Homophily, Contagion and Confounding: Pick Any Three

Cosma Shalizi

4 December 2008

The basic observation

> *People who are close to each other in social networks tend to be similar and vice versa*

Why?
Dynamics *on* the network (influence, diffusion, contagion)
vs.
Dynamics *of* the network (homophily, assortativity) vs.
Non-network dynamics

Did They Jump or Were They Pushed?

Difference between:

1. jumping off a bridge because your friend Joey did it (contagion)
2. being friends with Joey because you're both the kind of kid who'd jump off a bridge (homophily)
3. You and Joey both watch the wrong TV shows, and respond independently to them

Knowing which is which matters for understanding and for intervention

Making homophily and contagion look like causation

A lot of social science, theortical and applied, is about connecting long-term, hard-to-change social/economic status to more short-term, malleable cultural / political / consumer variables

- Political sociology
- Marketing and political strategy
- Historical materialism

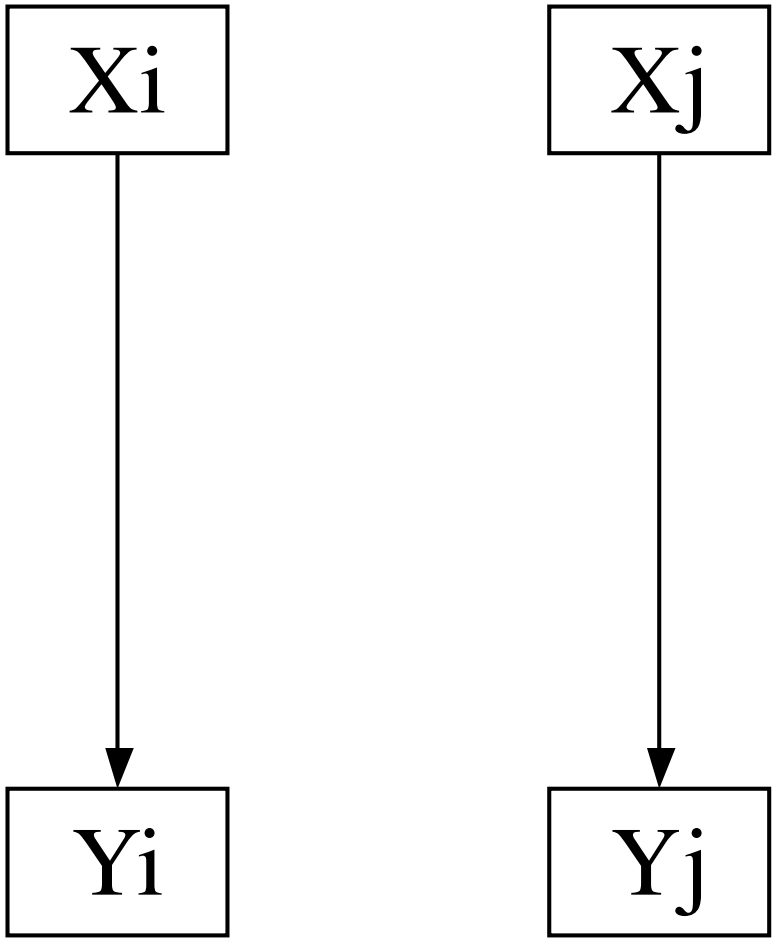
The idea is: how you live has a deep influence on what you think, feel and choose

> *Brecht: "Grub first, then ethics."*
> *Gellner: "Social structure is who you can marry, culture is what you wear at the weddding."*

What's the evidence?

