Romney Didn't Win Hearts In Last Night's Debate, According To Twitter

Julie Bort | Oct. 4, 2012, 12:03 PM | 8,223 | 72

- Romney had 2.1 million mentions, compared to 1.6 million for Obama. Volumes peaked during the live debate, with Romney getting almost double Obama's mentions (approximately 1.1 million to 600,000).
- Negative sentiment towards Romney far outweighed the positive. Obama had more positive sentiment.

Romney

Obama

- Negative remarks about Romney centered on the perception that he was rude (20.6%) and that he "promised to cut help" (10.2%), an apparent reference to his views on social programs.
- The positive stuff said about Obama included "right choice" (18%) and "best president" (8.7%).
- The negative stuff said about Obama included "lose debate" (30.1%) and "nervous" (7.6%).

- Almost half of the positive comments about Romney used terms like "win debate" (47.6%). People also liked his hair (9%).

The phrase “Big Bird” was appearing 17,000 times every minute on Twitter. At midnight, CNN reported that mentions of Big Bird on Facebook were up an astronomical 800,000%.

Facebook later said Big Bird was the fourth most-mentioned topic on Facebook during the debate, getting more attention than topics like jobs, taxes, Jim Lehrer and Obamacare.

Some review…

http://www.colbertnation.com/the-colbert-report-videos/260955/january-07-2010/james-fowler

A question

- Homophily: similar nodes ~= connected nodes
- Which is cause and which is effect?
  - Do birds of a feather flock together? **(Associative sorting)**
  - Do you change your behavior based on the behavior of your peers? **(Social contagion)**

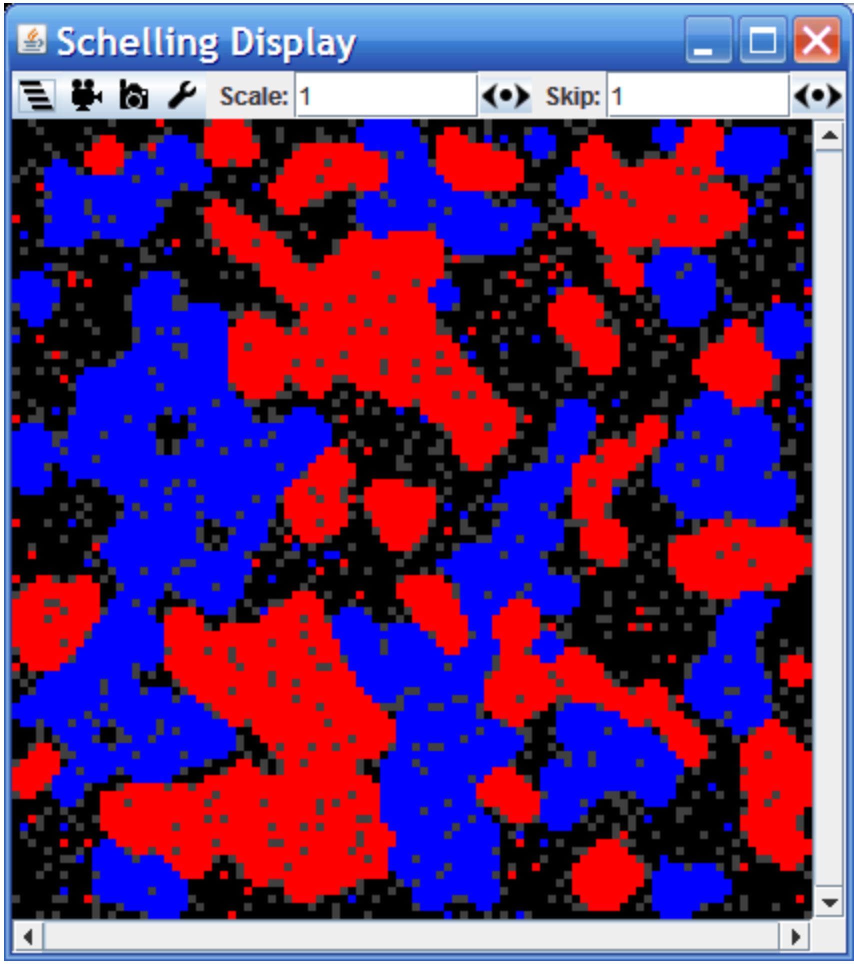
  - Note: Some authors use "homophily" only for associative sorting, some use it for observed correlation between attributes and connectivity.

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

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

Associative sorting example

- Network:
  - 2D grid, each point connected to immediate neighbors, each point has color (red or blue)
- Evolution: at each time t, each node will
  - Count colors of its neighbors
  - Move to a new (random) if it has <k neighbors of the same color
- Typical result: strong spatial segregation, even with weak preferences



k=3, Pr(red)=Pr(blue)=0.3

Social Contagion Example

- Lots of different reasons behavior might spread
  - Fads, cascades, …
- One reason: rational decisions made about products that have a “network effect”
    - I.e., the benefits and costs of the behavior are **not completely local** to the decision-maker
- Example: PowerPoint, …
- How can we analyze this?
  - From Easley & Kleinberg’s text, ch 16-17
  - We’ll go into this more later on….

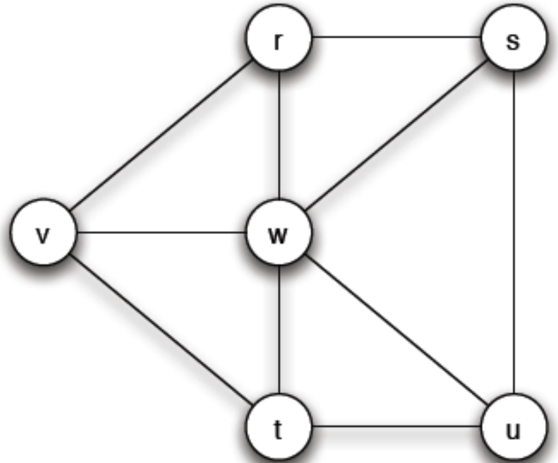
- if v and w both adopt behavior A, they each get a payoff of $a > 0$;
- if they both adopt B, they each get a payoff of $b > 0$; and
- if they adopt opposite behaviors, they each get a payoff of 0.

		w: A	w: B
v	A	a, a	$0, 0$
	B	$0, 0$	b, b

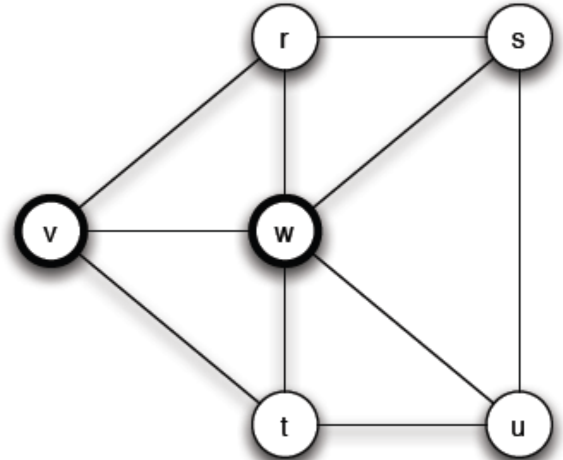
Figure 19.1: A-B Coordination Game

What if *v* is playing the game with many *w*' s ?

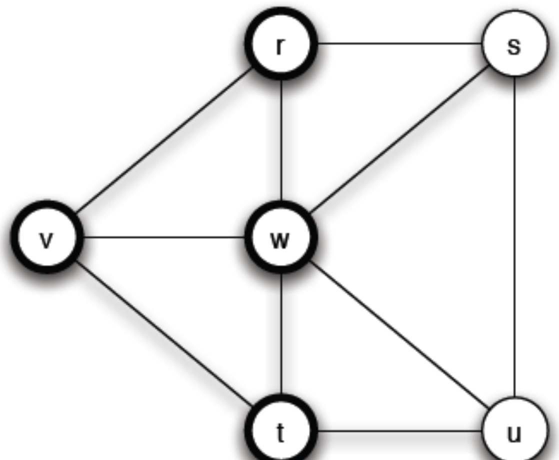
If *v* has *d* neighbors and *p*d of them* choose *A*, then *v* should chose *A* iff **pda>-(1-p)db** ie, iff **p>=b/(a+b)**



