

Observational Study of Contagion Effect  
oooooooooooo

Statistical identification of causal effect  
oooooooooooooooooooo

Experimental Study of Contagion Effect  
oooooooooooooooooooo

# SOSC 4300/5500: Homophily, Network, and Causal Inference

Han Zhang

Observational Study of Contagion Effect  
oooooooooooo

Statistical identification of causal effect  
oooooooooooooooooooo

Experimental Study of Contagion Effect  
oooooooooooooooooooo

## Outline

Observational Study of Contagion Effect

Statistical identification of causal effect

Experimental Study of Contagion Effect

## Empirical work on diffusion

- Last week we talked about **mathematical** models of diffusion
  - Simple contagion on small-world network
  - Complex contagion on complex networks
- Let us move on to **empirical** results regarding diffusion
- First, focus on observational data

Friends are similar to each other

- Countless evidence suggests that friends are usually more similar to each other than to random others
    - In mathematical terms, the **correlation** between friends' attitudes tend to be higher than the correlation between two random nodes
  - List to similar music; watch similar sports
  - Similar demographics (age, education, etc)
  - others?

## Observational Study of Contagion Effect

○○●○○○○○

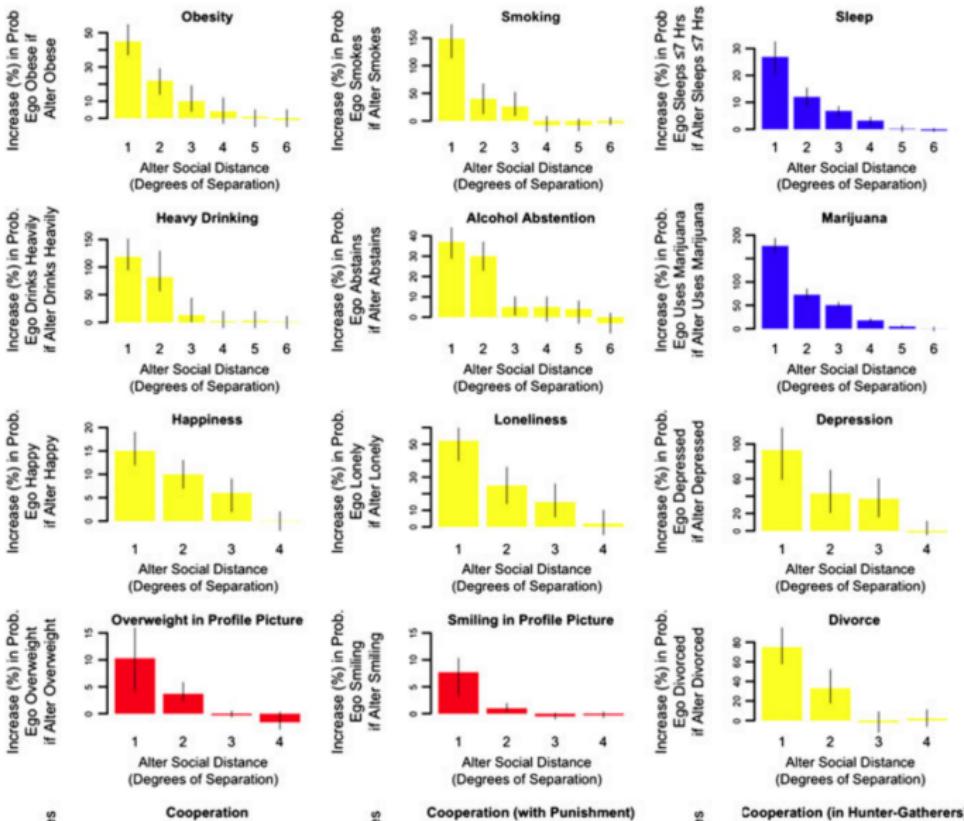
## Statistical identification of causal effect

○○○○○○○○○○○○○○○○

## Experimental Study of Contagion Effect

○○○○○○○○○○○○○○○○

# Friends are similar to each other



## Diffusion of Obesity: social influence

- Nicholas A. Christakis and James H. Fowler, *The Spread of Obesity in a Large Social Network over 32 Years*, New England Journal of Medicine **357** (2007), no. 4, 370–379
  - <https://www.youtube.com/watch?v=8aEtyRD1j5U>
  - Christakis and Fowler think it's because of **social influence**:
    - you follow your friends' behaviors
    - so obesity **diffuses** along social networks
    - or there is **contagion** of obesity along social networks
  - Other explanations?

