

Strata Conference

Understanding Social Contagion

Marcel Salathé, PhD

Center for Infectious Disease Dynamics (CIDD) | PSU

Department of Biology

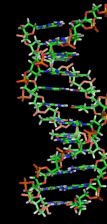
Department of Engineering and Computer Sciences

salathegroup.com @salathegroup

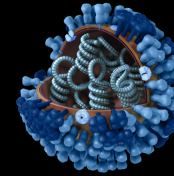
"Hire biologists! It doesn't
make sense for a high tech
company to have 3,000
software engineers but no
biologists."

George Dyson in WIRED Mar '12

Spread of genes

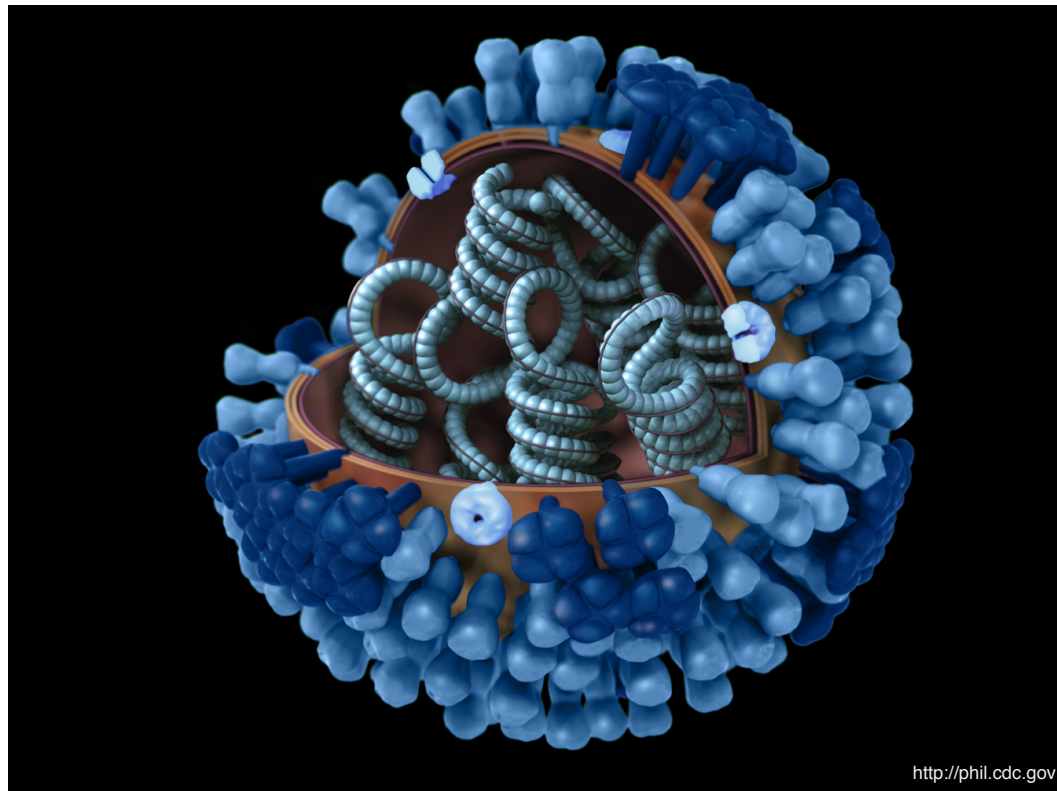


Spread of pathogens



Spread of memes

meme concept



<http://phil.cdc.gov>



Social contagion

The spread of behaviors

perhaps explain meme



<http://phil.cdc.gov>



<http://phil.cdc.gov>, <http://www.wikipedia.org>

WIRED

SUBSCRIBE »

SECTIONS »

BLOGS »

REVIEWS »

VIDEO »

HOW-TO »

Sign In | RSS Feeds



PHOTO GALLERY

The Ice Age Miracle
From Miracle Mile



VERSION CONTROL

How Wired Published
Its GitHub Story on
GitHub

FEATURES

The Jerusalem Syndrome:
Why Some Religious
Tourists Believe They Are
the Messiah

START

Proper Spelling? Its Tyme to
Let Luce!