- The stories sound good
- Casual empiricism
- Correlation/regression analyses; cultural choices are predictable from social positions

Probably even true a lot of the time
BUT usually ignores social networks and just looks at surveys



social

cultural

Responsible Just-So Story-telling

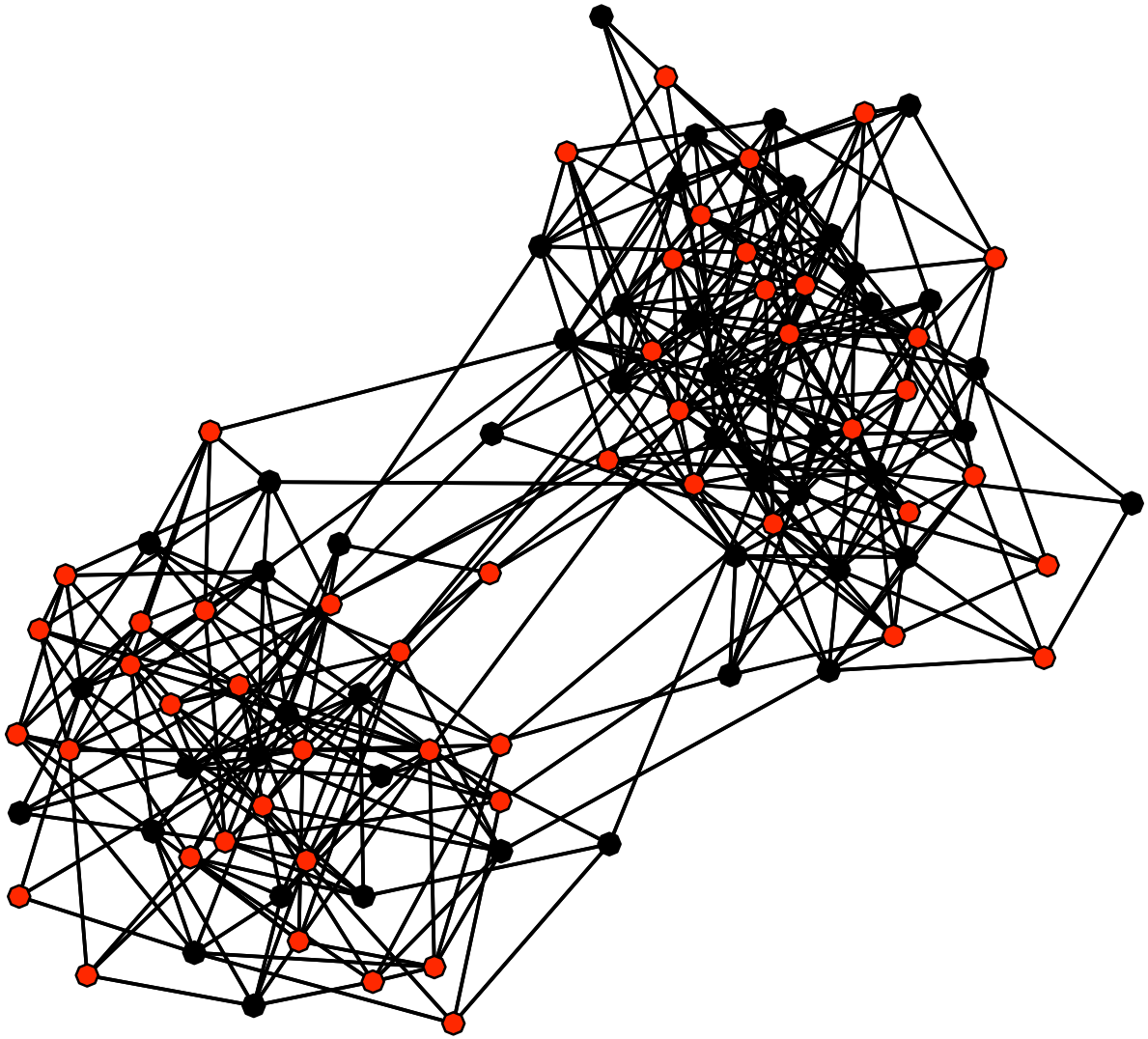
These accounts are usually adaptationist/functionalist
At the very least they are causal accounts
We should really check them
Biology suggests: a **neutral model**

