Network Analysis:

The Hidden Structures behind the Webs We Weave

17-213 / 17-668

Diffusion and Contagion

Tuesday, December 5, 2023

Patrick Park & Bogdan Vasilescu

Carnegie Mellon University
School of Computer Science

S3D
Software and Societal Systems Department

2-min Quiz, on Canvas

Information Diffusion

Spread the word: Viral marketing

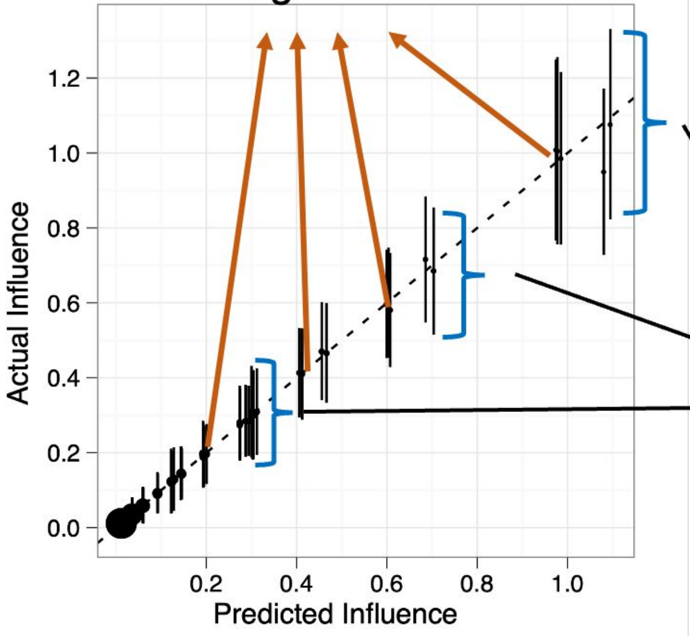
Question: Who should you target in a network to "maximize" information cascades?

- 74M separate diffusion events (Twitter retweets of URLs)
- Influence of the seed node: # of nodes in the diffusion tree
- Seed node's attributes (followers, friends, tweets) and previous success of the seed node most predictive of average influence scores of the leaf nodes (clusters) in the regression tree

Answer: Hard to predict

Diffusion is difficult to predict

Highly accurate predictions of within-leaf average influence scores



-Regression tree model not so predictive of individual influence scores
-Weak effect of the nature of the content
-With these “null” results, the paper pivots to asking a slightly different question: Who should you target to “optimize” information cascades (i.e., introduce cost constraint)?

Variance within leaf (cluster) too large to predict individual influence scores

Structural virality of diffusion

How do information cascades look like?

Structural virality (Wiener index)

- Average path length in a diffusion tree

$$\nu(T) = \frac{1}{n(n-1)} \sum_{i=1}^{n} \sum_{j=1}^{n} d_{ij}$$

Recall

In a tree

Structural virality of diffusion

Structural virality (Wiener index)

- Average path length in a diffusion tree

$$\nu(T) = \frac{1}{n(n-1)} \sum_{i=1}^{n} \sum_{j=1}^{n} d_{ij}$$

Structural virality of diffusion

Structural virality (Wiener index)