PLAY

Win at First-Person
Shooters, the Coward's Way

/ MAGAZINE

FEATURES 17.11

An Epidemic of Fear: How Panicked Parents Skipping Shots Endangers Us All

By Amy Wallace | October 19, 2009 | 3:00 pm | Wired Nov 2009



To hear his enemies talk,
you might think Paul Offit is
the most hated man in

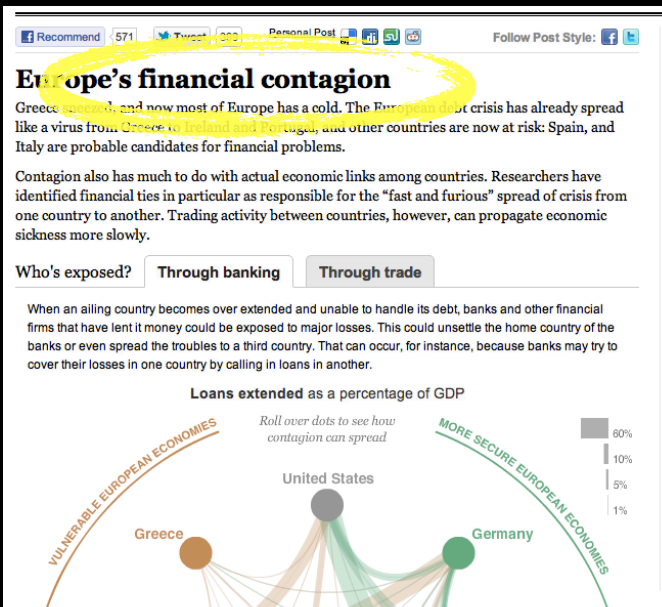
65

1

Tweet

+1

Share



The Economist[Log in](#) [Register](#) [Subscribe](#) [Digital & mobile](#) [Events](#) [Topics A-Z](#) [Newsletters](#) [Jobs](#)

Sunday February 26th 2012

[World politics](#) [Economics](#) [Science & technology](#) [Culture](#) [The World in 2012](#) [Blogs](#) [Debate](#) [Multimedia](#)

Social media in the 16th Century
How Luther went viral
Five centuries before Facebook and the Internet, social media helped bring about the Reformation

Dec 17th 2011 | from the print edition [Like](#) 4k [Tweet](#) 3,856



Recommendations
[Sign Up](#) Create an account or [log in](#) to see what your friends are recommending

 **How an economist says**
825 people recommend this.

 **What's America's real crime rate?**
89 people recommend this.

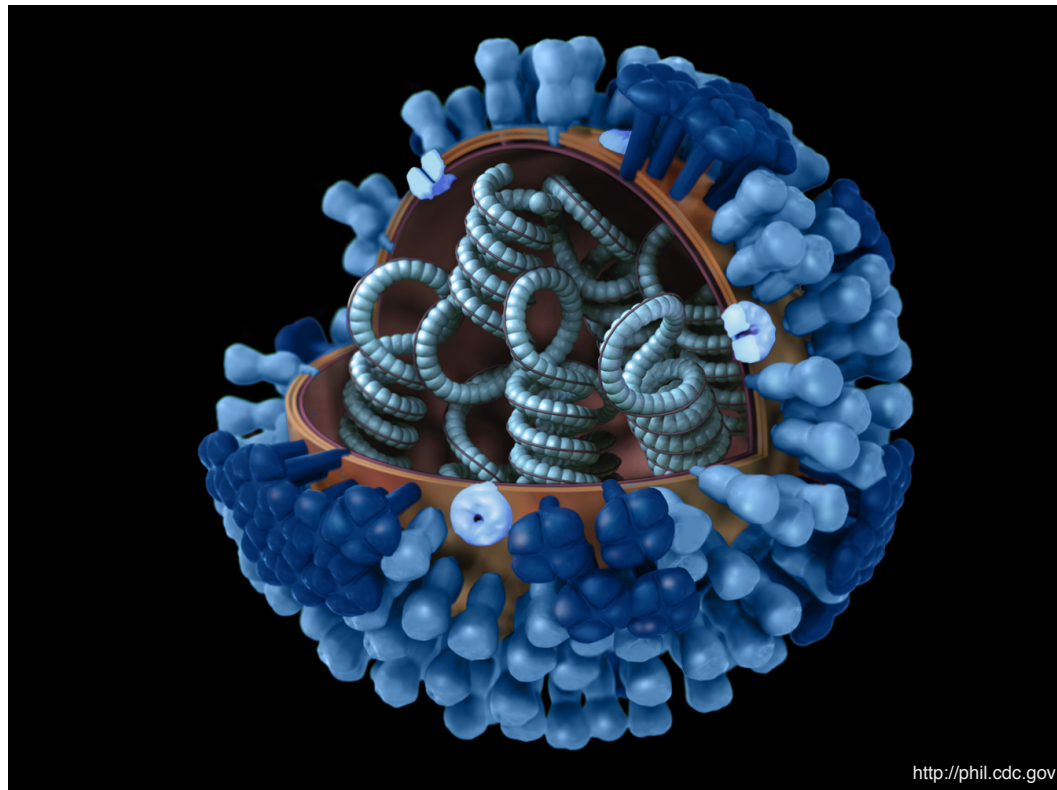
 **The comedic potential of "literally"**
61 people recommend this.

