, social networks are just large coupled dynamical assemblage...
Got interested in this problem with Sperber (1996) and especially Christakis and Fowler (2007)

My real interest: how much of the mechanism of a complex system can we reconstruct from observations?

As Yet Another Ex-Physicist, social networks are just large coupled dynamical assemblage. . .

Got interested in this problem with Sperber (1996) and especially Christakis and Fowler (2007)

. . . and expected to reach much more positive conclusions

My real interest: how much of the mechanism of a complex system can we reconstruct from observations?

As Yet Another Ex-Physicist, social networks are just large coupled dynamical assemblage. . .

Got interested in this problem with Sperber (1996) and especially Christakis and Fowler (2007)

. . . and expected to reach much more positive conclusions

Apologies in advance for social-scientific naivete

My real interest: how much of the mechanism of a complex system can we reconstruct from observations?

As Yet Another Ex-Physicist, social networks are just large coupled dynamical assemblage...

Got interested in this problem with Sperber (1996) and especially Christakis and Fowler (2007)

...and expected to reach much more positive conclusions

Apologies in advance for social-scientific naivete

Joint work with Andrew Thomas

Details: `http://arxiv.org/abs/1004.4704`

“If your friend Joey jumped off a bridge, would you jump too?”

“If your friend Joey jumped off a bridge, would you jump too?”

1. yes: Joey inspires you (social contagion or influence)

"If your friend Joey jumped off a bridge, would you jump too?"

1. yes: Joey inspires you (social contagion or influence)
2. yes: Joey infects you with a parasite which suppresses fear of falling (biological contagion)

"If your friend Joey jumped off a bridge, would you jump too?"

1. yes: Joey inspires you (social contagion or influence)
2. yes: Joey infects you with a parasite which suppresses fear of falling (biological contagion)
3. yes: you're friends *because* you both like to jump off bridges (manifest homophily)

"If your friend Joey jumped off a bridge, would you jump too?"

1. yes: Joey inspires you (social contagion or influence)
2. yes: Joey infects you with a parasite which suppresses fear of falling (biological contagion)
3. yes: you're friends *because* you both like to jump off bridges (manifest homophily)
4. yes: you're friends *because* you both like roller-coasters, and have a common risk-seeking propensity (latent homophily)

“If your friend Joey jumped off a bridge, would you jump too?”

1. yes: Joey inspires you (social contagion or influence)
2. yes: Joey infects you with a parasite which suppresses fear of falling (biological contagion)
3. yes: you’re friends *because* you both like to jump off bridges (manifest homophily)
4. yes: you’re friends *because* you both like roller-coasters, and have a common risk-seeking propensity (latent homophily)
5. yes: because sometimes jumping off a bridge is the only sane thing to do (external causation)

Wikipedia, s.v. "Tacoma Narrows Bridge (1940)"

Are these distinctions with *observational* differences?

Are these distinctions with *observational* differences?

1. Can't experiment by pushing Joey off the bridge

Are these distinctions with *observational* differences?

1. Can't experiment by pushing Joey off the bridge
2. Can't experiment by keeping Joey and Irene apart, or pushing them together

Contagion, Influence

Whether Irene does something is predicted by whether Irene's neighbors had already done it

Contagion, Influence

Whether Irene does something is predicted by whether Irene's neighbors had already done it

Contagion, Influence

Whether Irene does something is predicted by whether Irene's neighbors had already done it

- Diffusion of innovations
- Diffusion of ideologies

Contagion, Influence

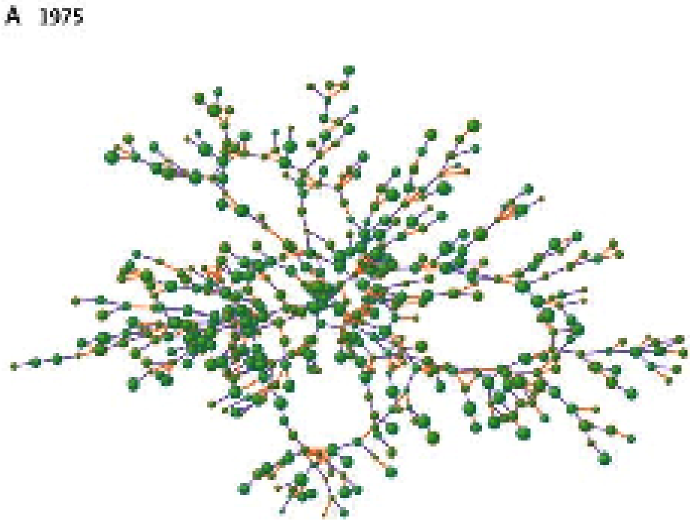
Whether Irene does something is predicted by whether Irene's neighbors had already done it