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

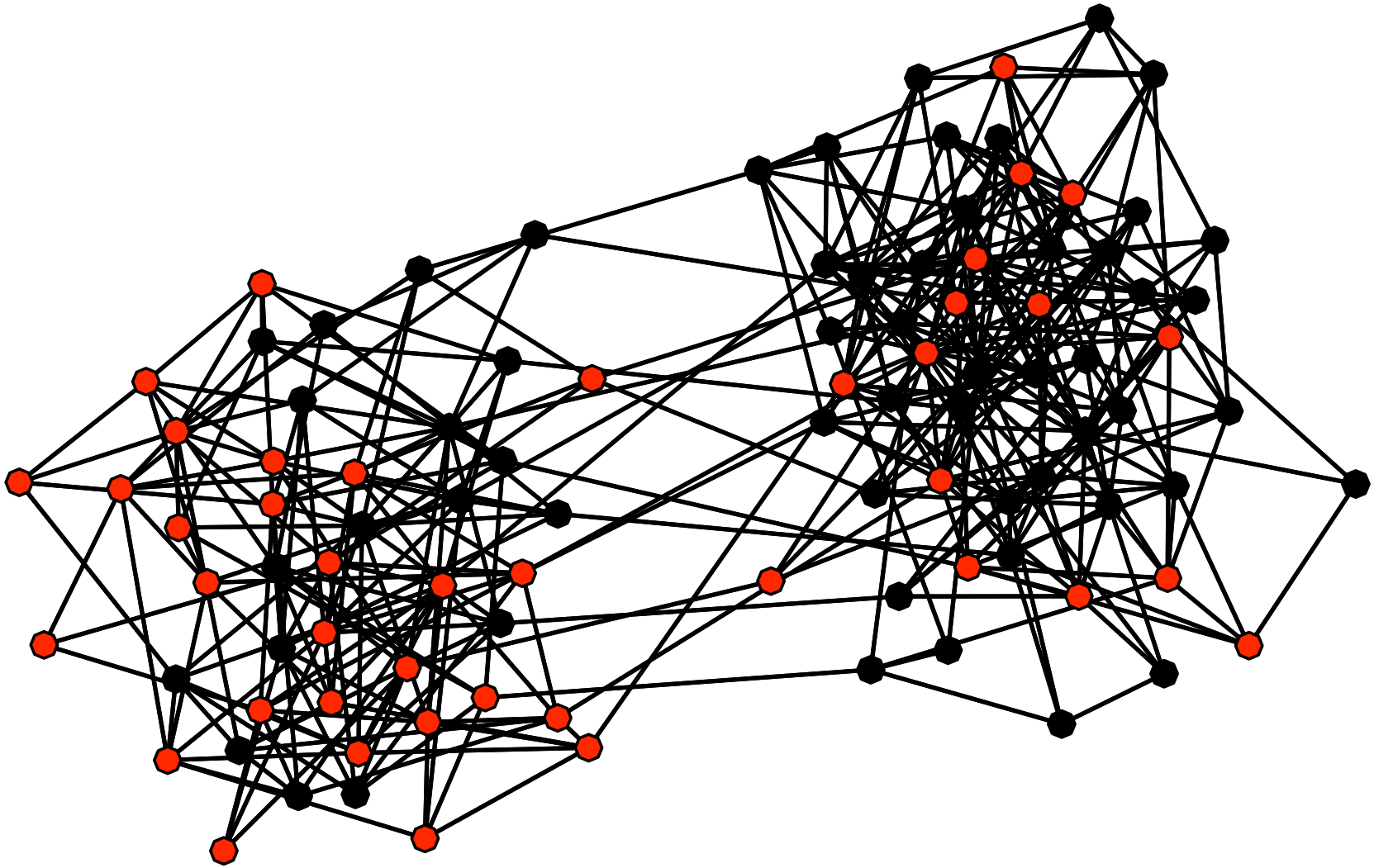
Caricature Neutral Model of Cultural Evolution