## Alternative mechanism 1: homophily

- Miller McPherson, Lynn Smith-Lovin, and James M. Cook, *Birds of a Feather: Homophily in Social Networks*, Annual Review of Sociology **27** (2001), 415–444
- **Homophily**: people are more likely to befriend with others similar to themselves
  - Primary homophily: overweight people become friends with other overweight people
  - Second homophily: overweight people share some other characteristics (e.g., lower-income family in the US), which drives homophily
- If your goal is to estimate contagion effect, then homophily is essentially a kind of selection bias

## Contagion vs. Homophily

- **Contagion (social influence) vs. homophily**
  - they result in the same correlation pattern: friends are similar to each other
  - but mechanisms are very different
- Social influences and homophily can both explain the below:
  - Friends are similar in their music taste
  - Friends usually consume similar political news (so-called filter bubbles or echo chambers)
  - Smoking/drinking
  - Emotion
  - Political beliefs
- Can you think of other examples?

## Alternative mechanism 2: contextual effect

- They just happen to share the same environment
  - A type of **omitted variable bias**, or confounding biases, or common external causes
- E.g., live in poor neighborhood with only fast food restaurants
- Contextual bias is *relatively* easier to control statistically
  - get more measures on confounding variables

## Compare three mechanisms

- Social influence/diffusion/contagion
- Homophily
- Same context/environment
- They all result in same correlation, but policy implications are very different
- [Question] Say, you observe that friends' political beliefs are very similar, and you are worried about echo chamber. What policies would you suggest if you want to reduce similarity in friends' political beliefs, if each of the three mechanism is true?

## Mechanisms and policy implications

- Social influence: encourage them to talk more about non-political contents?
- Homophily: let Facebook recommend some friends from the other political view in the beginning
- Context: try to downgrade the influence of some common information sources? Like ban some common low-quality information sources?
- As you can see, if we know which mechanism has more impact, we can design better policies

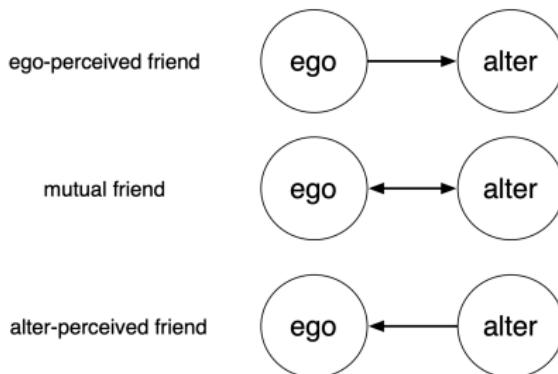
## Christakis and Fowler's statistical model

- Christakis and Fowler wants to demonstrate that social influence exists
- They used statistical model to prove that control for homophily and other con founders, social influence still exists
- Regress whether ego is obese at time  $t + 1$  based on:
  - alter obese at  $t$ 
    - outcome of interest
  - ego obese at  $t$ 
    - Control for autocorrelation
  - alter obese at  $t + 1$ 
    - Control for homophily
  - ego demographic variables (age, gender, education, etc.)
    - Control for observed confoundings

$$Y_{t+1}^{\text{ego}} = \alpha + \beta_1 y_t^{\text{ego}} + \beta_2 y_{t+1}^{\text{alter}} + \beta_3 y_t^{\text{alter}} + \sum_{i=1}^k \gamma_i x_i \quad (1)$$