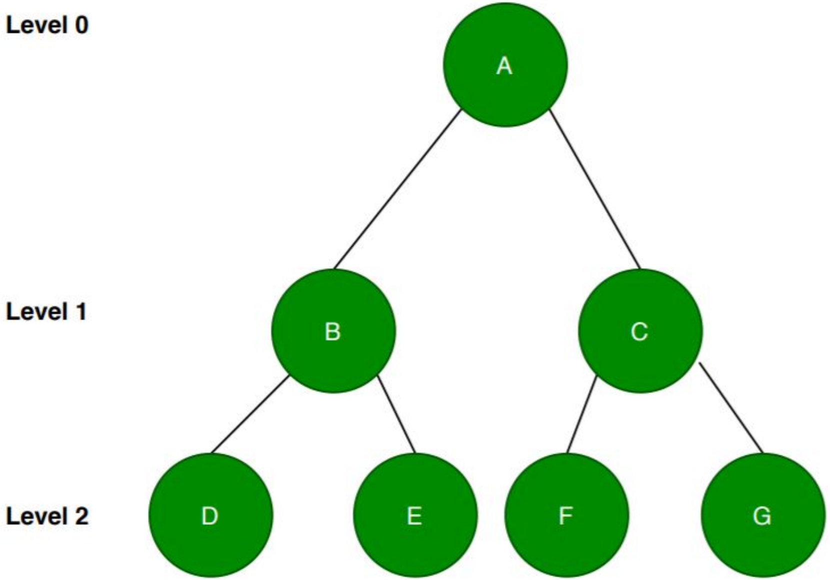
- Average path length in a diffusion tree

$$\nu(T) = \frac{1}{n(n-1)} \sum_{i=1}^{n} \sum_{j=1}^{n} d_{ij}$$

Recall, *d* ~ Ln(*N*) / Ln<*k*>

In a complete binary tree

- N=2^0+2^1+...2^h
- Ln(N) ~ h * Ln(2)
- h ~ Ln(N) / Ln(2) → <k> = 2
- h ~ Ln(N) / Ln<k>
- *d* ~ *h*



Structural virality of diffusion

Examples of information cascade trees in increasing order of virality

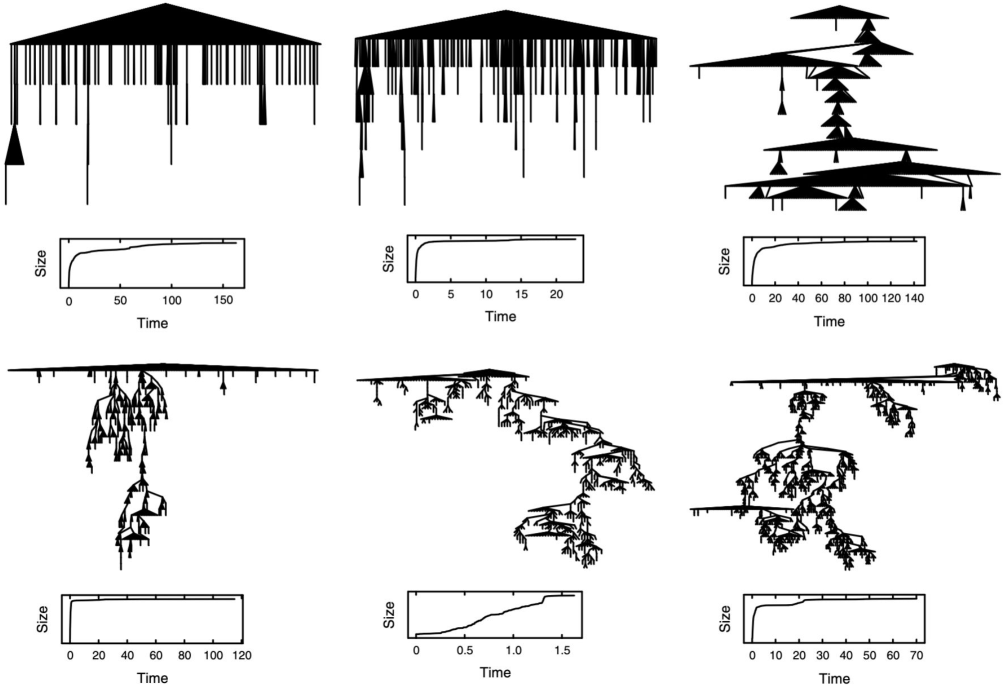
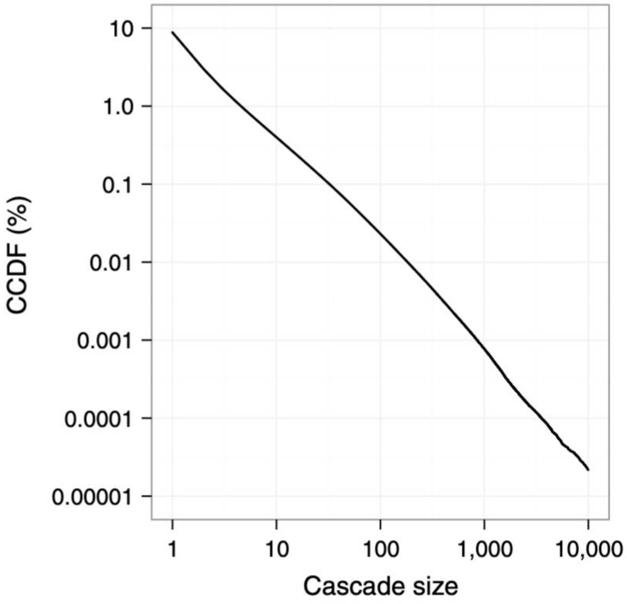