- Diffusion of innovations
- Diffusion of ideologies
- Infectious diseases

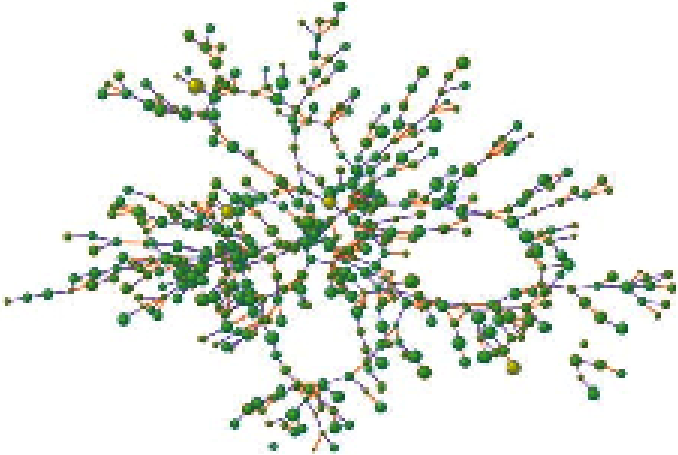
Contagion, Influence

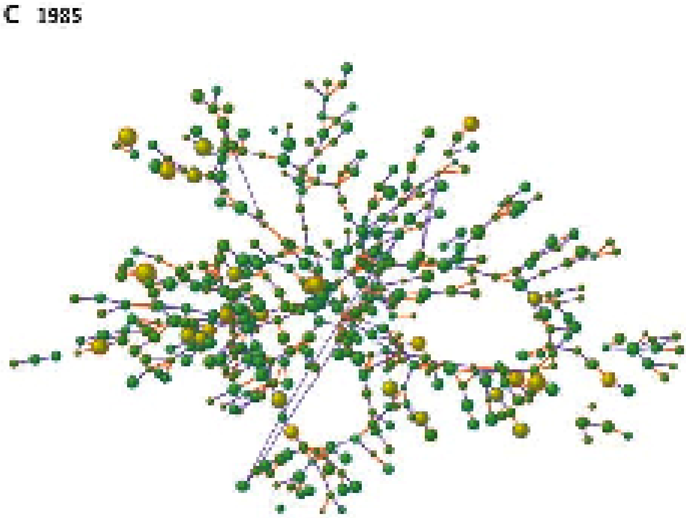
Whether Irene does something is predicted by whether Irene's neighbors had already done it

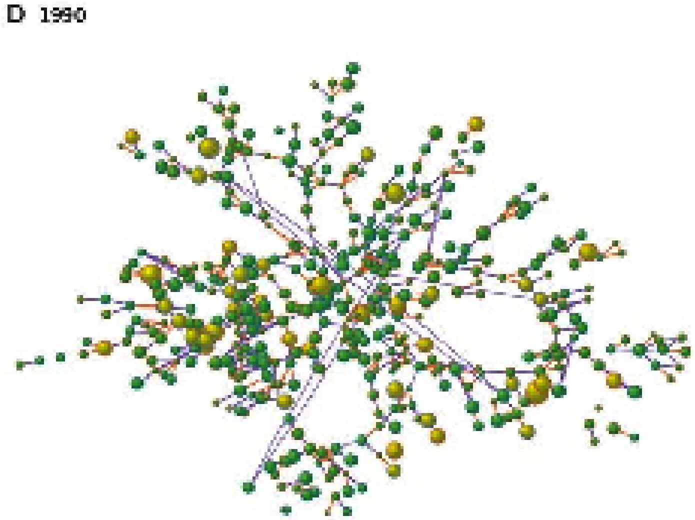
- Diffusion of innovations
- Diffusion of ideologies
- Infectious diseases
- Not-obviously-infectious conditions (e.g., obesity, loneliness, divorce) . . .

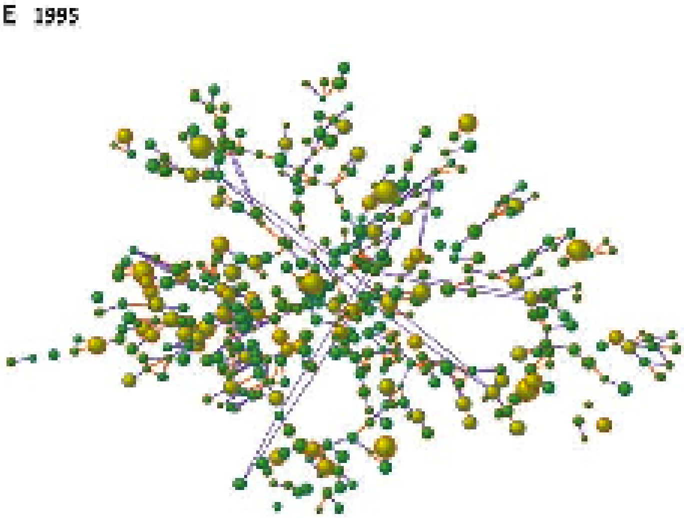

A 1975

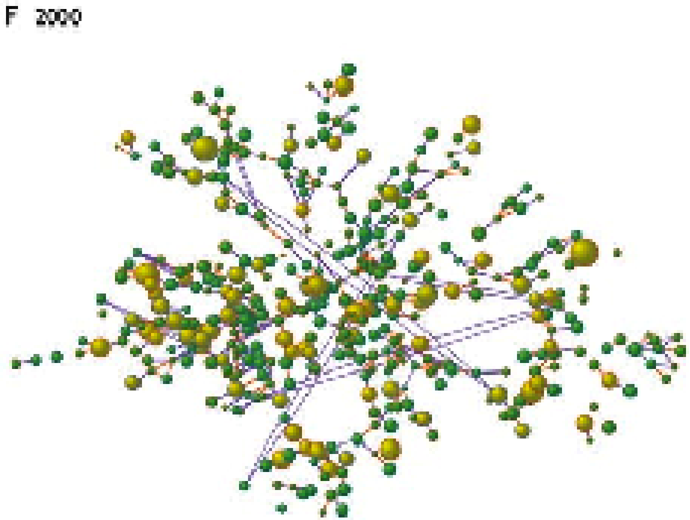
B 1980




C 1985


D 1990


E 1995


F 2000

Why Care?