## Unobserved factors

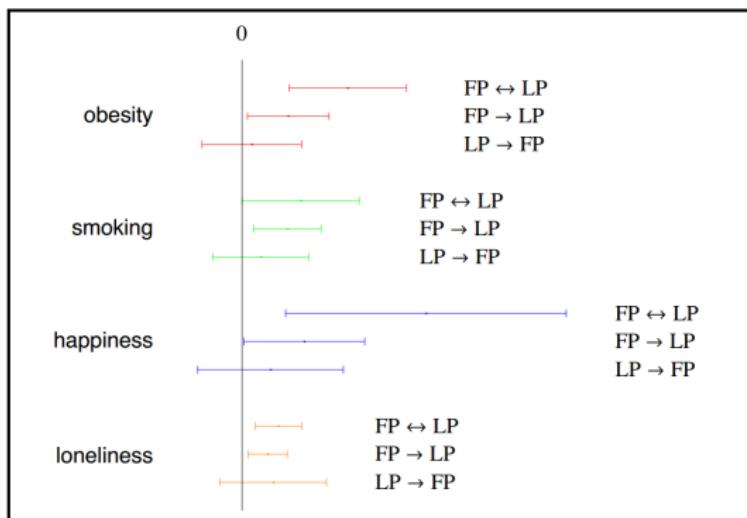
- Even after controlling for many things, we may still have **unobserved** factors
  - like where do people frequently go, which they have no data on
- To rule out unobserved shared environments: using **directionality** of ties:



## Directionality of ties and unobserved factors

- If social influence is the true mechanism, then
  - mutual friends are most likely to influence each other
  - ego-> alter the second
  - ego<- alter the least likely (you think I am your friends, but I, the survey respondent, do not think you are my friends); So I should not be impacted by you
- If unobserved context matter, then it's likely that you and I share the similar environment **regardless** of directionality of ties

## Directionality of ties and unobserved factors



- FP is ego and LP is alter
- Hence, what we observed is more likely due to social influence, not unobserved contexts

## “Towards Responsible Just-So Story Telling”

- Cosma Rohilla Shalizi and Andrew C. Thomas, *Homophily and Contagion Are Generically Confounded in Observational Social Network Studies*, arXiv:1004.4704 (2010)
- They offered a mathematical proof on:  
*Contagion effects are nonparametrically unidentifiable in the presence of latent homophily*
- Latent homophily are the **unobserved** similarities that push people to become friends
  - Like genetic similarities. There are arguments that genetically similar people are more likely to make friends with each other
  - But unless your survey have genetics data (mostly likely they don't), you don't have a measure for similarities in genes

## An area of debates

- Generally, it is very challenging to separate the three causal mechanisms from observational data
- Christakis and Fowler's data are already better than most because they are panel data and have directionality

Tyler J. VanderWeele, Elizabeth L. Ogburn, and Tchetgen Eric J. Tchetgen, *Why and When "Flawed" Social Network Analyses Still Yield Valid Tests of no Contagion*, Statistics, Politics, and Policy 3 (2012), no. 1

*Social network analyses of the type employed by Christakis and Fowler will still yield valid tests of the null of no social contagion, even though estimates and confidence intervals may not be valid.*

## Diffusion of Music Taste, or no?

- Kevin Lewis, Marco Gonzalez, and Jason Kaufman, *Social selection and peer influence in an online social network*, Proceedings of the National Academy of Sciences **109** (2012), no. 1, 68–72

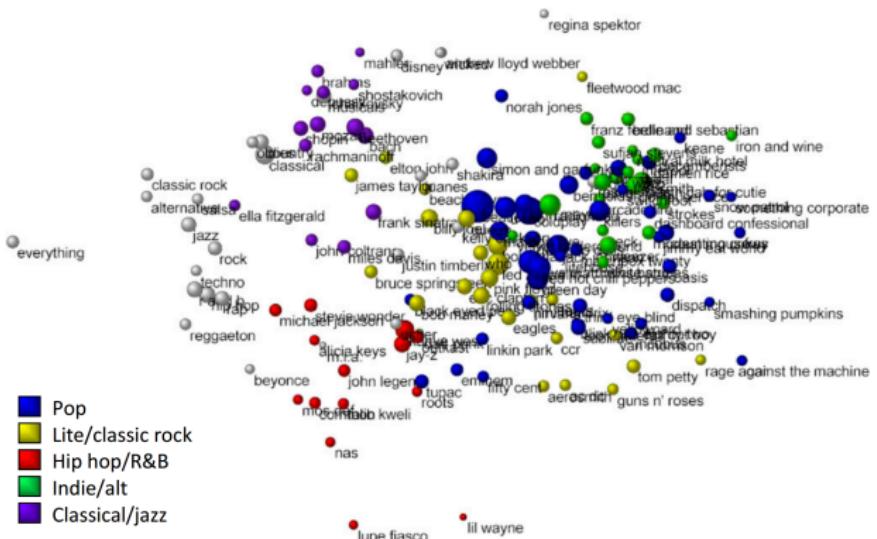
*Our data are based on the Facebook activity of a cohort of students at a diverse US college ( $n = 1,640$  at wave 1). Beginning in March 2006 (the students' freshman year) and repeated annually through March 2009 (the students senior year)*