Figure 2 **Distribution of Cascade Sizes on a Log–Log Scale, Aggregated Across the Four Domains We Study: Videos, News, Pictures, and Petitions**

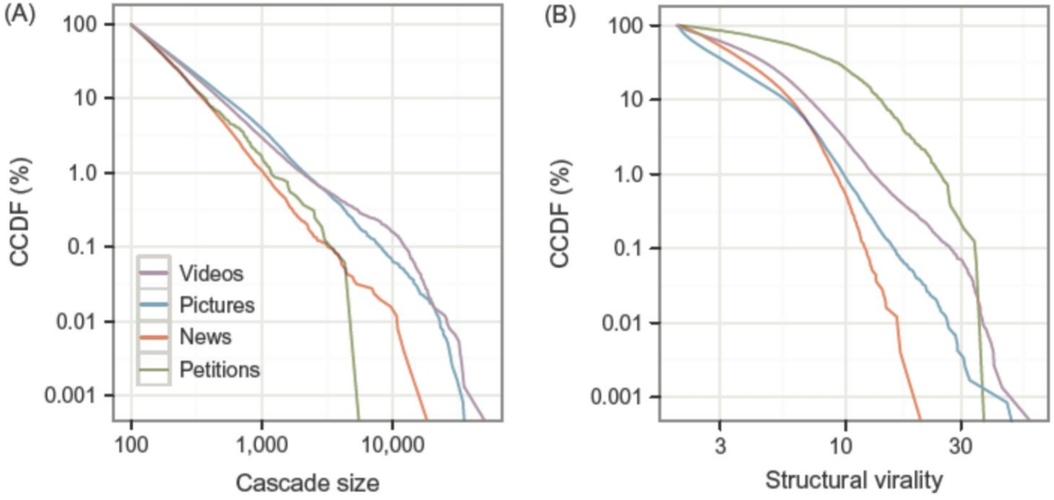


Note. CCDF, complementary cumulative distribution function.

Structural virality of diffusion

Does structural virality correlate
with cascade size?

Figure 4 **Size and Structural Virality Distributions on a Log–Log Scale for Cascades Containing at Least 100 Adopters, Separated by Domain**



Note. CCDF, complementary cumulative distribution function.