- Social engineering by targeting influential people
- Public health, marketing, crime, propaganda, etc.

Lots of interest for electronic social networks

Why Care?

- Social engineering by targeting influential people
- Public health, marketing, crime, propaganda, etc.

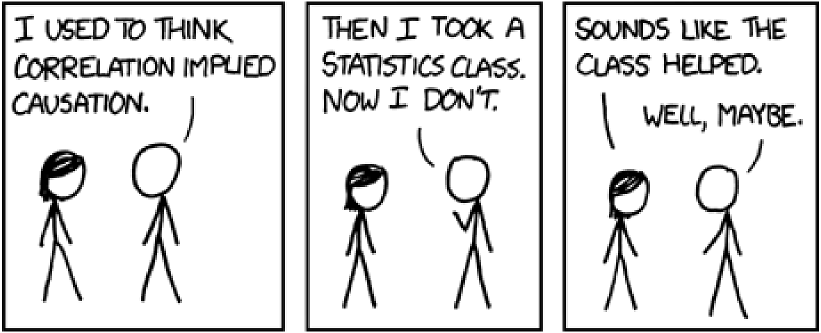
Lots of interest for electronic social networks
Can we actually figure out how influence/contagion there is?

Causal Inference

This is a causal inference question

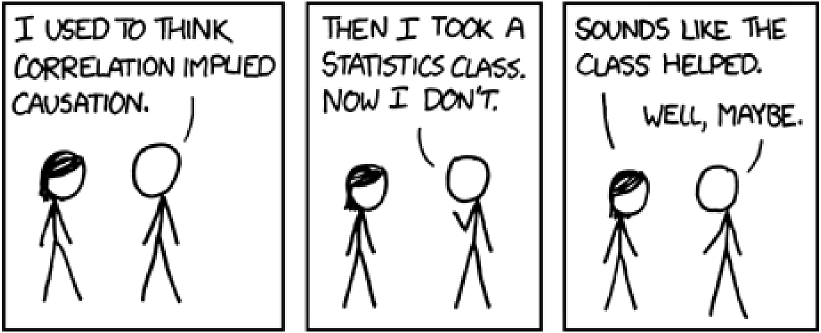
Causal Inference

This is a causal inference question



Causal Inference

This is a causal inference question



"Correlation doesn't imply causation, but it does waggle its eyebrows suggestively and gesture furtively while mouthing 'look over there'"

Looking Over There

Causal inference becomes a lot clearer once you start drawing pictures

Looking Over There

Causal inference becomes a lot clearer once you start drawing pictures
(Pearl, 2009; Morgan and Winship, 2007)

Looking Over There

Causal inference becomes a lot clearer once you start drawing pictures
(Pearl, 2009; Morgan and Winship, 2007)
Dots = variables, arrows = direct causal influence

Looking Over There

Causal inference becomes a lot clearer once you start drawing pictures
(Pearl, 2009; Morgan and Winship, 2007)
Dots = variables, arrows = direct causal influence
Do controls block off indirect paths between variables?

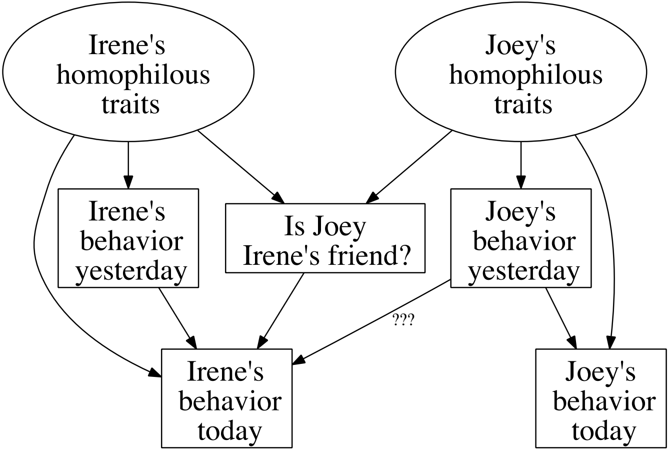
Looking Over There

Causal inference becomes a lot clearer once you start drawing pictures
(Pearl, 2009; Morgan and Winship, 2007)
Dots = variables, arrows = direct causal influence
Do controls block off indirect paths between variables?
Do controls *activate* indirect paths?

Looking Over There

Causal inference becomes a lot clearer once you start drawing pictures
(Pearl, 2009; Morgan and Winship, 2007)
Dots = variables, arrows = direct causal influence
Do controls block off indirect paths between variables?
Do controls *activate* indirect paths?