- X_i = unchanging status variable for node i ("social")
- Network is assortative on X
- $Y_i(t)$ = rapidly changing choice variable for i ("cultural")
  1. At each t, pick a random i, and a random neighbor j
  2. Set $Y_i(t) = Y_j(t-1)$
  3. Go to (1)

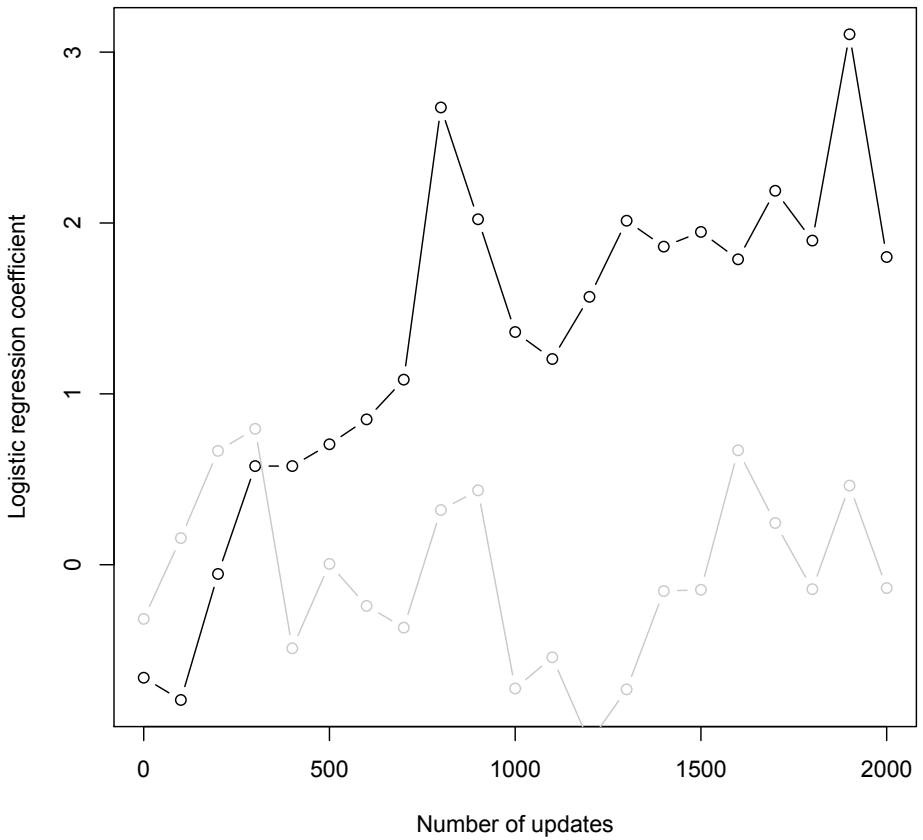
(= "voter model" of statistical mechanics on a stochastic graph)