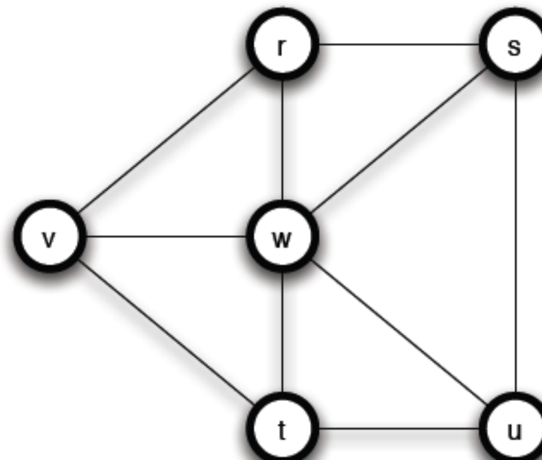
(a) *The underlying network*



(b) *Two nodes are the initial adopters*



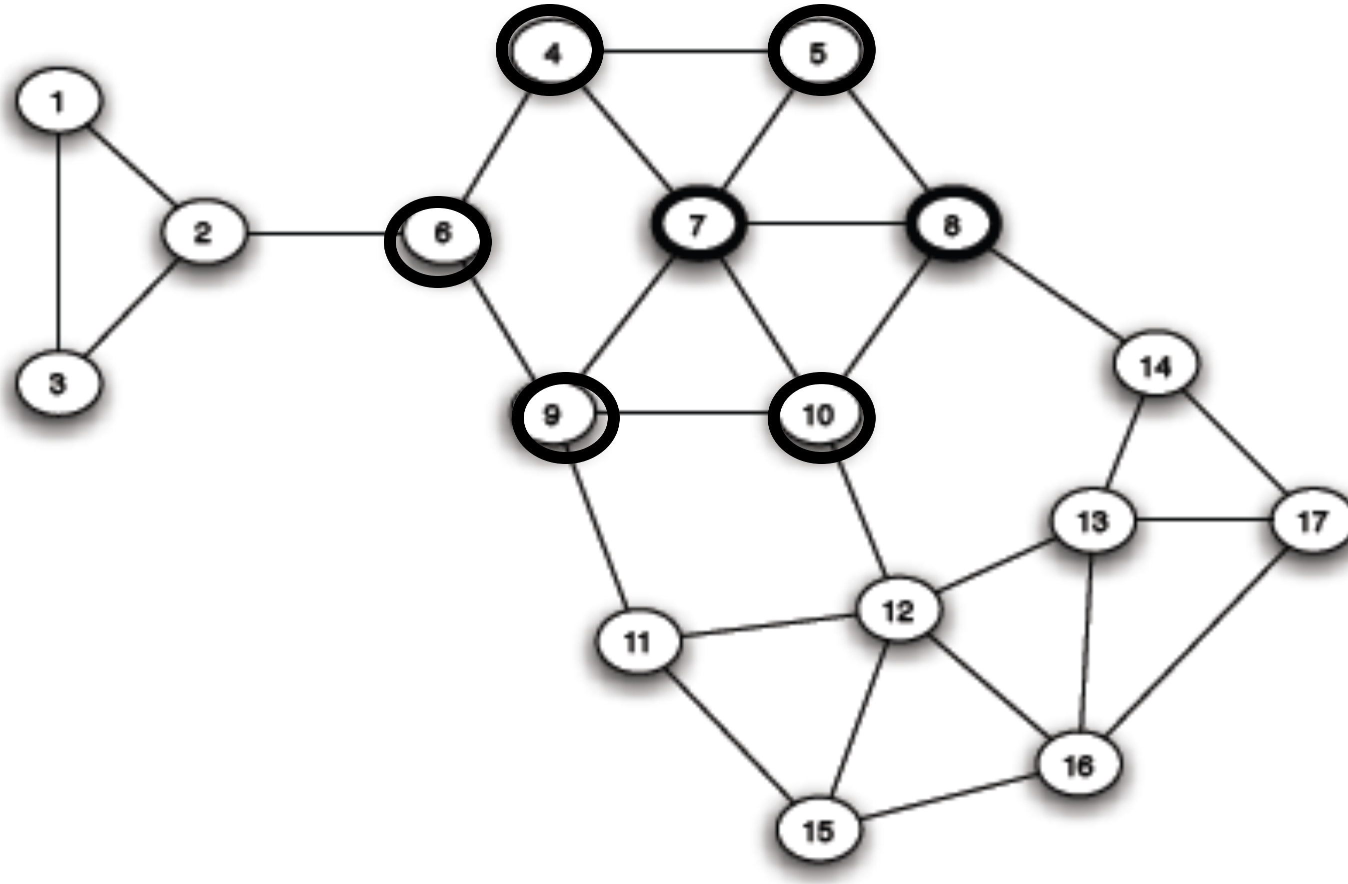
(c) *After one step, two more nodes have adopted*



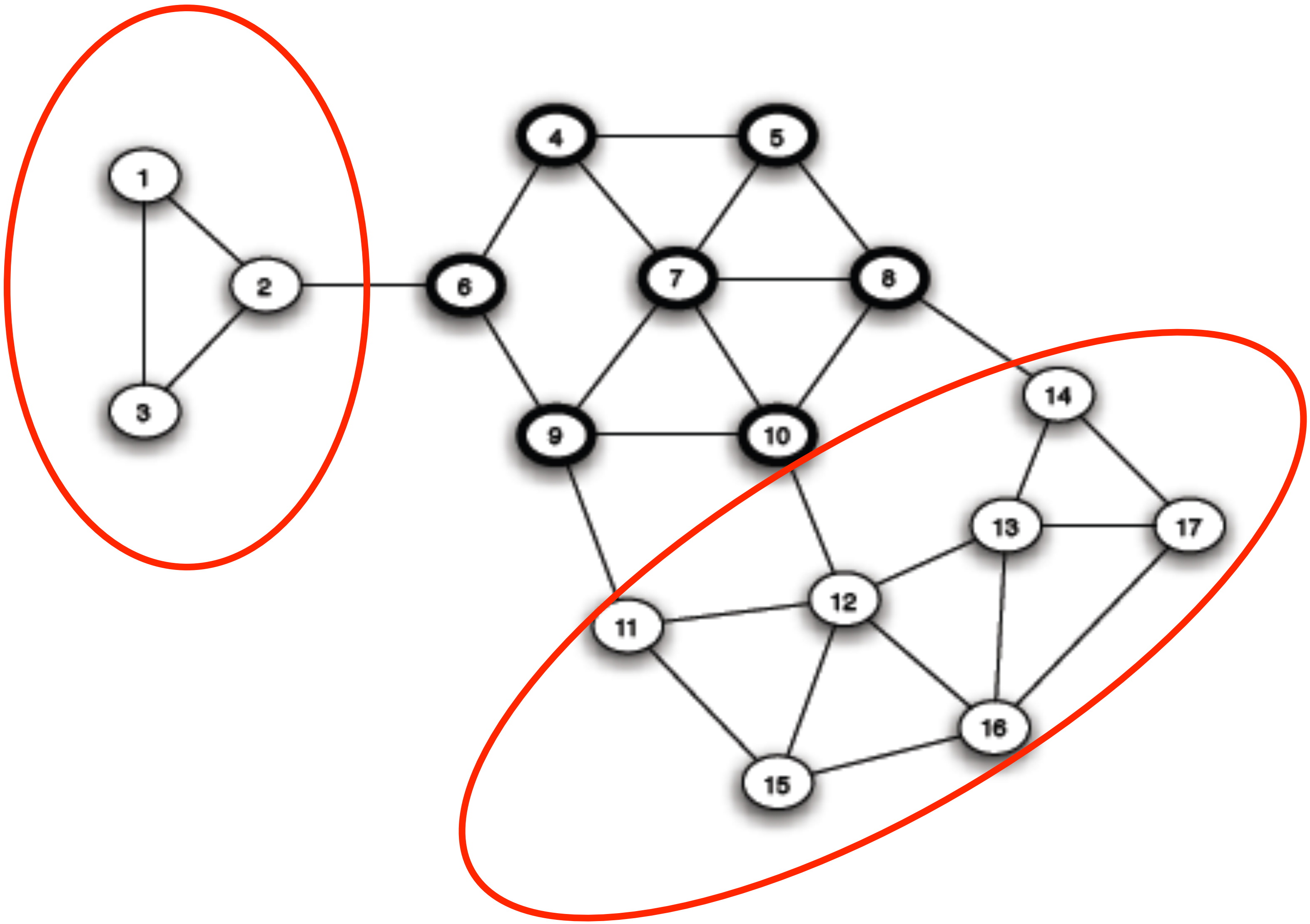
(d) *After a second step, everyone has adopted*

Threshold: switch if 40% of neighbors switched

Figure 19.3: Starting with v and w as the initial adopters, and payoffs $a = 3$ and $b = 2$, the new behavior A spreads to all nodes in two steps. Nodes adopting A in a given step are drawn with dark borders; nodes adopting B are drawn with light borders.



Threshold: switch if 40% of neighbors switched



General claim: dense clusters are less susceptible to cascades.

Thinking it through

1. Close-knit communities can halt a cascade of adoptions
   – Claim: a "complete cascade" happens iff there are no sufficiently close-knit clusters
2. A *small increase* in *a/(a+b)* might cause a big additional cascade.
3. Where the cascade *starts* might cause a big difference in the size of the cascade.
4. Marketing to specific individuals (e.g., in the middle of a cluster) might cause a cascade.