Separate question: what causal diagrams are compatible with the correlation pattern? (Spirtes *et al.*, 2001)


Irene's homophilous traits
Joey's homophilous traits
Irene's behavior yesterday
Is Joey Irene's friend?
Joey's behavior yesterday
???
Irene's behavior today
Joey's behavior today

Contagion Effects are Unidentifiable

1. Joey's behavior yesterday has information about Joey's traits

Contagion Effects are Unidentifiable

1. Joey's behavior yesterday has information about Joey's traits
2. Joey's traits have information about Irene's, *since* they are neighbors

Contagion Effects are Unidentifiable

1. Joey's behavior yesterday has information about Joey's traits
2. Joey's traits have information about Irene's, *since* they are neighbors
3. Irene's traits have information about Irene's behavior today

Contagion Effects are Unidentifiable

1. Joey's behavior yesterday has information about Joey's traits
2. Joey's traits have information about Irene's, *since* they are neighbors
3. Irene's traits have information about Irene's behavior today
4. $\therefore$ Joey's behavior yesterday predicts Irene's behavior today

Contagion Effects are Unidentifiable

1. Joey's behavior yesterday has information about Joey's traits
2. Joey's traits have information about Irene's, *since* they are neighbors
3. Irene's traits have information about Irene's behavior today
4. $\therefore$ Joey's behavior yesterday predicts Irene's behavior today *even if there is no direct causal effect*

Contagion Effects are Unidentifiable

1. Joey's behavior yesterday has information about Joey's traits
2. Joey's traits have information about Irene's, *since* they are neighbors
3. Irene's traits have information about Irene's behavior today
4. $\therefore$ Joey's behavior yesterday predicts Irene's behavior today *even if there is no direct causal effect*
5. $\therefore$ Homophily is confounded with contagion

Failed Escape Attempts

Adding covariates for Irene and Joey doesn't help

Failed Escape Attempts

Adding covariates for Irene and Joey doesn't help
Adding more time points doesn't help

Failed Escape Attempts

Adding covariates for Irene and Joey doesn't help
Adding more time points doesn't help
Letting social ties change over time *really* doesn't help

(Noel and Nyhan, 2010)

Possible Escapes

Really strong modeling assumptions *might* work (linearity doesn't seem strong enough)

Possible Escapes

Really strong modeling assumptions *might* work (linearity doesn't seem strong enough)

Observe and control for all the traits which affect friendship

Possible Escapes

Really strong modeling assumptions *might* work (linearity doesn't seem strong enough)

Observe and control for all the traits which affect friendship

Observe and control for all the traits which affect behavior

The Argument from Asymmetry

Focus on unreciprocated friendships

The Argument from Asymmetry

Focus on unreciprocated friendships

IRENE: *Joey is my friend*
JOEY: *Irene who?*

The Argument from Asymmetry

Focus on unreciprocated friendships

IRENE: *Joey is my friend*
JOEY: *Irene who?*

Suppose senders (Irene) are more predictable from receivers (Joey) than vice versa

The Argument from Asymmetry

Focus on unreciprocated friendships

> IRENE: *Joey is my friend*
> JOEY: *Irene who?*

Suppose senders (Irene) are more predictable from receivers (Joey) than vice versa
Doesn't this argue for direct influence?

The Argument from Asymmetry

Focus on unreciprocated friendships

IRENE: *Joey is my friend*
JOEY: *Irene who?*

Suppose senders (Irene) are more predictable from receivers (Joey) than vice versa
Doesn't this argue for direct influence?
Sounds plausible. . .

The Argument from Asymmetry

Focus on unreciprocated friendships

> IRENE: *Joey is my friend*
> JOEY: *Irene who?*

Suppose senders (Irene) are more predictable from receivers (Joey) than vice versa
Doesn't this argue for direct influence?
Sounds plausible. . .
. . . fails if senders and receivers have systematically different trait values

The Argument from Asymmetry

Focus on unreciprocated friendships

> IRENE: *Joey is my friend*
> JOEY: *Irene who?*

Suppose senders (Irene) are more predictable from receivers (Joey) than vice versa
Doesn't this argue for direct influence?
Sounds plausible. . .
. . . fails if senders and receivers have systematically different trait values

e.g., people similar friends but also like median friends

Making homophily and contagion look like causation

Long-term, hard-to-change social/economic status explains more short-term, malleable cultural / political / consumer variables

Making homophily and contagion look like causation

Long-term, hard-to-change social/economic status explains more short-term, malleable cultural / political / consumer variables

> *Gellner: "Social structure is who you can marry, culture is what you wear at the wedding."*

What’s the evidence?

What's the evidence?

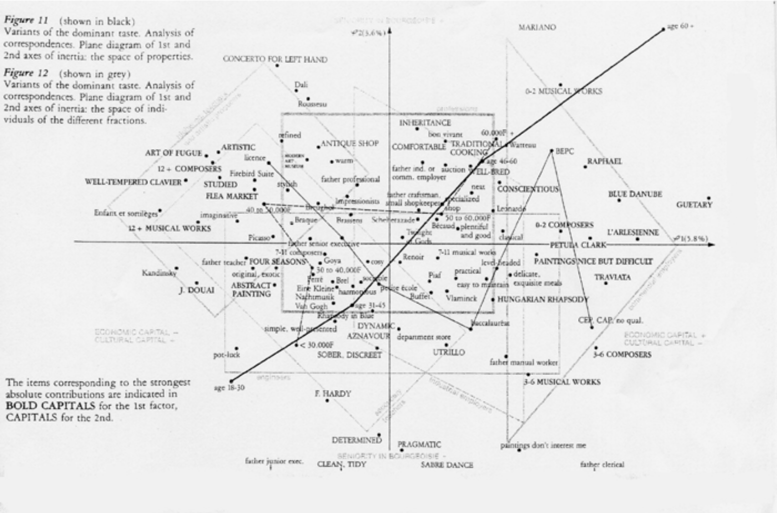
- The stories sound good

What's the evidence?

- The stories sound good
- Casual empiricism

What's the evidence?