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

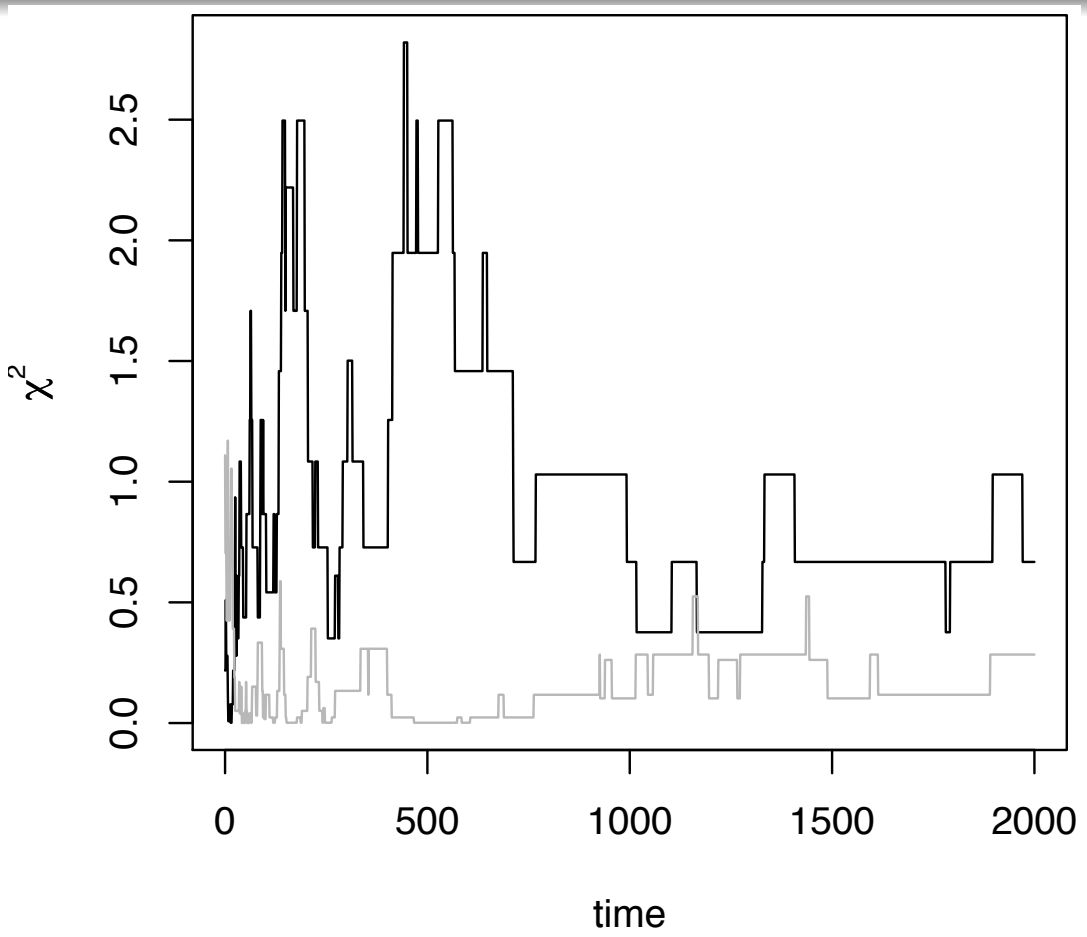


After 1000 updates

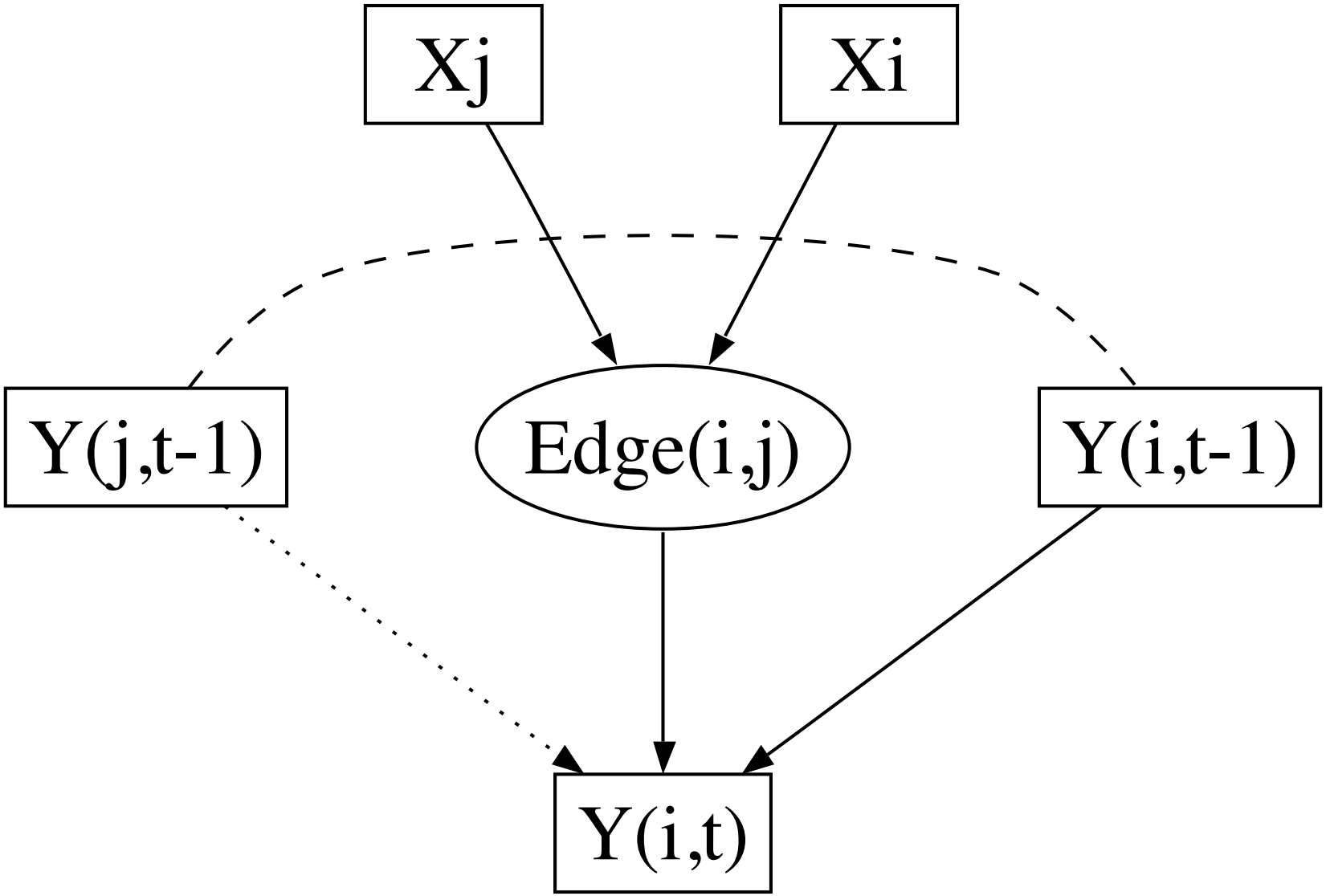


100 node network

Logistic regression of choice on status
black = assortative
grey = non-assortative



Same, χ^2 for status/choice contingency table


Xj
Xi
Y(j,t-1)
Edge(i,j)
Y(i,t-1)
Y(i,t)

— 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), since we're not using aggregated data
— Problem is that choices are not independent conditional on statuses
— Deconfound by conditioning on previous Y_j of neighbors

Can You Really Get Cooties From Touching Someone?

Finding: whether i does something at time t is well-predicted by whether i's neighbors had already done it at $t-1$