Thinking it Through

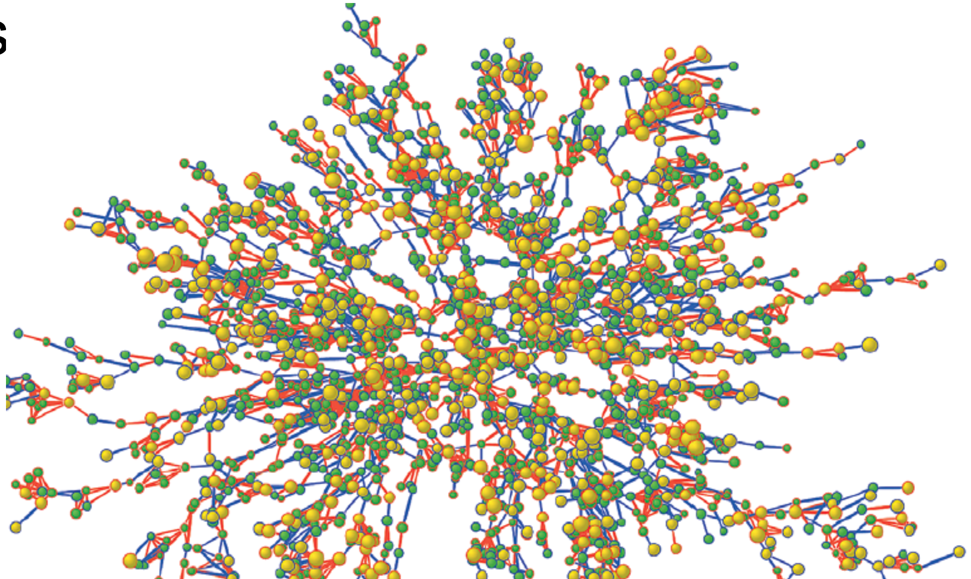
- You cane extend this to cover other situations, e.g., backward compatibility:

		w: A	w: B
v	A	a, a	a-ε,b
	B	$0, 0$	b, b

Figure 19.1: A-B Coordination Game

A complicated example

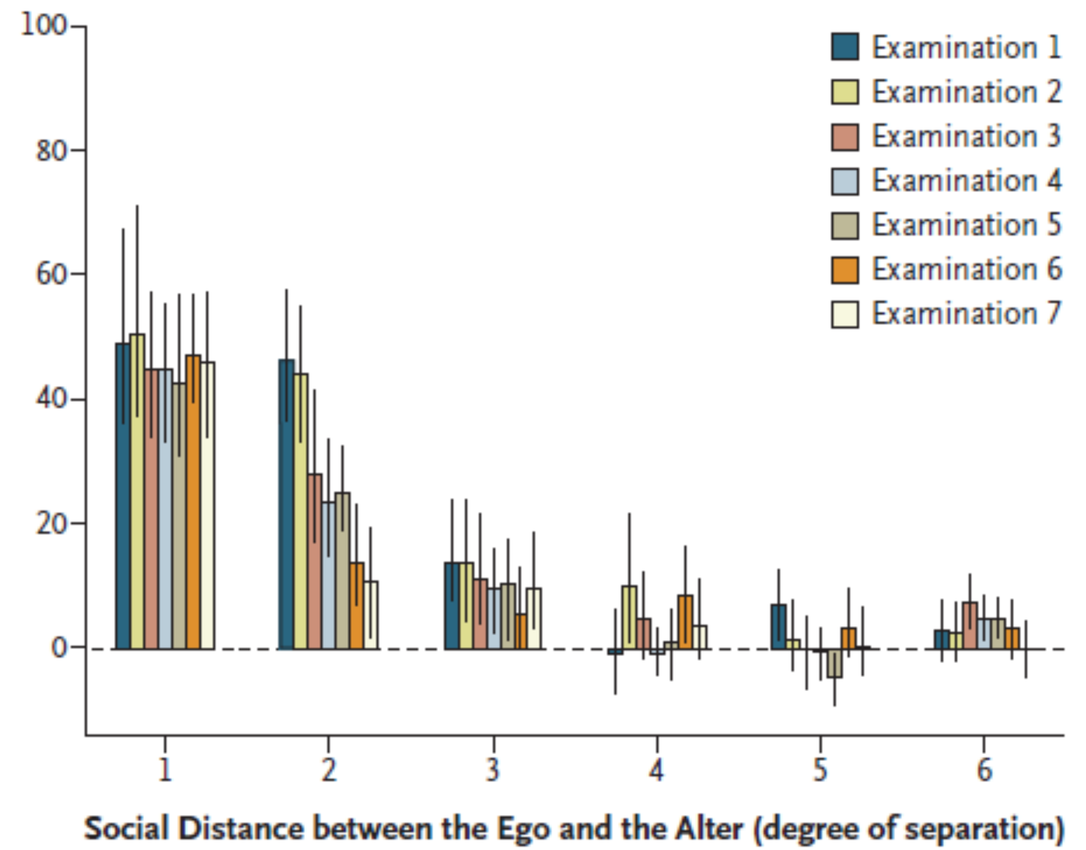
- NEJM, Christakis & Folwer, 2007: Spread of Obesity in A Large Social Network over 32 Years
- Statistical model: for x connected to w:
  - obesity(x,t) = F(age(x), sex(x), …, obesity(x,t-1),obesity(w,t-1))
- Linear regression model, so you can determine influence of a particular variable
- Looked at *asymmetric* links



A complicated example

Figure 3. Effect of Social and Geographic Distance from Obese Alters on the Probability of an Ego's Obesity in the Social Network of the Framingham Heart Study.

Panel A shows the mean effect of an ego's social proximity to an obese alter; this effect is derived by comparing the conditional probability of obesity in the observed network with the probability of obesity in identical networks (with topology preserved) in which the same number of obese persons is randomly distributed. The social distance between the alter and the ego is represented by degrees of separation (1 denotes one degree of separation from the ego, 2 denotes two degrees of separation from the ego, and so forth). The examination took place at seven time points. Panel B shows the mean effect of an ego's geographic proximity to an obese alter. We ranked all geographic distances (derived from geocoding) between the homes of directly connected egos and alters (i.e., those pairs at one degree of separation) and created six groups of equal size. This figure shows the effects observed for the six mileage groups (based on their average distance): 1 denotes 0 miles (i.e., closest to the alter's home), 2 denotes 0.26 mile, 3 denotes 1.5 miles, 4 denotes 3.4 miles, 5 denotes 9.3 miles, and 6 denotes 471 miles (i.e., farthest from the alter's home). There is no trend in geographic distance. I bars for both panels show 95% confidence intervals based on 1000 simulations. To convert miles to kilometers, multiply by 1.6.



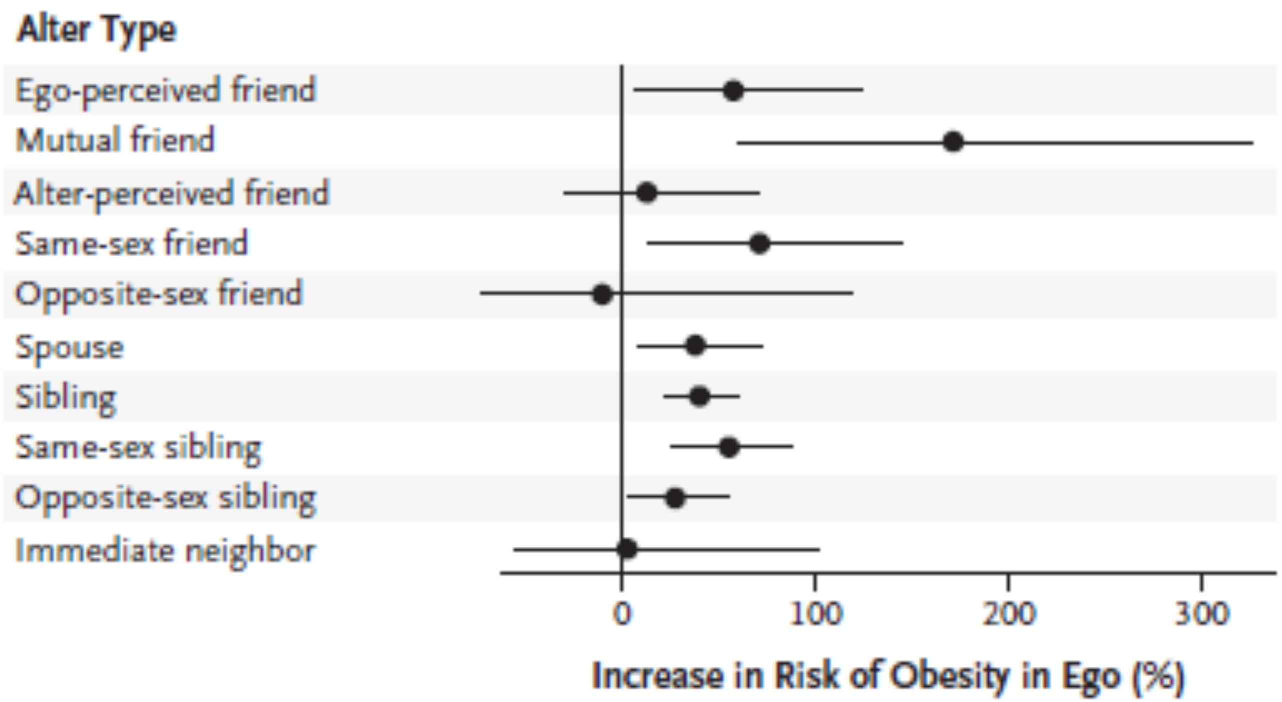
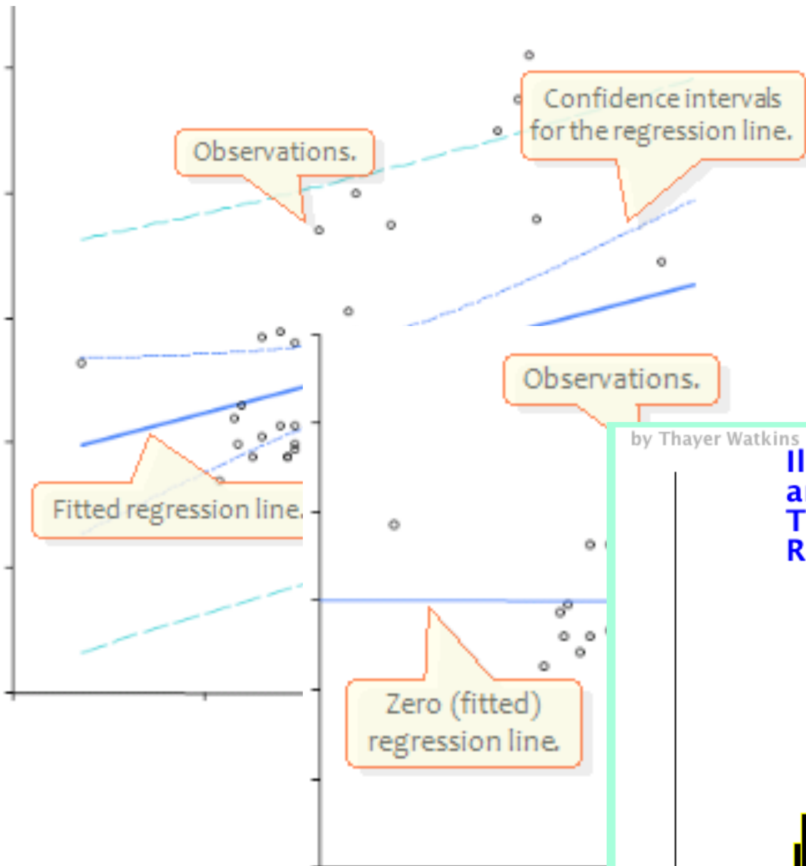