- The stories sound good
- Casual empiricism
- Correlation/regression analyses; cultural choices are predictable from social statuses (e.g. Bourdieu (1984))

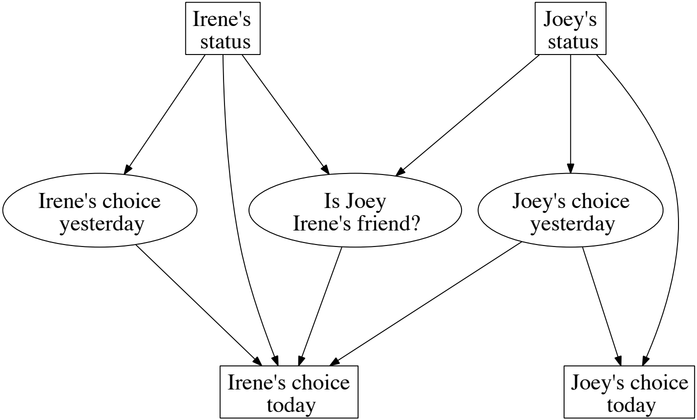


From Bourdieu (1984)

Probably true a lot of the time

Probably true a lot of the time

BUT usually ignores social networks and just looks at surveys


Irene's status
Joey's status
Irene's choice yesterday
Is Joey Irene's friend?
Joey's choice yesterday
Irene's choice today
Joey's choice today

More Confounding

Direct influence of Irene's status on Irene's choices is confounded with contagion:

More Confounding

Direct influence of Irene's status on Irene's choices is confounded with contagion:

1. Irene's status predicts who Irene's friends are

More Confounding

Direct influence of Irene's status on Irene's choices is confounded with contagion:

1. Irene's status predicts who Irene's friends are
2. $\therefore$ Irene's status predicts what Irene's friends chose yesterday

More Confounding

Direct influence of Irene's status on Irene's choices is confounded with contagion:

1. Irene's status predicts who Irene's friends are
2. $\therefore$ Irene's status predicts what Irene's friends chose yesterday
3. contagion: Joey's choice yesterday influences Irene's choice today if they are friends

More Confounding

Direct influence of Irene's status on Irene's choices is confounded with contagion:

1. Irene's status predicts who Irene's friends are
2. ∴ Irene's status predicts what Irene's friends chose yesterday
3. contagion: Joey's choice yesterday influences Irene's choice today if they are friends
4. ∴ Irene's status predicts Irene's choice, *even if* there is no direct influence

Responsible Just-So Story-telling

These accounts are usually adaptationist/functionalist
At the very least they are causal accounts
We should really check them
Biology suggests: a **neutral model**

Responsible Just-So Story-telling

These accounts are usually adaptationist/functionalist
At the very least they are causal accounts
We should really check them
Biology suggests: a **neutral model**

- Include all the evolutionary processes *except* adaptation

Responsible Just-So Story-telling

These accounts are usually adaptationist/functionalist
At the very least they are causal accounts
We should really check them
Biology suggests: a **neutral model**

- Include all the evolutionary processes *except* adaptation
- Work out expected behavior of this model

Responsible Just-So Story-telling

These accounts are usually adaptationist/functionalist
At the very least they are causal accounts
We should really check them
Biology suggests: a **neutral model**

- Include all the evolutionary processes *except* adaptation
- Work out expected behavior of this model
- Data departing from neutral model $\Rightarrow$ evidence of adapation

Caricature Neutral Model of Cultural Evolution

- Unchanging status for each node

Caricature Neutral Model of Cultural Evolution

- Unchanging status for each node
- Network is assortative on status (minimal departure from Erdős-Rényi)

Caricature Neutral Model of Cultural Evolution

- Unchanging status for each node
- Network is assortative on status (minimal departure from Erdős-Rényi)
- Binary choice for each node at each time

Caricature Neutral Model of Cultural Evolution

- Unchanging status for each node
- Network is assortative on status (minimal departure from Erdős-Rényi)
- Binary choice for each node at each time
- Start by tossing coins

Caricature Neutral Model of Cultural Evolution

- Unchanging status for each node
- Network is assortative on status (minimal departure from Erdős-Rényi)
- Binary choice for each node at each time
- Start by tossing coins
    1. At each time, pick a random node and a random neighbor

Caricature Neutral Model of Cultural Evolution

- Unchanging status for each node
- Network is assortative on status (minimal departure from Erdős-Rényi)
- Binary choice for each node at each time
- Start by tossing coins
  1. At each time, pick a random node and a random neighbor
  2. Copy the neighbor's choice

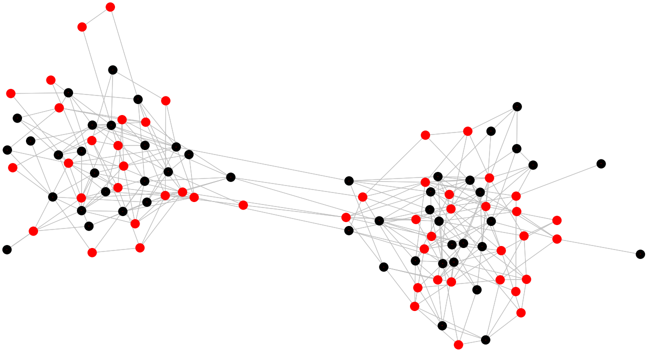
Caricature Neutral Model of Cultural Evolution

- Unchanging status for each node
- Network is assortative on status (minimal departure from Erdős-Rényi)
- Binary choice for each node at each time
- Start by tossing coins
  1. At each time, pick a random node and a random neighbor
  2. Copy the neighbor's choice
  3. Go to (1)