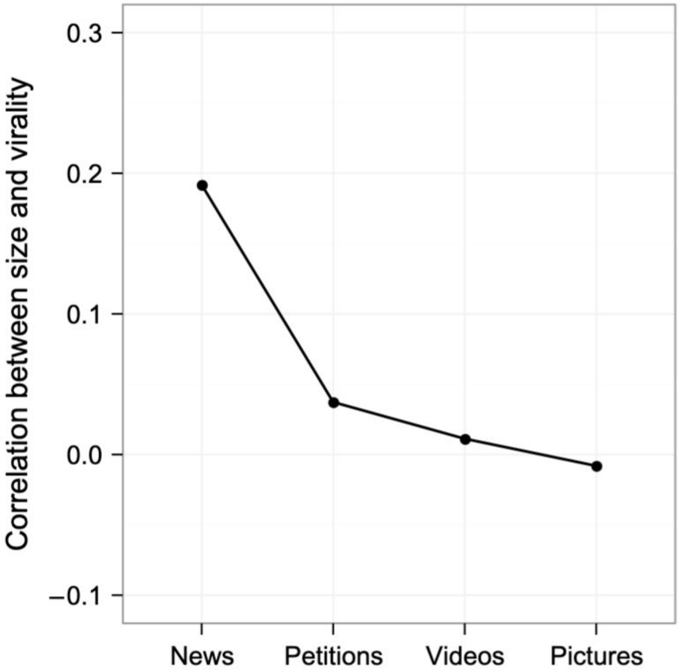
Structural virality of diffusion

Does structural virality correlate with cascade size?

- Not really

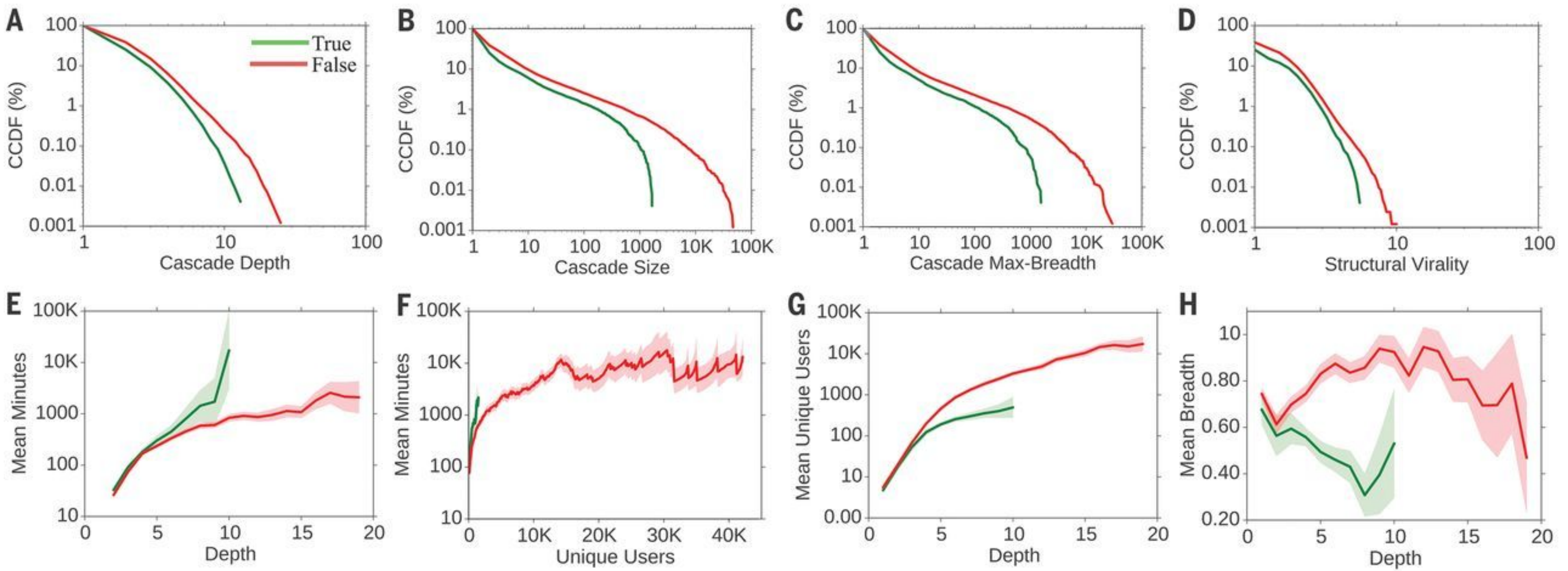
Predicting mass information diffusion is hard