Figure 4. Probability That an Ego Will Become Obese According to the Type of Relationship with an Alter Who May Become Obese in Several Subgroups of the Social Network of the Framingham Heart Study.

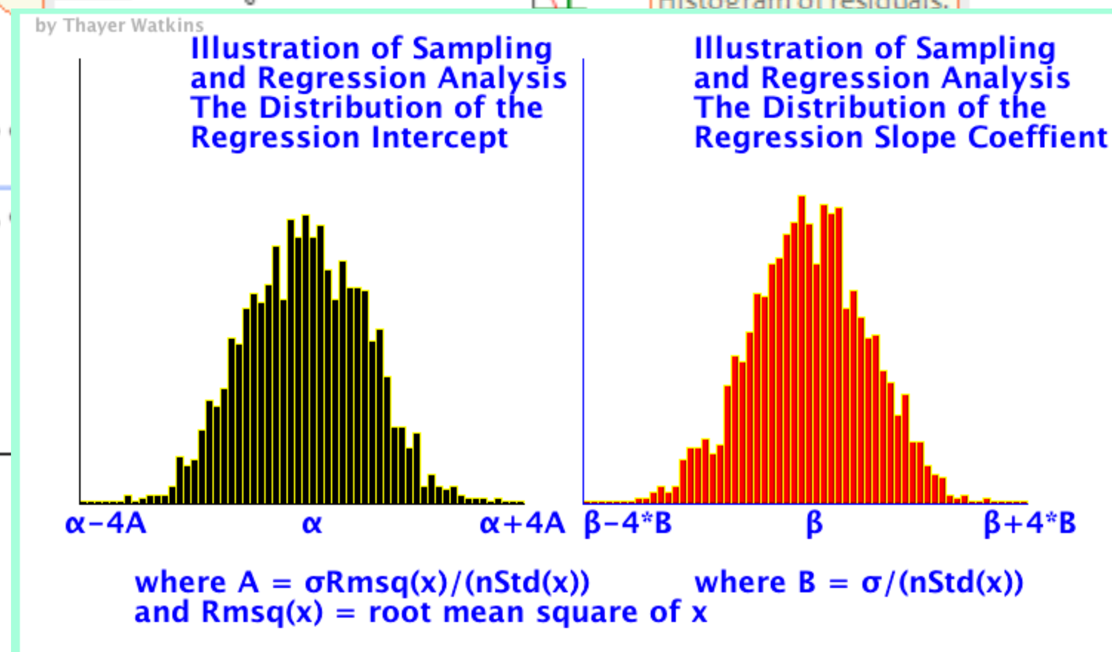
The closeness of friendship is relevant to the spread of obesity. Persons in closer, mutual friendships have more of an effect on each other than persons in other types of friendships. The dependent variable in each model is the obesity of the ego. Independent variables include a time-lagged measurement of the ego's obesity; the obesity of the alter; a time-lagged measurement of the alter's obesity; the ego's age, sex, and level of education; and indicator variables (fixed effects) for each examination. Full models and

Aside: linear regression



If true model for y is linear in x1, …, xn plus Gaussian noise then

- regression coefficients are normally distributed
- you can test to see if the influence of x is "real"



Another example

- NEJM, Christakis & Fowler, 2007: Spread of Obesity in A Large Social Network over 32 Years
- Statistical model: for x connected to w:
  - obesity(x,t) = F(age(x), sex(x), …, obesity(x,t-1),obesity(w,t-1))
  - "Granger causality"
- Linear regression model, so you can determine influence of a particular variable
  - But you're tied to a parametric model and it's assumptions
- Looked at *asymmetric* links
  - Seems like a clever idea but … what's the principle here?

Homophily, Contagion, Confounding: Pick Any Three

Cosma Shalizi

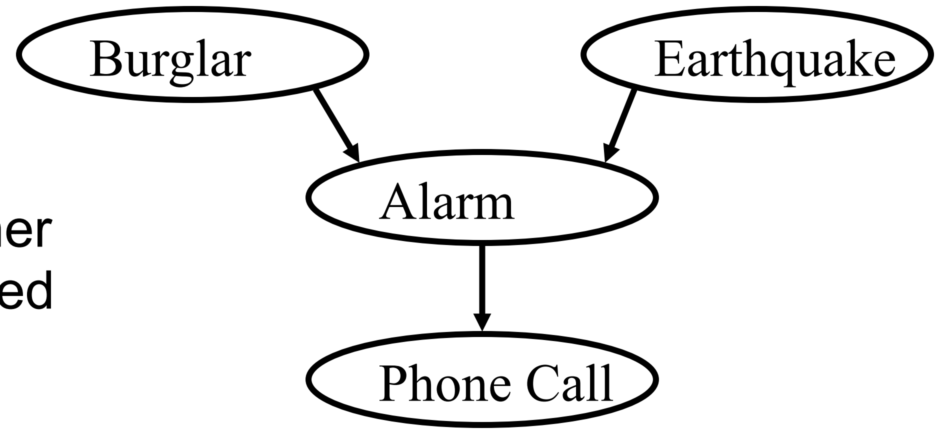
Statistics Department, Carnegie Mellon University

Santa Fe Institute

11 December 2009

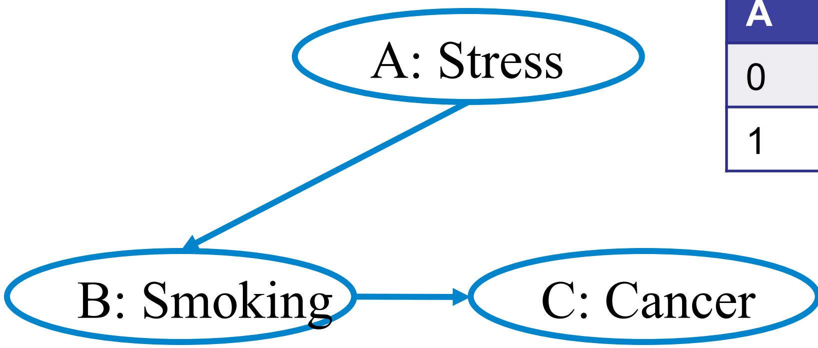
The “Burglar Alarm” example

- Your house has a twitchy burglar alarm that is also sometimes triggered by earthquakes.
- Earth arguably doesn’t care whether your house is currently being burgled
- While you are on vacation, one of your neighbors calls and tells you your home’s burglar alarm is ringing. Uh oh!

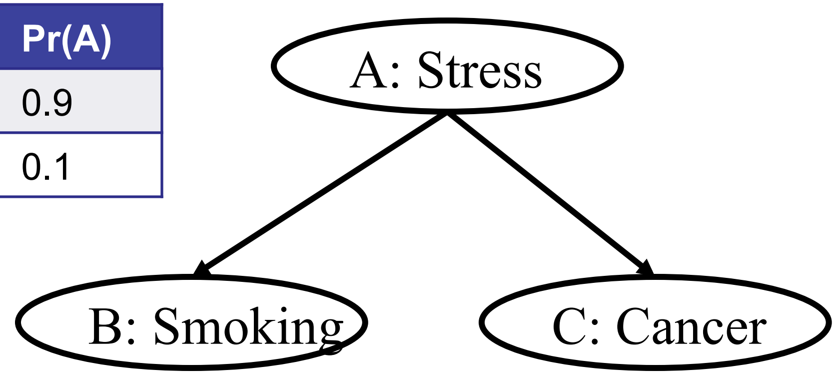


- “A node is independent of its non-descendants given its parents”
- “T wo nodes are independent *unless* they have a common unknown cause, are linked by an chain of unknown causes, or have a common known effect”

Causality and Graphical Models



A	Pr(A)
0	0.9
1	0.1



Pr(A,B,C)=Pr(C|B)Pr(B|A)Pr(A)