## Observational Study of Contagion Effect

## Statistical identification of causal effect

## Experimental Study of Contagion Effect

# Diffusion of Music Taste, or no?



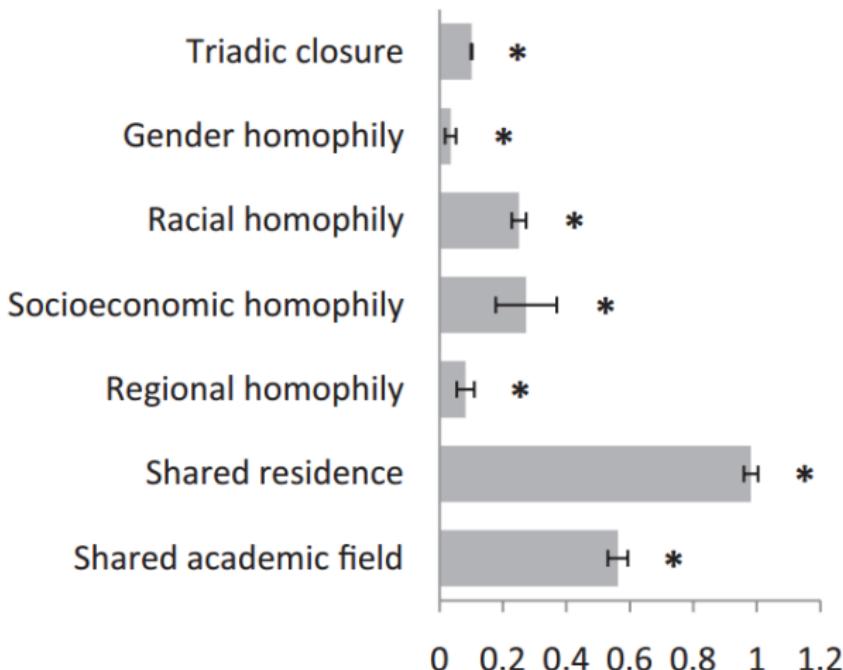
## Social influence vs. homophily again

- Are similar music taste between friends
  - social influence
  - homophily?
- A differnet approach: **stochastic actor-oriented network models** (SAOM)
  - Krzysztof Nowicki and Tom A. B. Snijders, *Estimation and Prediction for Stochastic Blockstructures*, Journal of the American Statistical Association **96** (2001), no. 455, 1077–1087
  - <https://www.stats.ox.ac.uk/~snijders/siena/>

## SAOM

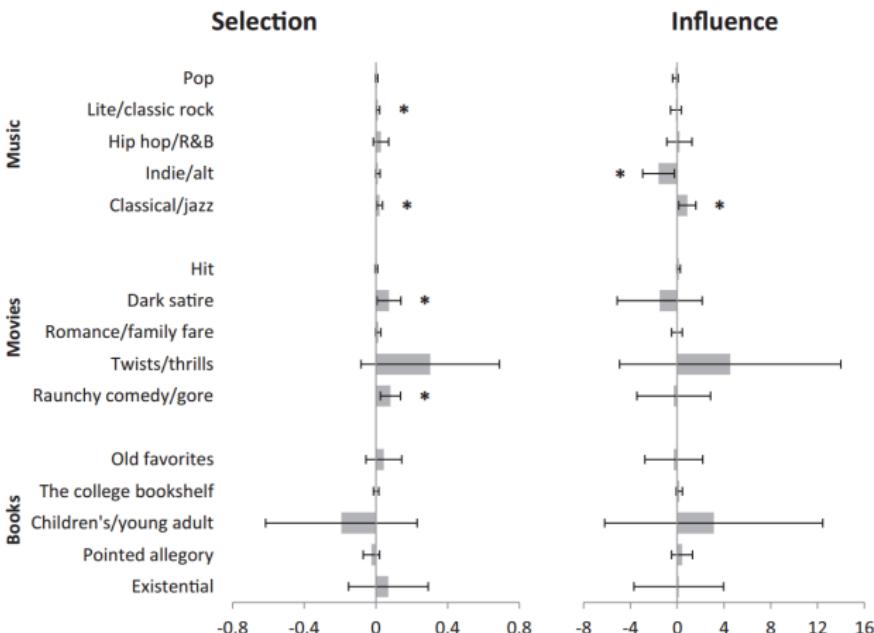
- **Simulate** what a network would look like, if both homophily and social influence exists
  - Step 1: fix network structure; allow music taste to diffuse along networks (similar to the examples you have seen last week)
  - Step 2: fix the music tastes; allow people to change their friends, based on similarity of music tastes (homophily)
  - Repeat Step 1 and 2 until convergence
- Now, if we only run step 1 or step 2, then we would know what network looks like if there were no social influence, or no homophily
- Then compare these two simulated networks with observed network