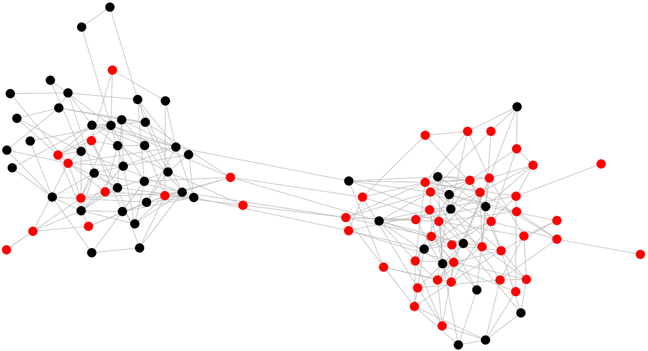
Caricature Neutral Model of Cultural Evolution

- Unchanging status for each node
- Network is assortative on status (minimal departure from Erdős-Rényi)
- Binary choice for each node at each time
- Start by tossing coins
  1. At each time, pick a random node and a random neighbor
  2. Copy the neighbor's choice
  3. Go to (1)

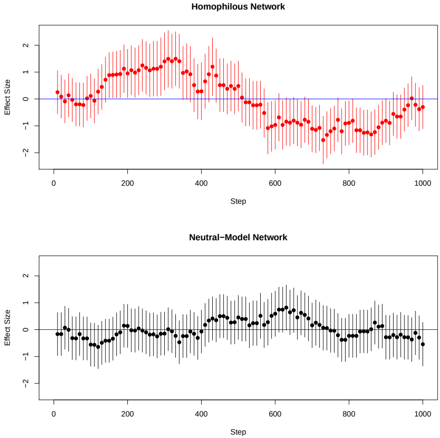
(= "voter model" of statistical mechanics)



100 node network, homophily for status (2 groups), initial choices



After 300 updates



Coefficients for logistic regression of choice on status, $\pm$ 95% confidence intervals.
Red, homophilous network; black, matched non-assortative network

Contagion + Homophily Looks Like Causation

- Neutral diffusion + homophily looks like a real connection between social status and cultural choices

Contagion + Homophily Looks Like Causation

- Neutral diffusion + homophily looks like a real connection between social status and cultural choices
- Problem is *not* the ecological fallacy (red-state/blue-state fallacy)

Contagion + Homophily Looks Like Causation

- Neutral diffusion + homophily looks like a real connection between social status and cultural choices
- Problem is *not* the ecological fallacy (red-state/blue-state fallacy)
- Choices are still dependent after controlling for status

Contagion + Homophily Looks Like Causation

- Neutral diffusion + homophily looks like a real connection between social status and cultural choices
- Problem is *not* the ecological fallacy (red-state/blue-state fallacy)
- Choices are still dependent after controlling for status
- Need to control for neighbors' *previous choices*

What To Do?

How can we go forward with studying contagion when there is homophily?

- Experiment: on choice/behavior or social ties or both

What To Do?

How can we go forward with studying contagion when there is homophily?

- Experiment: on choice/behavior or social ties or both
  this is Science, and is Hard

What To Do?

How can we go forward with studying contagion when there is homophily?

- Experiment: on choice/behavior or social ties or both
  this is Science, and is Hard
- No problem if we control for the homophilous traits

What To Do?

How can we go forward with studying contagion when there is homophily?

- Experiment: on choice/behavior or social ties or both
  this is Science, and is Hard
- No problem if we control for the homophilous traits
  $\therefore$ figure out what they are and measure them

What To Do?

How can we go forward with studying contagion when there is homophily?

- Experiment: on choice/behavior or social ties or both
  this is Science, and is Hard
- No problem if we control for the homophilous traits
  $\therefore$ figure out what they are and measure them
  this is Science, and is Hard

What To Do?

How can we go forward with studying contagion when there is homophily?

- Experiment: on choice/behavior or social ties or both
  this is Science, and is Hard
- No problem if we control for the homophilous traits
  $\therefore$ figure out what they are and measure them
  this is Science, and is Hard
- Bounds: even if we can't point-identify, maybe we can pin down a range

What To Do?

How can we go forward with studying contagion when there is homophily?

- Experiment: on choice/behavior or social ties or both
  this is Science, and is Hard
- No problem if we control for the homophilous traits
  $\therefore$ figure out what they are and measure them
  this is Science, and is Hard
- Bounds: even if we can't point-identify, maybe we can pin down a range
- Clustering: figure out the traits from the social network

Bounds and Partial Identification

Unidentifiable parameter $\equiv$ multiple values of the parameter yield the *same* observational distribution

Bounds and Partial Identification

Unidentifiable parameter $\equiv$ multiple values of the parameter yield the *same* observational distribution
$\therefore$ even infinite data does not pin down the parameter

Bounds and Partial Identification

Unidentifiable parameter $\equiv$ multiple values of the parameter yield the *same* observational distribution
$\therefore$ even infinite data does not pin down the parameter
Partial identification (Manski, 2007): range of parameter values yielding one distribution might be limited

Bounds and Partial Identification

Unidentifiable parameter $\equiv$ multiple values of the parameter yield the *same* observational distribution
$\therefore$ even infinite data does not pin down the parameter
Partial identification (Manski, 2007): range of parameter values yielding one distribution might be limited
$\therefore$ infinite data *bounds* the parameter

Bounds and Partial Identification

Unidentifiable parameter $\equiv$ multiple values of the parameter yield the *same* observational distribution
$\therefore$ even infinite data does not pin down the parameter
Partial identification (Manski, 2007): range of parameter values yielding one distribution might be limited
$\therefore$ infinite data *bounds* the parameter
Could we bound contagion effects when there is homophily?