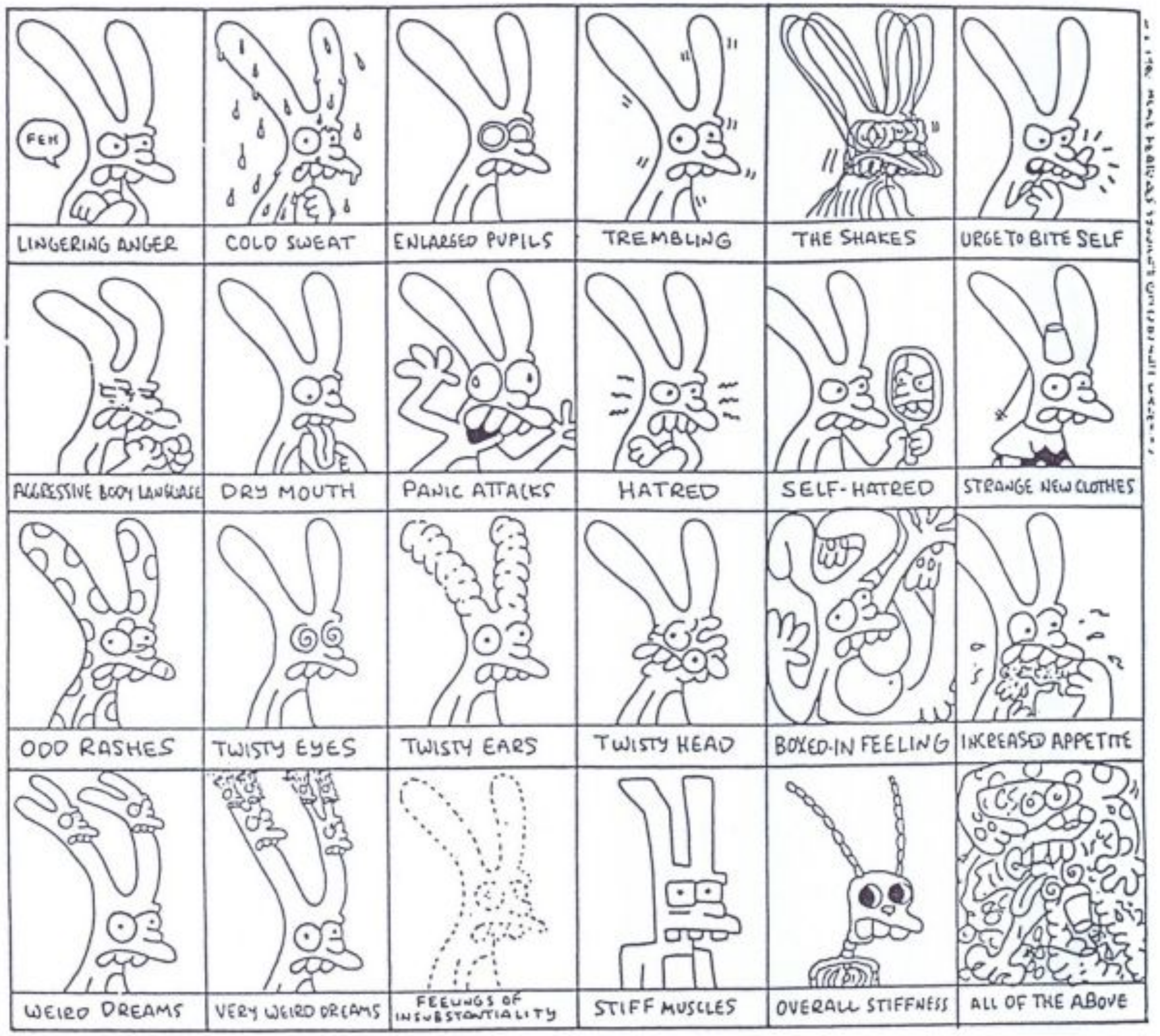
Pr(A,B,C)=Pr(C|A)Pr(B|A)Pr(A)

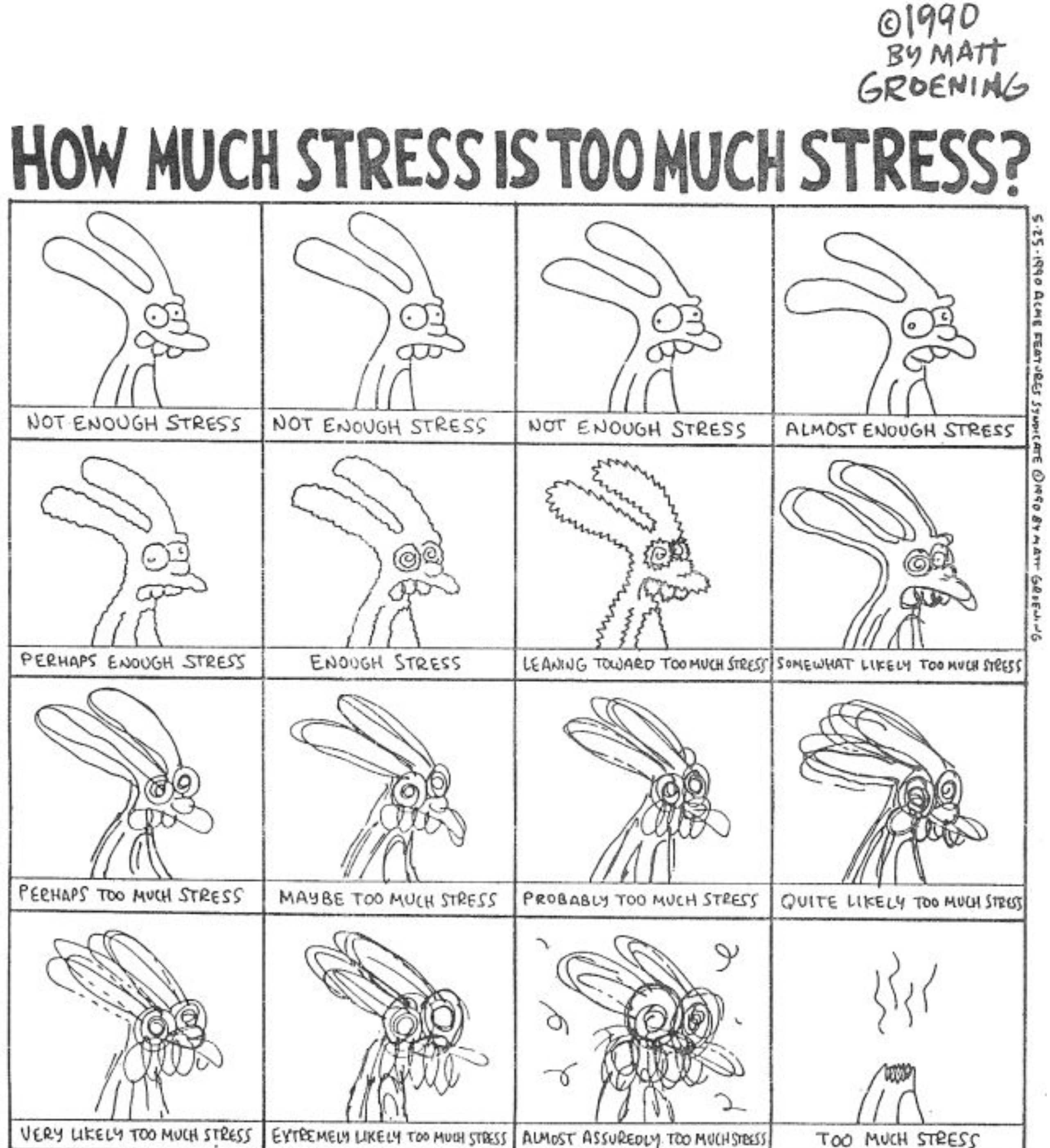
B	C	Pr(C|B)
0	0	0.1
0	1	0.9
1	0	0.1
1	1	0.9

A	B	Pr(B|A)
0	0	0.1
0	1	0.9
1	0	0.1
1	1	0.9

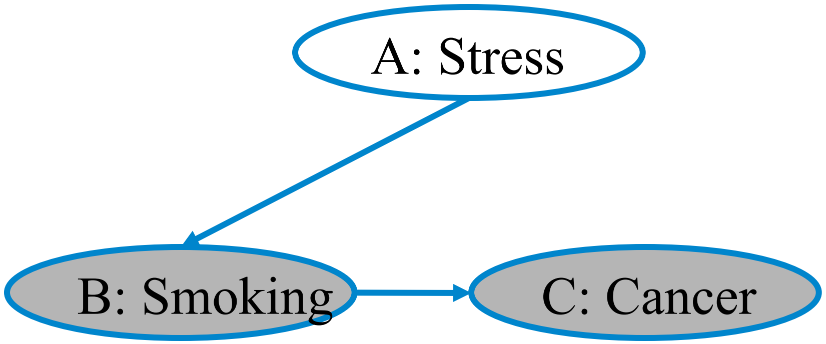
A	C	Pr(C|A)
0	0	0.1
0	1	0.9
1	0	0.1
1	1	0.9