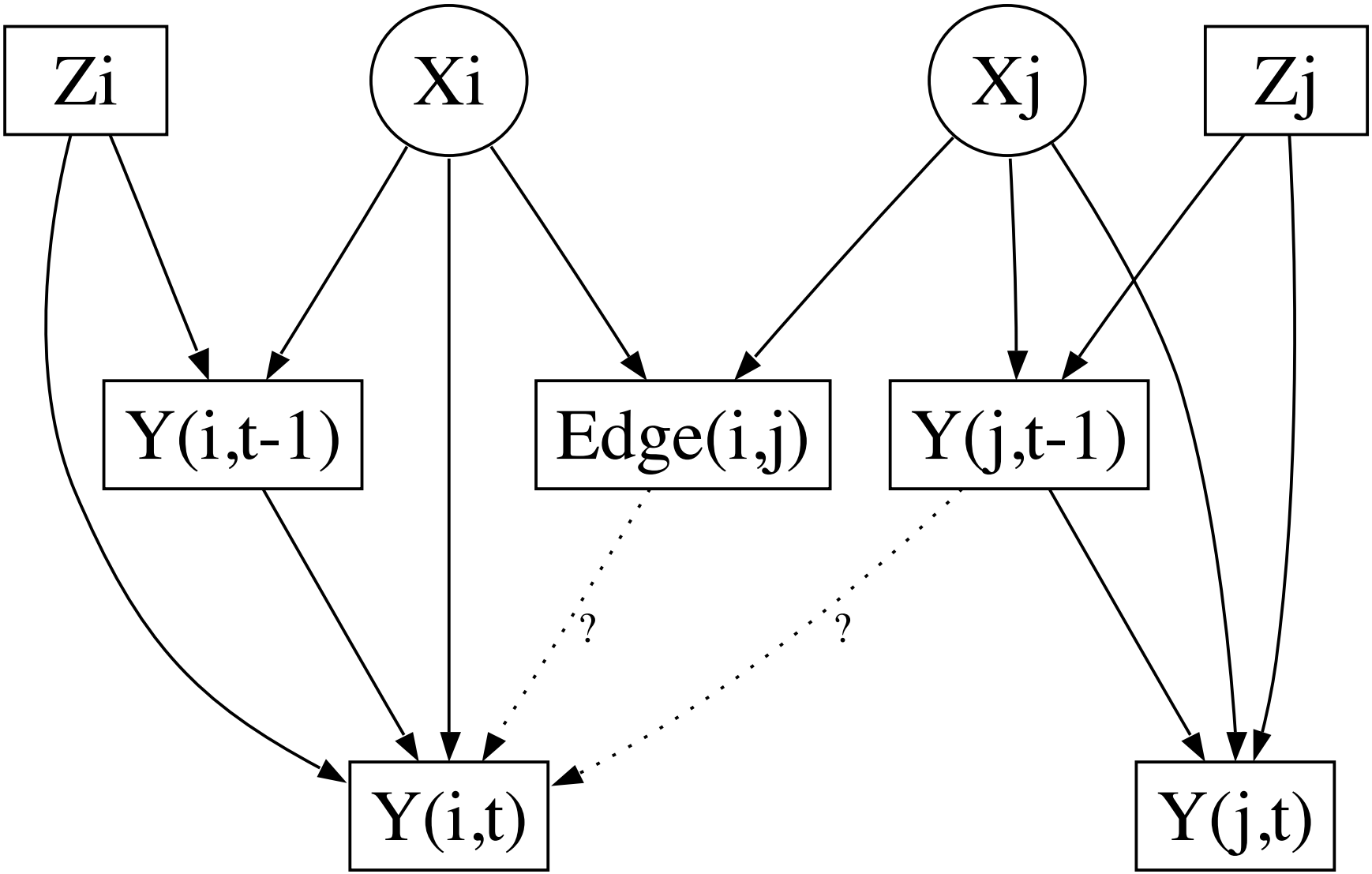
- Diffusion of innovations
- Infectious diseases
- Not-obviously-infectious medical conditions (e.g., obesity)

...

Again, often correct
Again, how hard is it to fake this without real contagion?

Suppose that:

- X_i = persistent trait/traits of node i, *not directly observable*
- Homophily on X
- Z_i = other, non-assortative but observable traits
- $Y_i(t)$ = whether i has adopted at t, τ_i = time of adoption
- X_i influences whether/when i adopts


Zi
Xi
Xj
Zj
Y(i,t-1)
Edge(i,j)
Y(j,t-1)
?
?
Y(i,t)
Y(j,t)

Item: Y_i is a cue for X_i
Item: Neighbor's X_j are cues for X_i
$\therefore$ Neighbors' Y_j are cues for X_i
$\therefore$ Y_i depends statistically on previous Y_j of neighbors, whether there's a direct edge or not
$\therefore$ Latent homophily is confounded with contagion
—Conditioning on $Y_i(t-1)$, $Y_j(t-1)$ does not screen off homophily

except for homophily due to adoption status

—Need to condition on the homophilous trait X_i

Or other covariates Z such that $X_i \perp X_j | (Z_i, Z_j)$, if they exist

Conclusion

- Homophily + contagion looks like causality
- Homophily + causality looks like contagion
- Contagion looks like homophily (especially with causality)