Bounds and Partial Identification

Unidentifiable parameter $\equiv$ multiple values of the parameter yield the *same* observational distribution
$\therefore$ even infinite data does not pin down the parameter
Partial identification (Manski, 2007): range of parameter values yielding one distribution might be limited
$\therefore$ infinite data *bounds* the parameter
Could we bound contagion effects when there is homophily?
Work in progress, only negative results so far…

Partial Control by Clustering?

Could we work out the trait from the network?

Partial Control by Clustering?

Could we work out the trait from the network?
Homophily $\Rightarrow$ you tend to resemble your neighbors

Partial Control by Clustering?

Could we work out the trait from the network?
Homophily $\Rightarrow$ you tend to resemble your neighbors
$\Rightarrow$ Especially likely if you all have lots of neighbors in common who all have lots of neighbors in common, etc.

Partial Control by Clustering?

Could we work out the trait from the network?
Homophily $\Rightarrow$ you tend to resemble your neighbors
$\Rightarrow$ Especially likely if you all have lots of neighbors in common who all have lots of neighbors in common, etc.
$\Rightarrow$ Modules/communities

Partial Control by Clustering?

Could we work out the trait from the network?
Homophily $\Rightarrow$ you tend to resemble your neighbors
$\Rightarrow$ Especially likely if you all have lots of neighbors in common who all have lots of neighbors in common, etc.
$\Rightarrow$ Modules/communities
Use community membership as a proxy for the traits

Partial Control by Clustering?

Could we work out the trait from the network?
Homophily $\Rightarrow$ you tend to resemble your neighbors
$\Rightarrow$ Especially likely if you all have lots of neighbors in common who all have lots of neighbors in common, etc.
$\Rightarrow$ Modules/communities
Use community membership as a proxy for the traits
Can't remove confounding

Partial Control by Clustering?

Could we work out the trait from the network?
Homophily $\Rightarrow$ you tend to resemble your neighbors
$\Rightarrow$ Especially likely if you all have lots of neighbors in common who all have lots of neighbors in common, etc.
$\Rightarrow$ Modules/communities
Use community membership as a proxy for the traits
Can't remove confounding

unless estimated community assignments are a sufficient statistic for the traits

Partial Control by Clustering?

Could we work out the trait from the network?
Homophily ⇒ you tend to resemble your neighbors
⇒ Especially likely if you all have lots of neighbors in common who all have lots of neighbors in common, etc.
⇒ Modules/communities
Use community membership as a proxy for the traits
Can't remove confounding
unless estimated community assignments are a sufficient statistic for the traits
but *should* reduce it, tighten bounds

Partial Control by Clustering?

Could we work out the trait from the network?
Homophily $\Rightarrow$ you tend to resemble your neighbors
$\Rightarrow$ Especially likely if you all have lots of neighbors in common who all have lots of neighbors in common, etc.
$\Rightarrow$ Modules/communities
Use community membership as a proxy for the traits
Can't remove confounding
unless estimated community assignments are a sufficient statistic for the traits
but *should* reduce it, tighten bounds
. . . or make it worse if the relationship isn't simple homophily

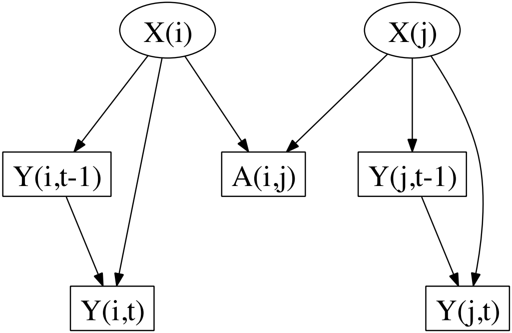
Conclusion

1. Homophily + causal influence looks like contagion
2. Homophily + contagion looks like causal influence
3. Need scientific knowledge and/or blind faith in assumptions
   Technical trickery won't do (alas)
4. *May* be possible to *limit* confounding

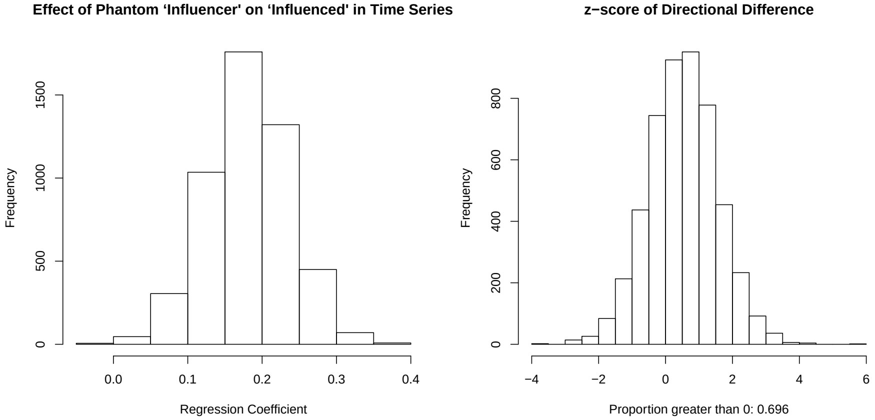
Toy Example

Try to predict $Y(i,t)$ from $Y(j,t)$ and vice versa when
$A_{ij} = 1, A_{ji} = 0$
$X(i) \sim \mathcal{U}(0,1)$
Edges form with probability $\propto \mathrm{logit}^{-1}(-3|X(i) - X(j)|)$
i nominates j from among neighbors, $\propto \mathrm{logit}^{-1}(-|X(j) - 0.5|)$

$$
\begin{aligned}
Y(i,0) &= (X(i) - 0.5)^3 + \mathcal{N}(0,(0.02)^2) \\
Y(i,1) &= Y(i,0) + 0.3X_i + \mathcal{N}(0,(0.02)^2)
\end{aligned}
$$