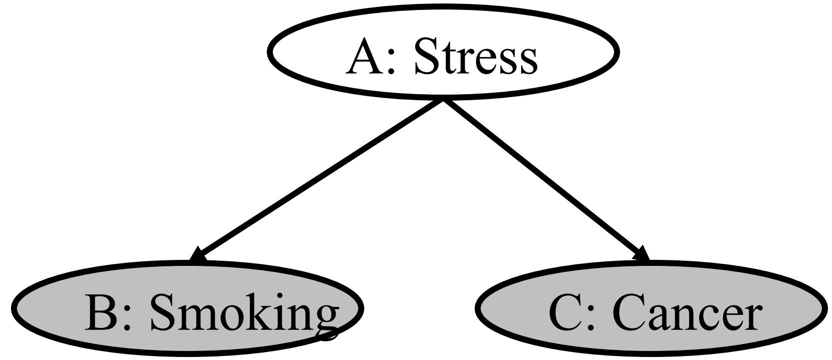
THE 24 WARNING SIGNS OF STRESS




©1990 BY MATT GROENING
HOW MUCH STRESS IS TOO MUCH STRESS?
NOT ENOUGH STRESS
NOT ENOUGH STRESS
NOT ENOUGH STRESS
ALMOST ENOUGH STRESS
PERHAPS ENOUGH STRESS
ENOUGH STRESS
LEANING TOWARD TOO MUCH STRESS
SOMEWHAT LIKELY TOO MUCH STRESS
PERHAPS TOO MUCH STRESS
MAYBE TOO MUCH STRESS
PROBABLY TOO MUCH STRESS
QUITE LIKELY TOO MUCH STRESS
VERY LIKELY TOO MUCH STRESS
EXTREMELY LIKELY TOO MUCH STRESS
ALMOST ASSUREDLY TOO MUCH STRESS
TOO MUCH STRESS
5-25-1990 ACME FEATURES SYNDICATE ©1990 BY MATT GROENING

Causality and Graphical Models





Pr(A,B,C)=Pr(C|B)Pr(B|A)Pr(A)

To estimate:
- Pr(B=b) for b=0,1
- Pr(C=c|B=c) for b=0,1 and c=0,1

The estimates for Pr(B) and Pr(C|B) are correct with *either* underlying model.

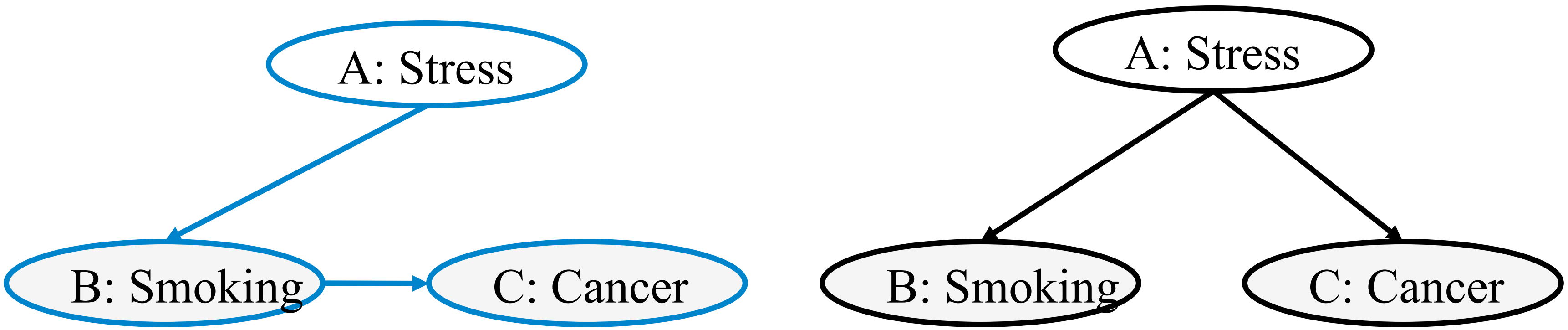
Pr(A,B,C)=Pr(C|A)Pr(B|A)Pr(A)

$$\Pr(C \mid B) = \sum_a \Pr(C \mid a)\Pr(a \mid B)$$

$$= \sum_a \Pr(C \mid a)\frac{\Pr(B \mid a)\Pr(a)}{\Pr(B)}$$

To estimate:
- Pr(B=b) for b=0,1
- Pr(C=c|B=c) for b=0,1 and c=0,1

Causality and Graphical Models

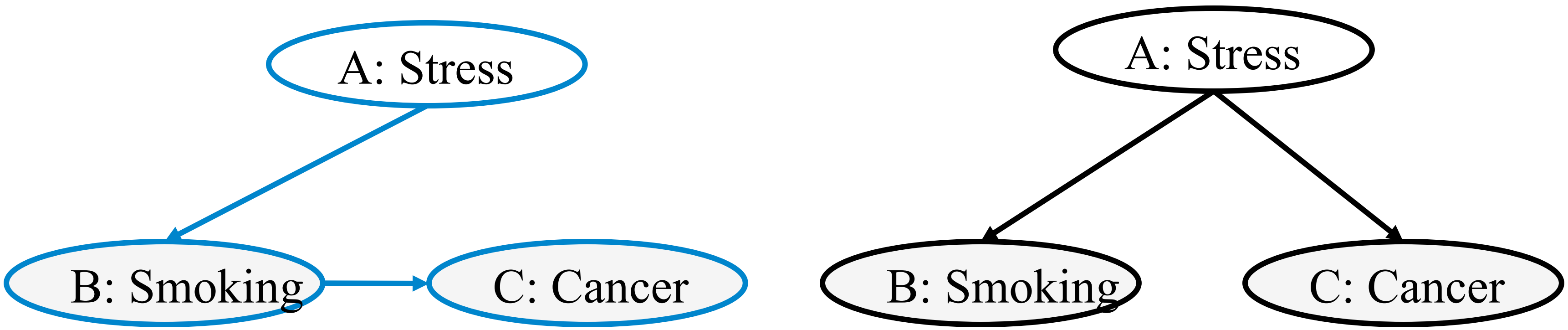


These two models are not "identifiable" from samples of (B,C) only.

Def: A class of models is *identifiable* if you can learn the *true* parameters of any m in M from sufficiently many samples.

Corr: A class of models M is *not* identifiable if there are some distributions generated by M that could have been generated by more than one model in M.

Causality and Graphical Models



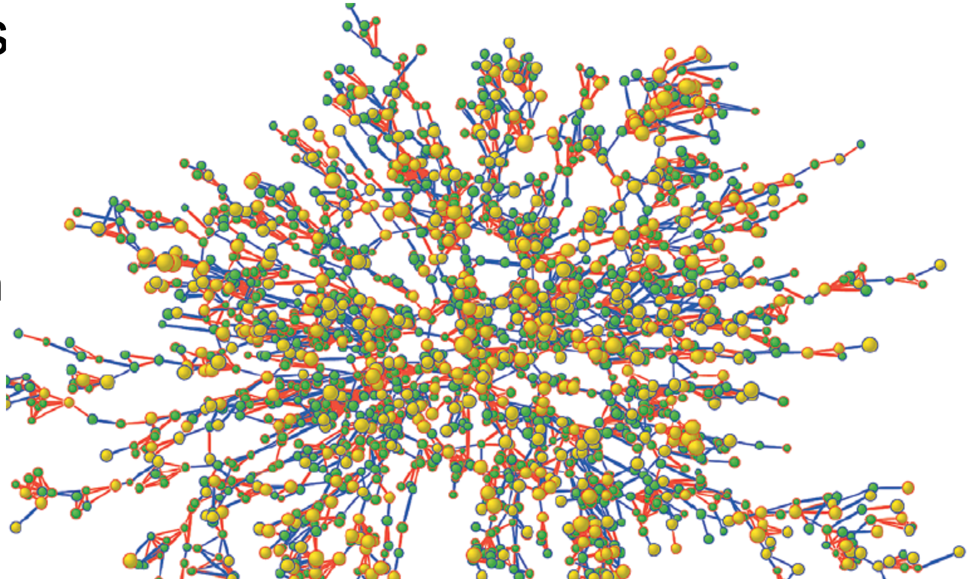
How could you tell the models apart without seeing A?

- Step 1: Interpret the arrows as "direct causality"
- Step 2: Do a manipulation
  - Split the population into Sample and Control
  - Do something to make the Sample stop smoking
  - Watch and see if Cancer rates change in the Sample versus the control

A complicated example

- NEJM, Christakis & Folwer, 2007: Spread of Obesity in A Large Social Network over 32 Years
- Statistical model: for x connected to w:
  - obesity(x,t) = F(age(x), sex(x), …, obesity(x,t-1),obesity(w,t-1))
- Linear regression model, so you can determine influence of a particular variable
- Looked at *asymmetric* links