## What factors predict friendship



- Shared space is the mostly salient predictor

# Homophily vs. social influence



- Some evidence for selection/homophily
- But no evidence for social influence

## ERGM

- There is a separate branch of statistical analysis that simulates networks and then estimate effects
- Exponential random graph models (ERGM)
- Per Block, Johan Koskinen, James Hollway, Christian Steglich, and Christoph Stadtfeld, *Change we can believe in: Comparing longitudinal network models on consistency, interpretability and predictive power*, Social Networks **52** (2018), 180–191
- SOAM slightly better, but “both models perform poorly in out-of-sample prediction compared to trivial predictive models.”

## Take-away messages

- Friends are similar to each other
- But it may be due to different reasons
  - homophily
  - contagion/social influence
  - common environment
- Contagion effect is difficult to estimate in observational data when people can choose their friends and may be exposed to environmental changes that we don't measure
- To do it right, you need to take more statistics class
  - and even the best models so far have problems
- Or alternatively, consider running experiments
- Both are frontiers of science, now!

Observational Study of Contagion Effect  
oooooooooo

Statistical identification of causal effect  
oooooooooooooooooooo

Experimental Study of Contagion Effect  
●oooooooooooooooooooo

## Randomized controlled experiments

- Randomized controlled experiments
  - Split subjects into treated/control groups randomly
  - Provide treatment to treated users, not to control users
  - Compare **difference in means**
- Simple and straightforward, compared with observational studies

Observational Study of Contagion Effect  
oooooooooo

Statistical identification of causal effect  
ooooooooooooooo

Experimental Study of Contagion Effect  
○●ooooooooooooooo

## Two types of experiments

- Lab experiment
  - Pros: full control
  - Cons: lack of external validity; full of convenient samples
- Field experiment
  - Pros: external validity;
  - Cons: take a lot of resources
- Internet era has led to **online lab/field experiment**

Observational Study of Contagion Effect  
oooooooooo

Statistical identification of causal effect  
ooooooooooooooo

Experimental Study of Contagion Effect  
oo●oooooooooooo

## Emotion contagion

- Hypothesis: seeing happy messages from friends also make you happier
- Counter-arguments
  - From homophily perspective?
  - From shared context/environment perspective?

## Emotion contagion

- Adam D. I. Kramer, Jamie E. Guillory, and Jeffrey T. Hancock, *Experimental evidence of massive-scale emotional contagion through social networks*, Proceedings of the National Academy of Sciences **111** (2014), no. 24, 8788–8790
- Experimental design (**online field experiment**)
  - Around 700,000 people
  - Three groups:
    - positive posts in Facebook News Feed reduced (by 10% to 90%)
    - negative posts in Facebook News Feed reduced
    - control
  - Post scored as positive or negative based on LIWC dictionary
  - Outcome: proportion of words posted that were positive or negative in 1 week of experiment