Causal graph of the model with no contagion, but asymmetry in regression coefficients


Effect of Phantom ‘Influencer' on ‘Influenced' in Time Series
Frequency
0
500
1000
1500
0.0
0.1
0.2
0.3
0.4
Regression Coefficient
z−score of Directional Difference
Frequency
0
200
400
600
800
−4
−2
0
2
4
6
Proportion greater than 0: 0.696

Results:

- $Y(i, 1)$ is well-predicted from $Y(j, 0)$
- *Nominees* are disproportionately in the middle; $i \rightarrow j, j \not\rightarrow i$ suggests i is more peripheral
- For asymmetric pairs, regression of sender on receiver differs from that of receiver on sender

An Analogy For Community Control

Gene association studies: does having this genetic variant influence this trait/change this risk?

Real populations are structured

Sub-populations differ (due to reproductive isolation etc.)

⇒ genes are correlated

⇒ random biases and inflated vvariances (vs. usual formulas)

⇒ many bogus results

Population structure substantial even for e.g. Germany (Steffens *et al.*, 2006) or Italy, never mind “white Americans”

Responses: (1) pedigrees; (2) “genomic control” by estimating over-dispersion empirically (Devlin *et al.*, 2001); (3) clustering — the diffusion maps in Lee *et al.* (2009) look *a lot* like Newman (2006)

Bourdieu, Pierre (1984). *Distinction: A Social Critique of the Judgement of Taste*. Cambridge, Massachusetts: Harvard University Press.

Christakis, Nicholas A. and James H. Fowler (2007). "The Spread of Obesity in a Large Social Network over 32 Years." *The New England Journal of Medicine*, **357**: 370–379. URL `http://content.nejm.org/cgi/content/abstract/357/4/370`.

Devlin, B., Kathryn Roeder and Larry Wasserman (2001). "Genomic Control, a New Approach to Genetic-Based Association Studies." *Theoretical Population Biology*, **60**: 155–166. URL `http://www.stat.cmu.edu/tr/tr749/tr749.html`. doi:10.1006/tpbi.2001.1542.

Lee, Ann B., Diana Luca, Lambertus Klei, Bernie Devlin and

Kathryn Roeder (2009). "Discovering Genetic Ancestry Using Spectral Graph Theory." *Genetic Epidemiology*, **34**: 51–59. URL `http://www.stat.cmu.edu/~roeder/pubs/llkdr2009.pdf`. doi:10.1002/gepi.20434.

Manski, Charles F. (1993). "Identification of Endogeneous Social Effects: The Reflection Problem." *Review of Economic Studies*, **60**: 531–542. URL `http://www.jstor.org/pss/2298123`.

— (2007). *Identification for Prediction and Decision*. Cambridge, Massachusetts: Harvard University Press.

Morgan, Stephen L. and Christopher Winship (2007). *Counterfactuals and Causal Inference: Methods and Principles for Social Research*. Cambridge, England: Cambridge University Press.

Newman, Mark E. J. (2006). "Finding Community Structure in

Networks Using the Eigenvectors of Matrices." *Physical Review E*, **74**: 036104. URL http://arxiv.org/abs/physics/0605087.

Noel, Hans and Brendan Nyhan (2010). "The "Unfriending" Problem The Consequences of Homophily in Friendship Retention for Causal Estimates of Social Influence." Online preprint. URL http://www-personal.umich.edu/~bnyhan/unfriending.pdf.

Pearl, Judea (2009). *Causality: Models, Reasoning, and Inference*. Cambridge, England: Cambridge University Press, 2nd edn.

Sperber, Dan (1996). *Explaining Culture: A Naturalistic Approach*. Oxford: Basil Blackwell.

Spirtes, Peter, Clark Glymour and Richard Scheines (2001).

Causation, Prediction, and Search. Cambridge, Massachusetts: MIT Press, 2nd edn.

Steffens, Michael, Claudia Lamina, Thomas Illig, Thomas Bettecken, Rainer Vogler, Patricia Entz, Eun-Kyung Suk, Mohammad Reza Toliat, Norman Klopp, Amke Caliebe, Inke R. König, Karola Köhler, Jan Lüdemann, Amalia Diaz Lacava, Rolf Fimmers, Peter Lichtner, Andreas Ziegler, Andreas Wolf, Michael Krawczak, Peter Nürnberg, Jochen Hampe, Stefan Schreiber, Thomas Meitinger, H.-Erich Wichmann, Kathryn Roeder, Thomas F. Wienker and Max P. Baur (2006). "SNP-Based Analysis of Genetic Substructure in the German Population." *Human Heredity*, **62**: 20–29. doi:10.1159/000095850.