Not a clinical trial with an intervention



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

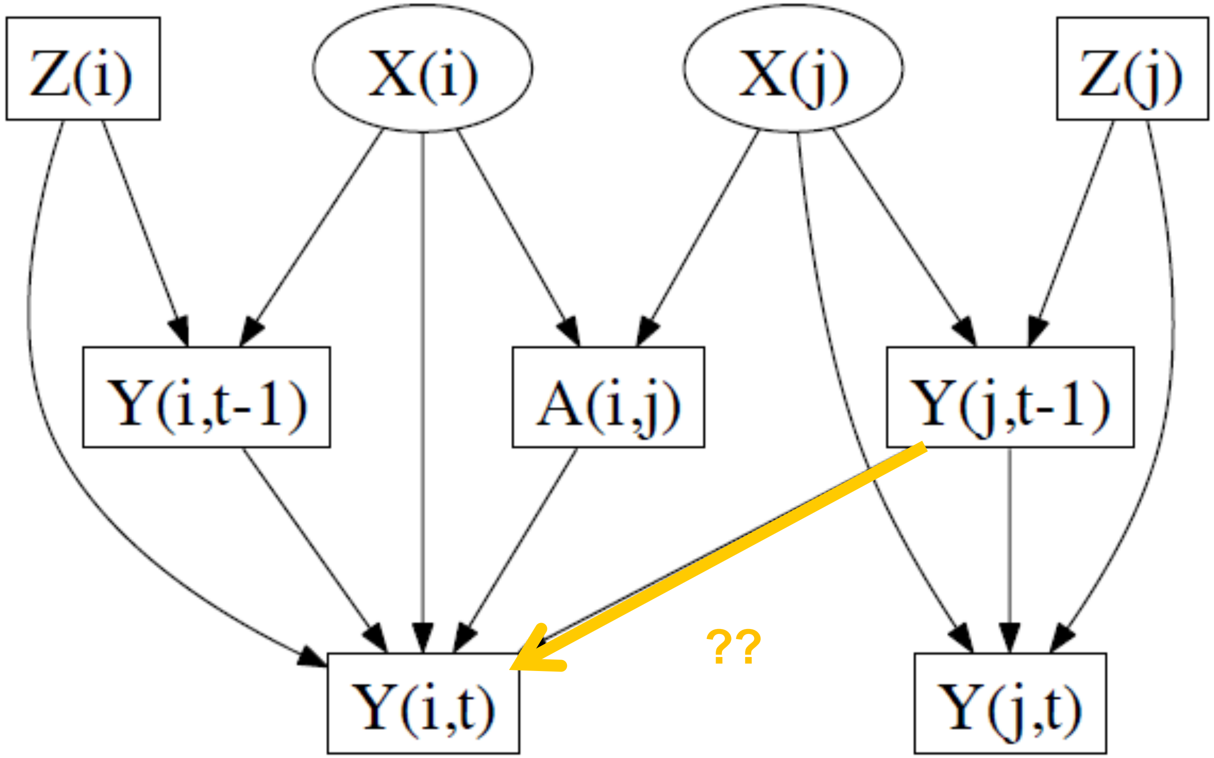
1. yes: Joey inspires you (social contagion or influence)
2. yes: Joey infects you with a parasite which suppresses fear of falling (actual contagion)
3. yes: you’re friends *because* you both like to jump off bridges (manifest homophily)
4. yes: you’re friends *because* you both like roller-coasters, and have a common risk-seeking propensity (latent homophily)
5. yes: because you’re both on it when it starts collapsing and that’s the only way off (external causation)

Notation:

- $Y(i,t) =$ does node i show condition/behavior at time t?
- $X(i) =$ *latent* persistent trait of i
- $Z(i) =$ other, manifest persistent traits
- $A(i,j) =$ whether there is an edge from j to i

We suppose that:

- $Y(i,t-1)$ has a direct influence on $Y(i,t)$
- $X(i)$ has a direct influence on whether/when i adopts
- $Z(i)$ has a direct influence on $Y(i,t)$ (possibly null)
- $Y(j,t-1)$ *may* have a direct influence on $Y(i,t)$, but only if $A(i,j) = 1$
- Homophily: $X(i)$ and $X(j)$ both directly influence $A(i,j)$


Z(i)
X(i)
X(j)
Z(j)
Y(i,t-1)
A(i,j)
Y(j,t-1)
Y(i,t)
??
Y(j,t)

d-separation

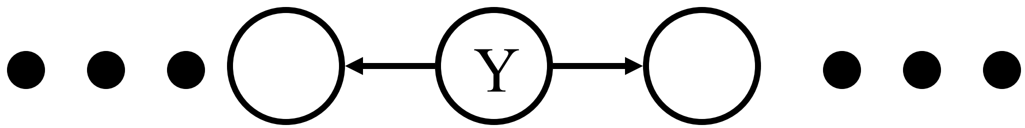
- Fortunately, there is a relatively simple algorithm for determining whether two variables in a Bayesian network are conditionally independent: *d-separation*.

- Definition: *X* and *Z* are *d-separated* by a set of evidence variables *E* iff every undirected path from *X* to *Z* is "blocked", where a path is "blocked" iff one or more of the following conditions is true: ...

ie. X and Z are dependent iff there exists an unblocked path

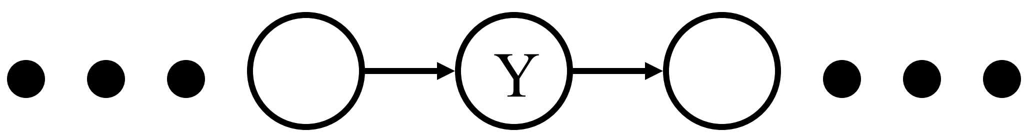
A path is “blocked” when...

- There exists a variable *Y* on the path such that
  - it **is** in the evidence set *E*
  - the arcs putting *Y* in the path are “tail-to-tail”



unknown “common causes” of X and Z impose dependency

- Or, there exists a variable *Y* on the path such that
  - it **is** in the evidence set *E*
  - the arcs putting *Y* in the path are “tail-to-head”

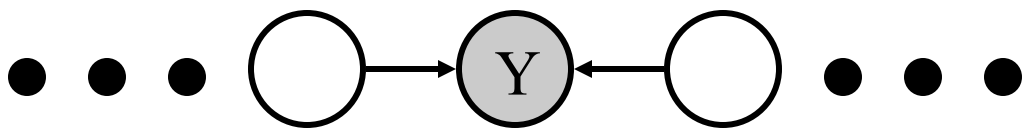


unknown “causal chains” connecting X an Z impose dependency

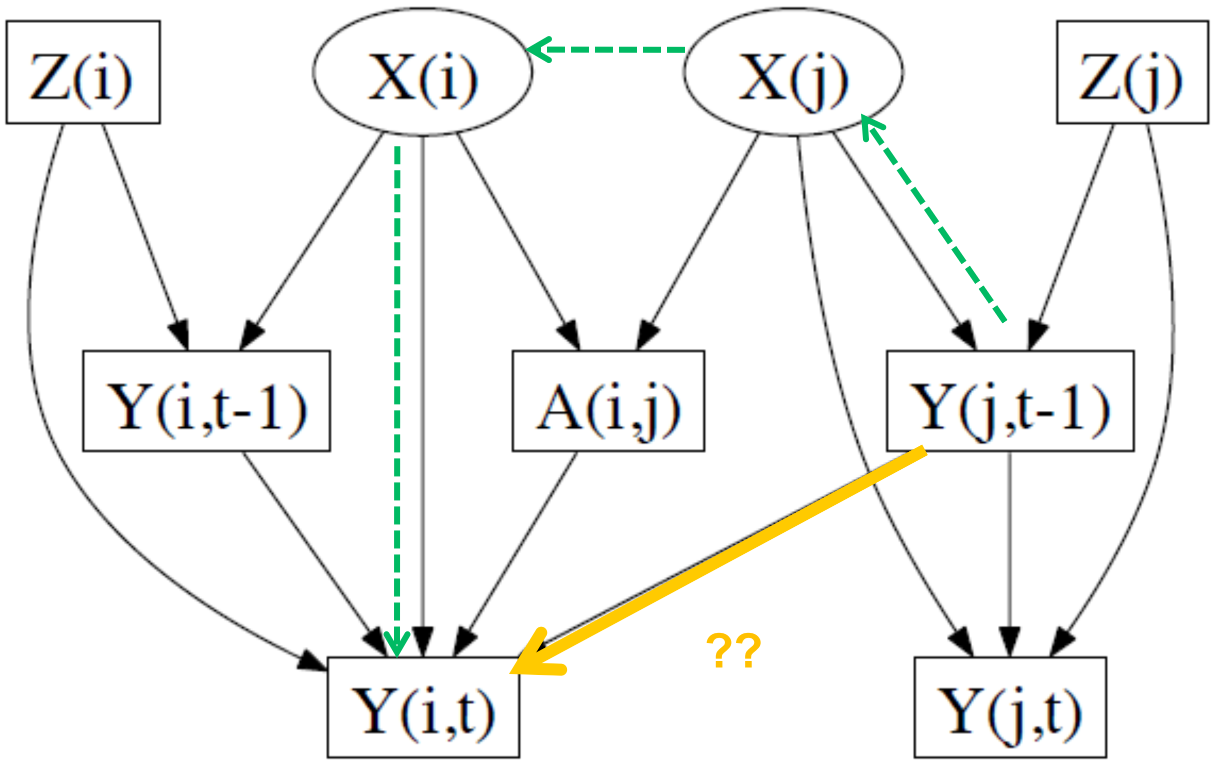
- Or, ...