Observational Study of Contagion Effect  
oooooooooooo

Statistical identification of causal effect  
oooooooooooooooooooo

Experimental Study of Contagion Effect  
oooo●oooooooooooo

## Emotion Contagion

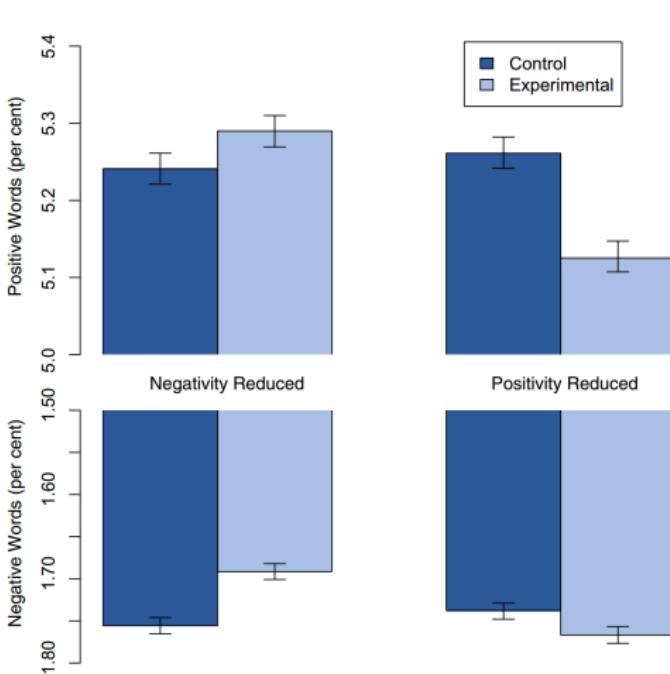


Fig. 1. Mean number of positive (*Upper*) and negative (*Lower*) emotion words (percent) generated people, by condition. Bars represent standard errors.

Observational Study of Contagion Effect  
oooooooooo

Statistical identification of causal effect  
oooooooooooooo

Experimental Study of Contagion Effect  
ooooo●oooooooooo

## Questions

- Is posting on Facebook a good measure of how we feel?
- Is LIWC dictionary methods a good way to quantify emotional content of a Facebook post?

Observational Study of Contagion Effect  
oooooooooo

Statistical identification of causal effect  
ooooooooooooooo

Experimental Study of Contagion Effect  
oooooooo●oooooooo

## Ethical concerns

- Most importantly, should this experiment ever have happened?
- Facebook reveals news feed experiment to control emotions; protests over secret study involving 689,000 users in which friends' postings were moved to influence moods
- Stop complaining about the Facebook study. It's a golden age for research
- A collection of articles on the ethical debates of Facebook's emotion contagion study

[http://laboratorium.net/archive/2014/06/30/the\\_facebook\\_emotional\\_manipulation\\_study\\_source](http://laboratorium.net/archive/2014/06/30/the_facebook_emotional_manipulation_study_source)

## Simple and complex contagion

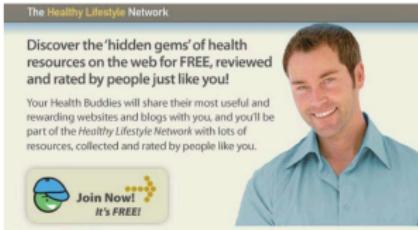
- Last week, we learned the differences between simple and complex contagion

|                   | Regular network (lattice) | Small-world network |
|-------------------|---------------------------|---------------------|
| Simple contagion  | slow                      | fast                |
| Complex contagion | fast                      | slow                |

- Can this theory be empirically tested?
- Damon Centola, *The Spread of Behavior in an Online Social Network Experiment*, Science **329** (2010), no. 5996, 1194–1197
- This is an example of **online lab experiment**

## Recruitment of research subjects

- “I created an Internet-based health community, containing 1528 participants recruited from health-interest World Wide Web sites.”



# What this health community looks like?

Here's How It Works - Mozilla Firefox

File Edit View History Bookmarks Tools Help

http://healthylifestyle.nw.harvard.edu/ Google

The Healthy Lifestyle Network

You are: John-672  Finish

Your health interests:

- \* Weight loss and dieting
- \* Lowering cholesterol
- \* Exercise programs
- \* Stress reduction and relaxation