Figure 6 **Correlation Between Cascade Size (Popularity) and Structural Virality Across Four Domains**



True vs. False information diffusion

False news diffuses much faster, reaches broader audience, and penetrates more deeply



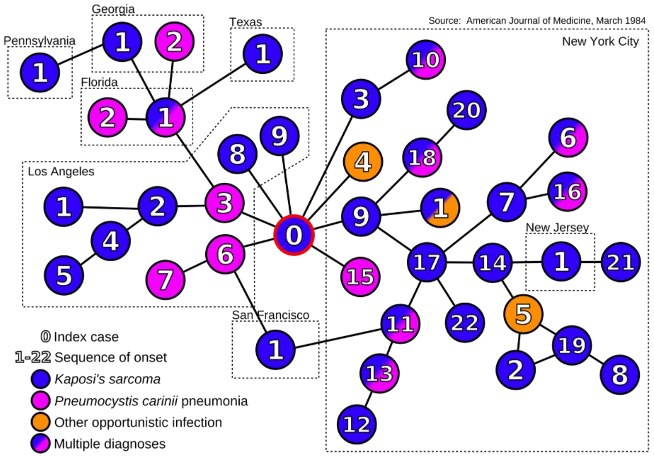
Social Contagion: Costly spreading

Simple Contagion

A single contact leads to adoption/contagion (e.g., virus)

Spreads quickly in networks with low CPL (e.g., small-world)

Individual with a diverse egonetwork can “infect” disproportionately (e.g., super spreaders)



Dynamics of Behavioral Change

Model the effect of network structure on the spread and adoption of behaviors through network ties

Three Mechanisms of social adoption
-Common environmental influence

-Homophily (e.g., similar taste)

-Social influence

-

Very difficult to disentangle these mechanisms with observational data

(e.g., Framingham study of the spread of obesity)

Dynamics of Behavioral Change

Identification strategy: **experimental** approach

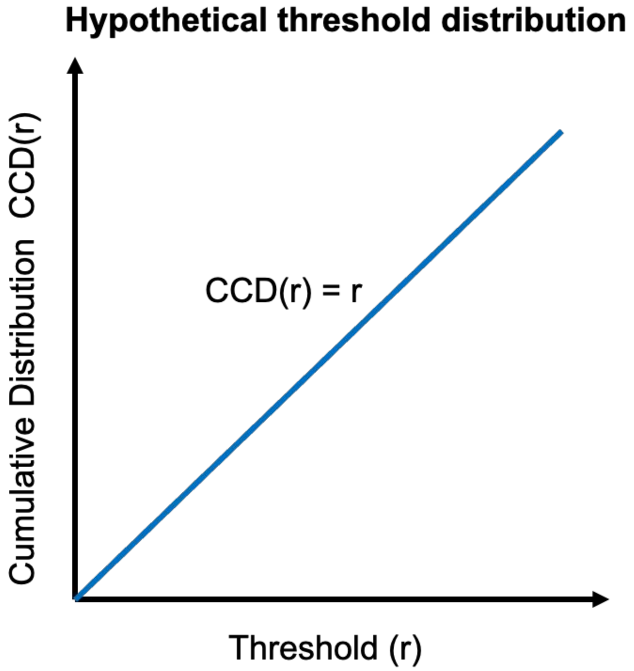
- Create two separate worlds, with and without social influence
- Observe adoption behavior in the two worlds
- Example: The Music Lab experiment

Dynamics of Behavioral Change

Weak influence condition

Strong influence condition

Threshold models of adoption



Some social behaviors require more than single exposure for adoption

- Individuals can have different levels of reluctance/resistance (thresholds)

- Variance in norms, preferences, utility lead to a distribution of thresholds

- Toy example: If an initial adoption occurs, adoption will reach 100% (saturation)