A path is “blocked” when… (the funky case)

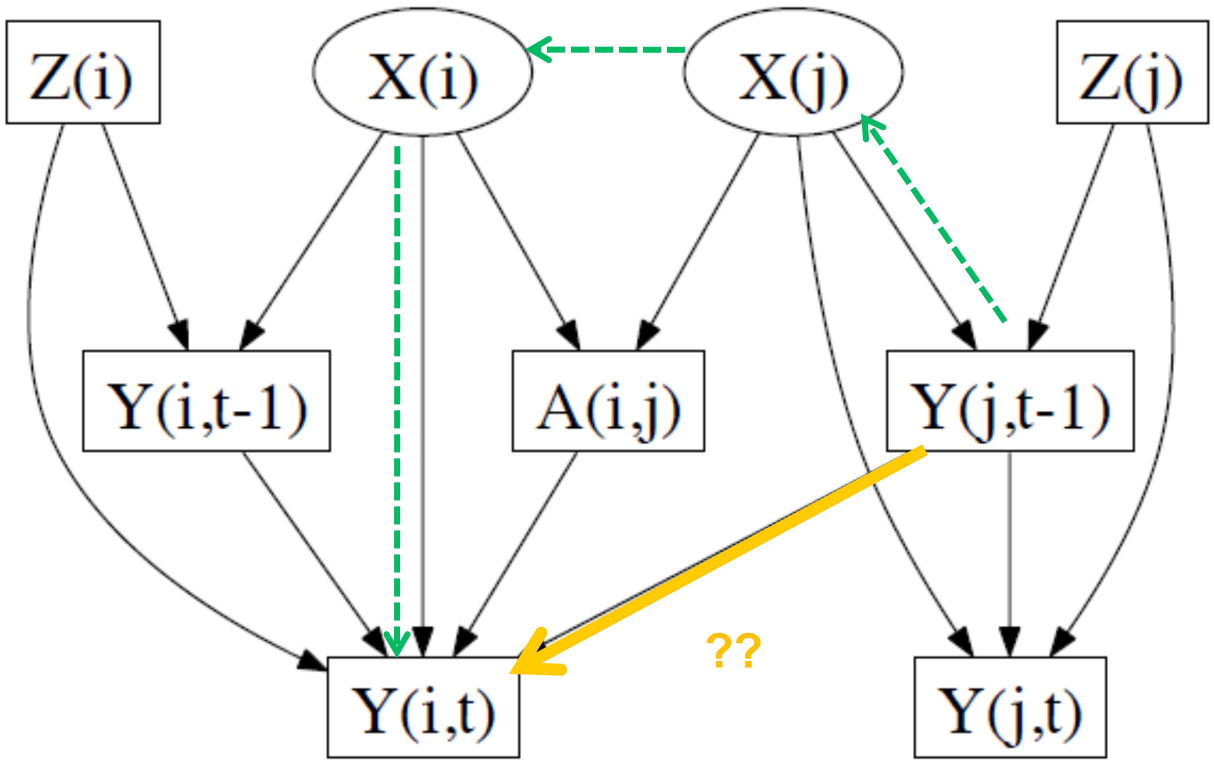
- … Or, there exists a variable *V* on the path such that
  - it **is NOT** in the evidence set *E*
  - **neither are any of its descendants**
  - the arcs putting *Y* on the path are “head-to-head”



<u>Known</u> “common symptoms” of X and Z impose dependencies… X may “explain away” Z

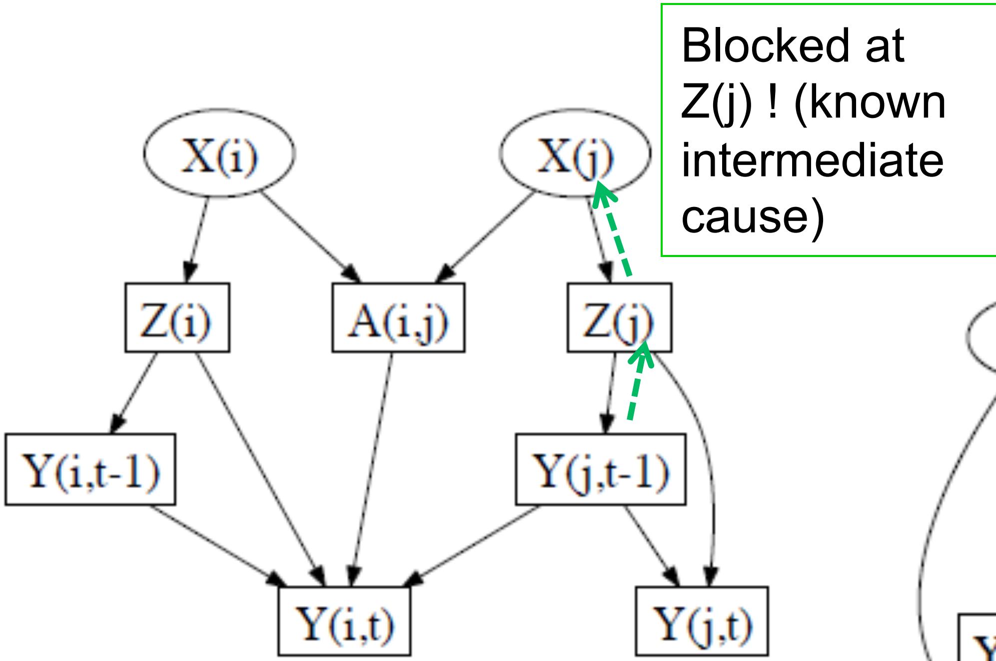


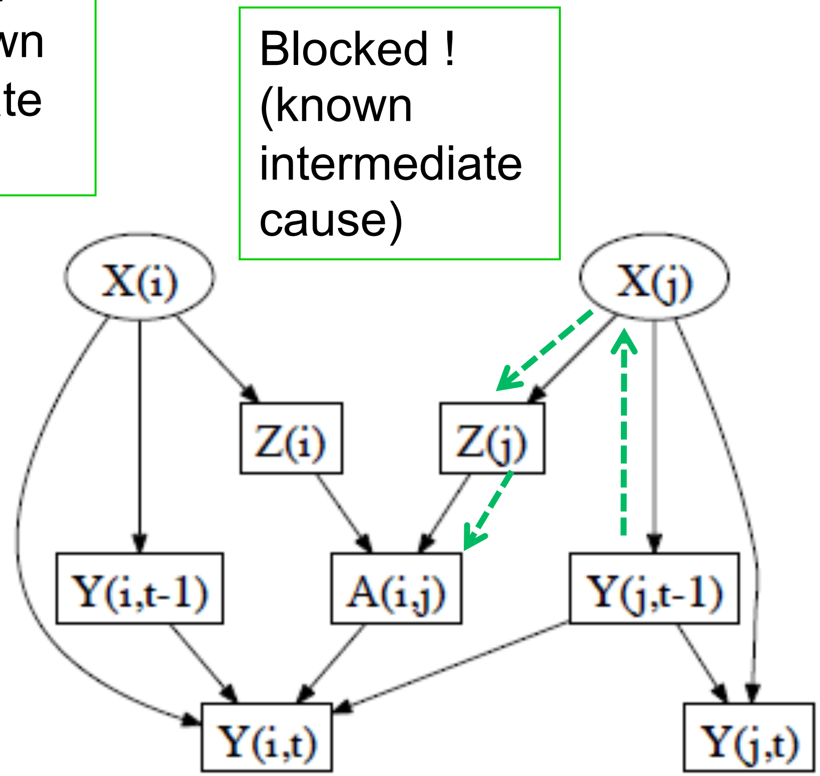
- “A node is independent of its non-descendants given its parents”
- “T wo nodes are independent *unless* they have a common unknown cause or a common known effect”



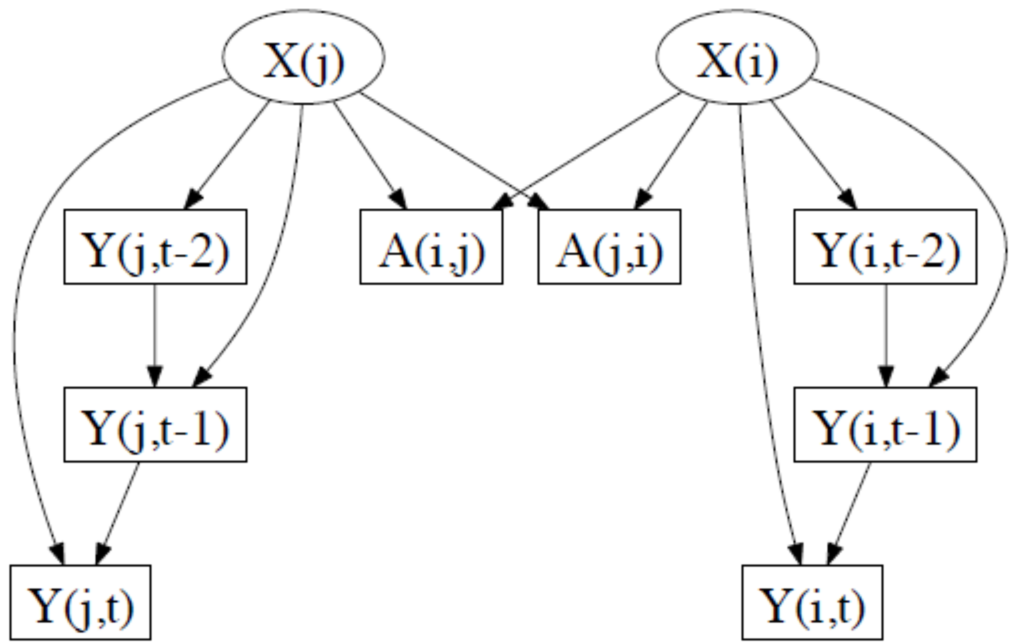
- Conclusion:
  - Y(j,t-1) influences Y(i,t) through latent homophily via the unblocked green path
  - There's no way of telling this apart from the orange path (without parametric assumptions) – model is not "identifiable"

Some fixes


Blocked at Z(j) ! (known intermediate cause)
X(i)
X(j)
Z(i)
A(i,j)
Z(j)
Y(i,t-1)
Y(j,t-1)
Y(i,t)
Y(j,t)


Blocked ! (known intermediate cause)
X(i)
X(j)
Z(i)
Z(j)
Y(i,t-1)
A(i,j)
Y(j,t-1)
Y(i,t)
Y(j,t)

A consequence



One can instantiate this model to show the same effects observed by Christakis and Fowler … even though there is no social contagion