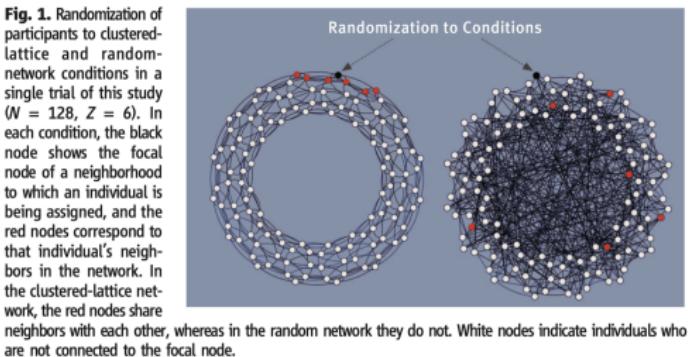
These are your health buddies:

|  |   |  |
|--|---|--|
| <br>Toan-502<br>Health interests: <ul style="list-style-type: none"><li>* Stress reduction and relaxation</li><li>* Exercise programs</li><li>* Alcohol use and stress factors</li></ul>  | <br>Jeff-459<br>Health interests: <ul style="list-style-type: none"><li>* Exercise programs</li><li>* Stress reduction and relaxation</li><li>* Avoiding environmental pollutants</li></ul>        | <br>David-370<br>Health interests: <ul style="list-style-type: none"><li>* Weight loss and dieting</li><li>* Exercise programs</li><li>* Using vitamin supplements</li></ul>                                  |
| <br>Joshua-150<br>Health interests: <ul style="list-style-type: none"><li>* Stress reduction and relaxation</li><li>* Exercise programs</li><li>* Finding where and how to get screenings</li><li>* Limiting sun exposure</li></ul> | <br>Jake-424<br>Health interests: <ul style="list-style-type: none"><li>* Lowering cholesterol</li><li>* Stress reduction and relaxation</li><li>* Tobacco quitting and avoiding relapse</li></ul> | <br>Jeremy-388<br>Health interests: <ul style="list-style-type: none"><li>* Weight loss and dieting</li><li>* Lowering cholesterol</li><li>* Nutrition and meal planning</li><li>* Yoga and pilates</li></ul> |

Done

## Experiment design

- Some random nodes are reserved for researchers (the seed nodes)
- Respondents were assigned to other nodes and randomly to two conditions



- “The network typologies were created before the participants arrived, and the participants could not alter the typology in which they were embedded”
  - effectively removing the homophily/selection effect

## Behavior to diffuse

- Seed nodes will adopt some behaviors first
- Centola think these health behaviors requires multiple confirmation from friends; it's a type of complex contagion

**The Healthy Lifestyle Network - Mozilla Firefox**  
File Edit View History Bookmarks Tools Help  
<http://healthylifestyle.nis.harvard.edu/forum.php>      Google

User: John-672

### The Healthy Lifestyle Network

#### Community Forum

[Home](#) | [Healthy Lifestyle](#) | [Fitness](#) | [Nutrition](#) | [Smoking Cessation](#) | [Weight Loss](#)

Welcome to the community forum! This site provides recommended resources for finding out about tools and programs for improving your lifestyle. Please click on the links to view the sites, and provide ratings on their usefulness.

#### New Recommendations

**Nutrition Source**  
★★★★★ (rating of 4.3 out of 50 votes)  
Easy to understand state-of-the-art information about diet and nutrition from the department of nutrition at the Harvard School of Public Health

**Mayo Clinic Fitness Center**  
★★★★★ (rating of 3.4 out of 14 votes)  
Information on exercise basics, plans, overcoming fitness obstacles, and injury prevention.

**Tufts Nutrition Research**  
★★★★ (rating of 2.6 out of 14 votes)  
Current research on nutrition and diet.

#### Recommended Resources

**Nutrition Source**  
★★★★★ (rating of 4.3 out of 50 votes)  
Easy to understand state-of-the-art information about diet and nutrition from the department of nutrition at the Harvard School of Public Health

**American Cancer Society: Kick the Habit**  
★★★★★ (rating of 4.2 out of 25 votes)  
Provides both general information and specific guidelines for quitting smoking, including motivation for cessation, craving control, and finding the best way to quit.

**My Pyramid**  
★★★★ (rating of 4.1 out of 29 votes)  
Provides information to understand nutritional guidelines, and interactive tools and plans to apply the guidelines to daily life.

**Discovery Health Diet and Fitness Center**  
★★★★ (rating of 4.1 out of 21 votes)  
Information on dieting and fitness for weight loss and health, replete with tools, forums, and recipes.