Threshold models of adoption

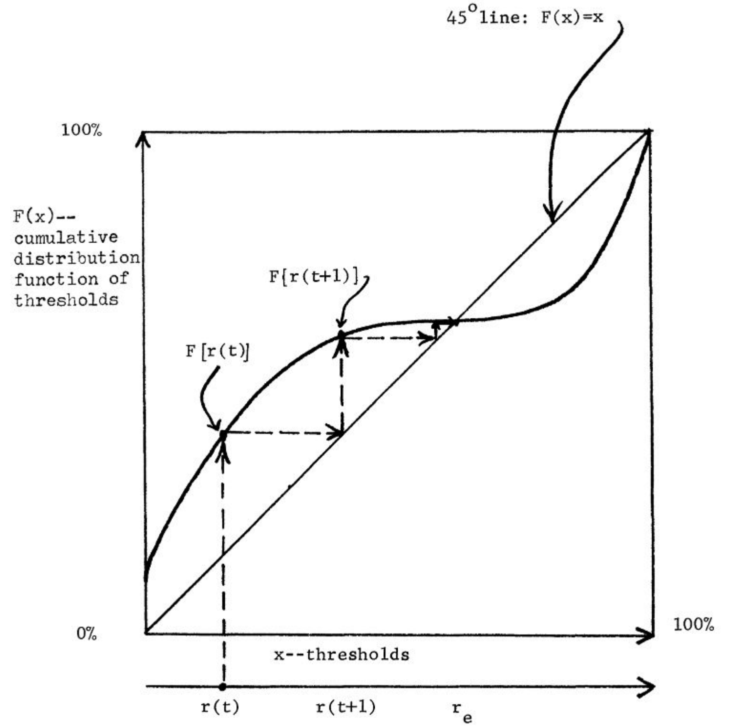


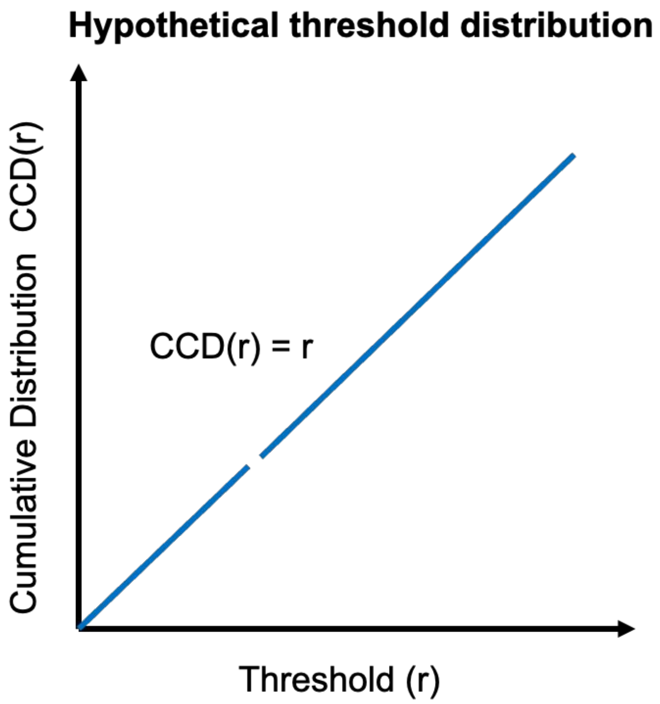
FIG. 1.—Graphical method of finding the equilibrium point of a threshold distribution. $r(t)$ = proportion having rioted by time t.