 **ZuRich**
1,668 people recommend this.

[Facebook social plugin](#)

Most commented **Most recommended**

1. Nuclear proliferation: Bombing Iran
2. Official statistics: Don't lie to me, Argentina
3. China's new sports problems: China's...



<http://phil.cdc.gov>

Does social contagion exist?

1. Personal experience (peer influence)
2. Temporal dynamics
3. Social dynamics

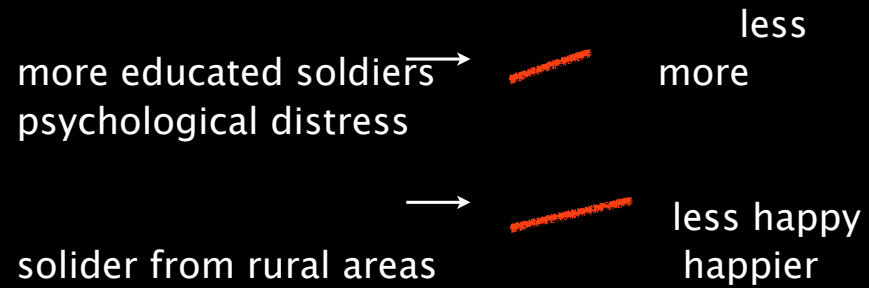
Just so stories...

survey of 600,000 soldiers in WWII:

more educated soldiers	→	more
psychological distress		
	→	
soldier from rural areas		happier

Just so stories...

survey of 600,000 soldiers in WWII:

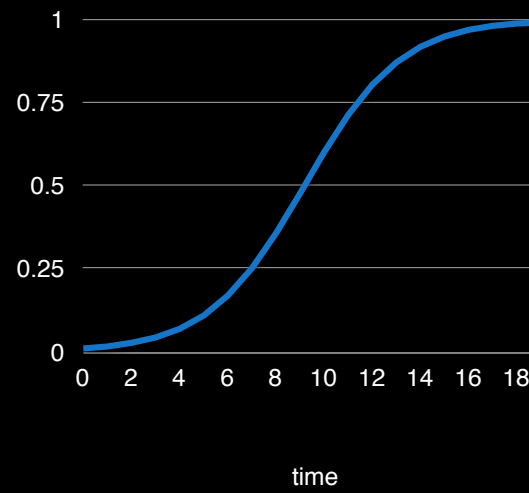


Just so stories...