**Harvard Vanguard**  
★★★★ (rating of 4.1 out of 24 votes)  
Information and advice about weight loss, diet, and nutrition from practicing physicians

Jake-424

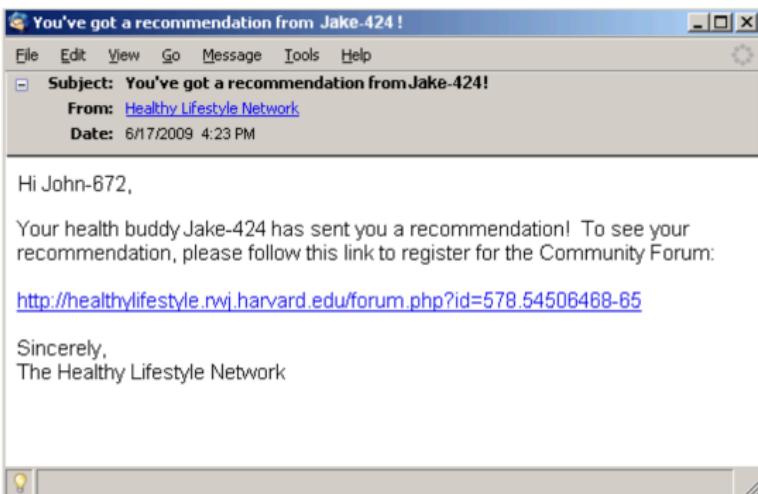
**Health Buddies Recommend**

**Nutrition Source**  
★★★★ (rating of 4.0)  
Easy to understand state-of-the-art information about diet and nutrition from the department of nutrition at the Harvard School of Public Health

**Done**

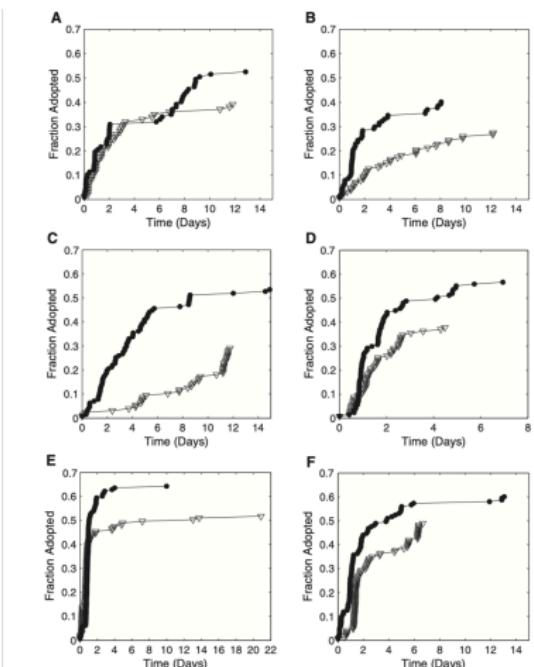
## Signals

- Each time a participant adopted the vbehavior, messages were sent to hear health buddies inviting them to adopt



# Results

- As expected, contagion is faster on regular networks



**Fig. 2.** Time series showing the adoption of a health behavior spreading through clustered-lattice (solid black circles) and random (open triangles) social networks. Six independent trials of the study are shown, including (A)  $N = 98$ ,  $Z = 6$ , (B to D)  $N = 128$ ,  $Z = 6$ , and (E and F)  $N = 144$ ,  $Z = 8$ . The success of diffusion was measured by the fraction of the total network that adopted the behavior. The speed of the diffusion process was evaluated by comparing the time required for the behavior to spread to the greatest fraction reached by both conditions in each trial.

## What to like about this research

- Controls the effects of network topology, independent of factors such as homophily, geographic proximity
- “Study the spread of a health-related behavior that is unknown to the participants before the study, thereby eliminating the effects of nonnetwork factors from the diffusion dynamics, such as advertising, availability, and pricing”
- “allows the same diffusion process to be observed multiple times, under identical structural conditions”, thus being more robust
- What are the potential limitations?

## Take-away messages

- To make explanatory arguments (e.g., social influences causes high correlation between attributes of friends)
- Experimental approach usually provides cleaner analysis and more powerful results, compared with statistical analysis on observational data
- But experiments are harder to implement
  - Online field experiment: best if you are insider or know someone in the company
    - Even companies do not want to publish these type of research (though they are still running these experiments everyday)
  - Online lab experiment: Centola spent years to build the website he used