Some social behaviors require more than single exposure for adoption

- Assumption 1: People have perfect information about adoption at time *t*

- Assumption 2: Individual's threshold pertains to population adoption, not local adoption

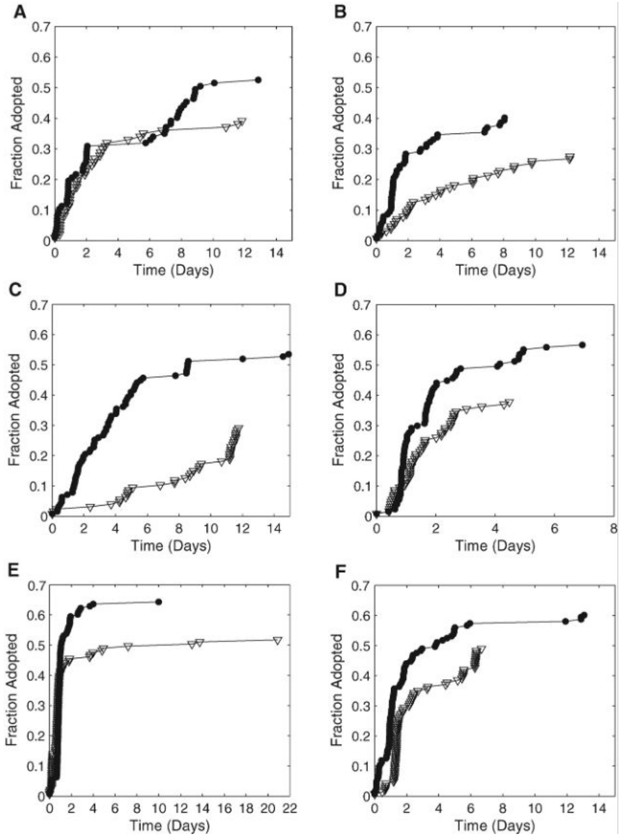
Threshold models of adoption



Sensitivity of collective behavior

- A negligible change to the threshold distribution can lead to vastly different equilibria

Complex Contagion



Adoption/infection probability increases with the number of neighbors who already adopted

Builds on the ideas of thresholds and social reinforcement

Initially studied as a simulation model (Centola and Macy 2007)

Centola reproduced the results through real-world experiments

Complex Contagion

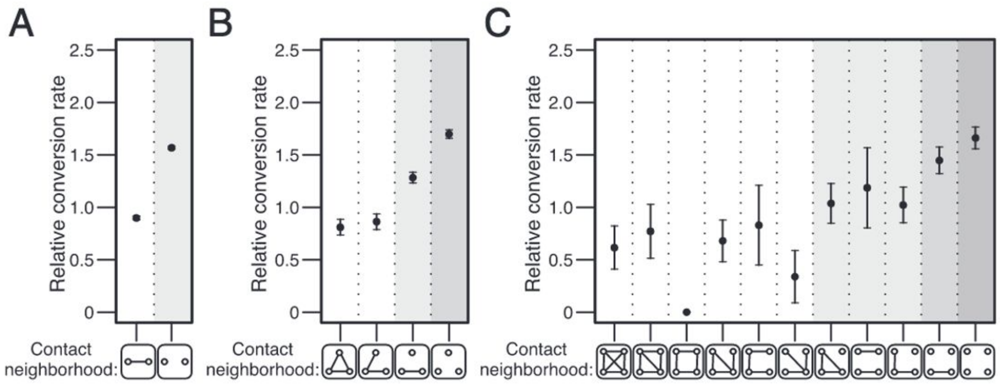


Fig. S2. Recruitment conversion for demographically homogeneous neighborhoods, as a function of (*A*) two-node, (*B*) three-node, and (*C*) four-node contact neighborhood graphs. The conversion scale is the same as for Fig. 1 in the main text. Error bars represent 95% confidence intervals.

Ugander et al. 2012

Open questions:

For a focal individual, is a closed or open triad more conducive to social contagion? (e.g., Facebook adoption study)

Complex Contagion

Social contagion is an endogenous process:

- Homophily → adoption
- Embeddedness → adoption
- Tie strength → adoption

Similar people form strong ties

Embedded relations tend to be strong ties

Tie strength can potentially increase similarity

Tie strength can generate embedded relations

Result: **Difficult to estimate causal effect on adoption**

Summary

-