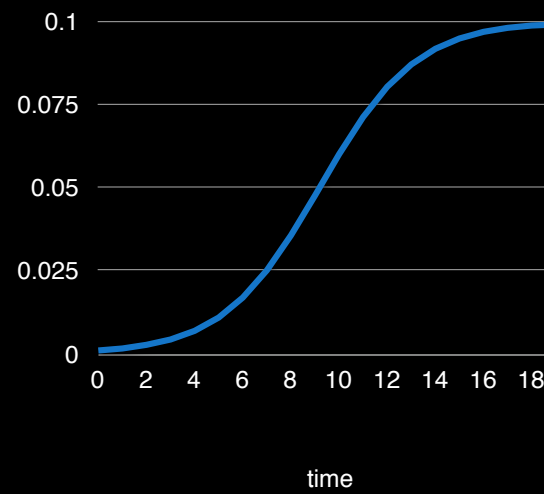
survey of 600,000 soldiers in WWII:



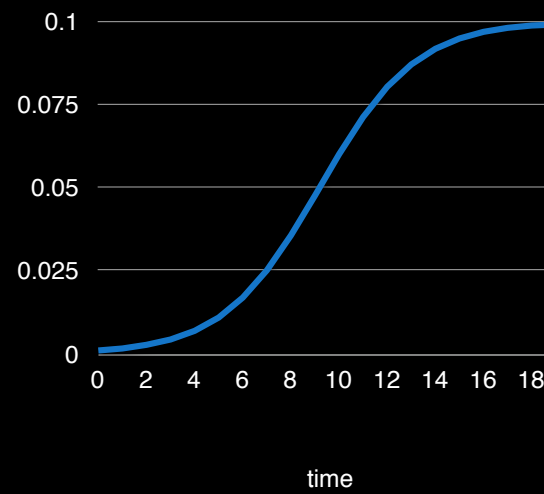
Diffusion



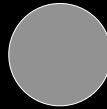
Example 1: influenza



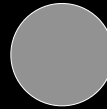
Example 2: sunburn



Contagion



A

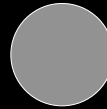


B

Contagion



A



B

Contagion



A



B

NOT Contagion

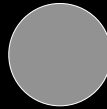


A

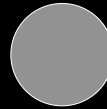


B

Contagion

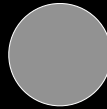


A



B

NOT Contagion



A

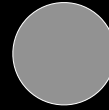


B

Contagion



A

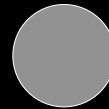


B

NOT Contagion



A



B

Contagion



A



B

NOT Contagion

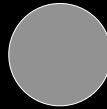


A

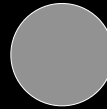


B

Contagion

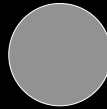


A

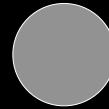


B

Contagion



A

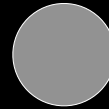


B

Contagion



A



B

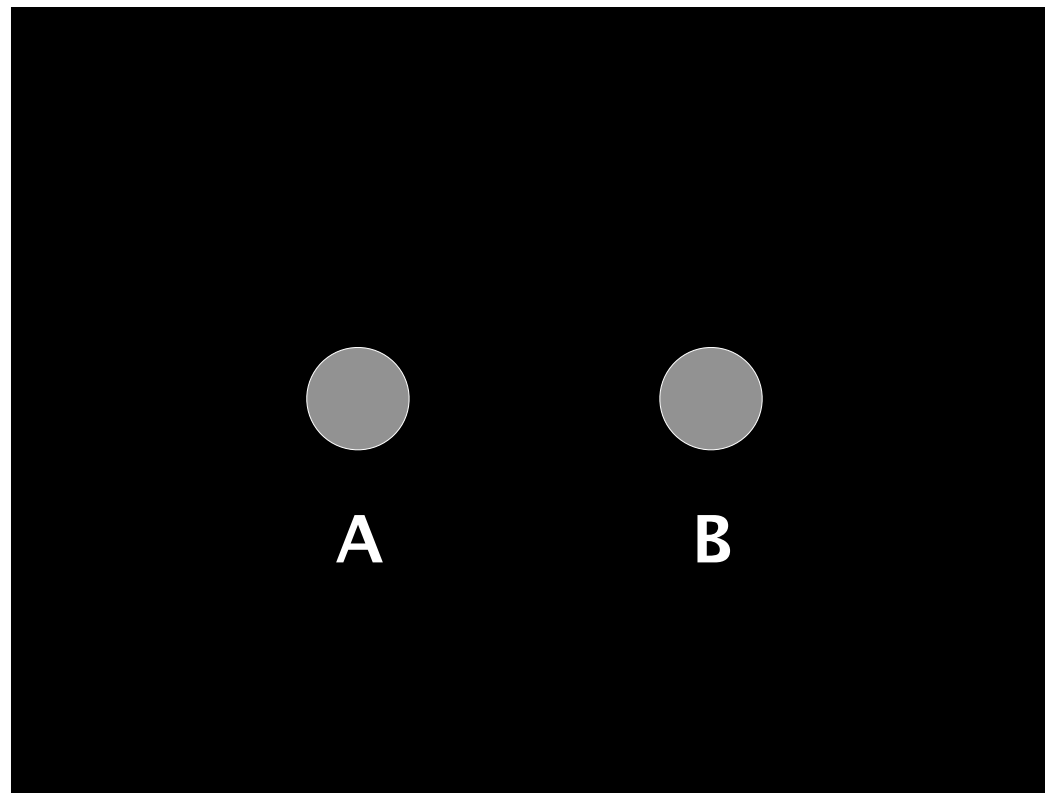
NOT Contagion



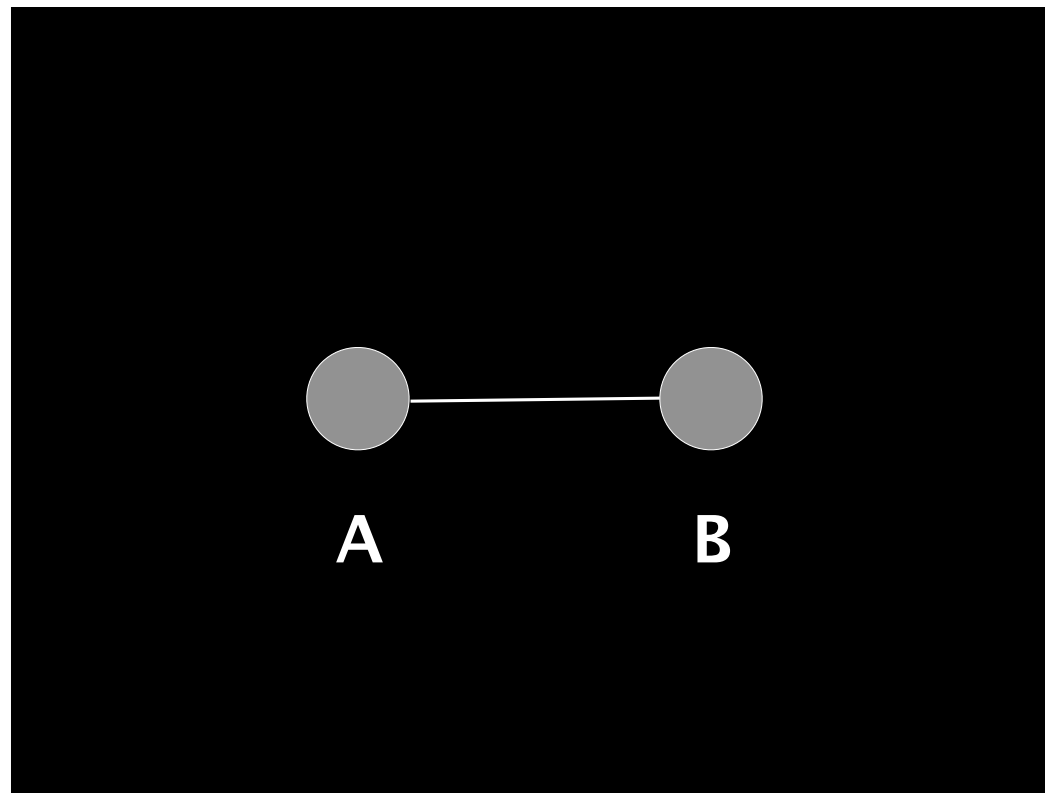
A



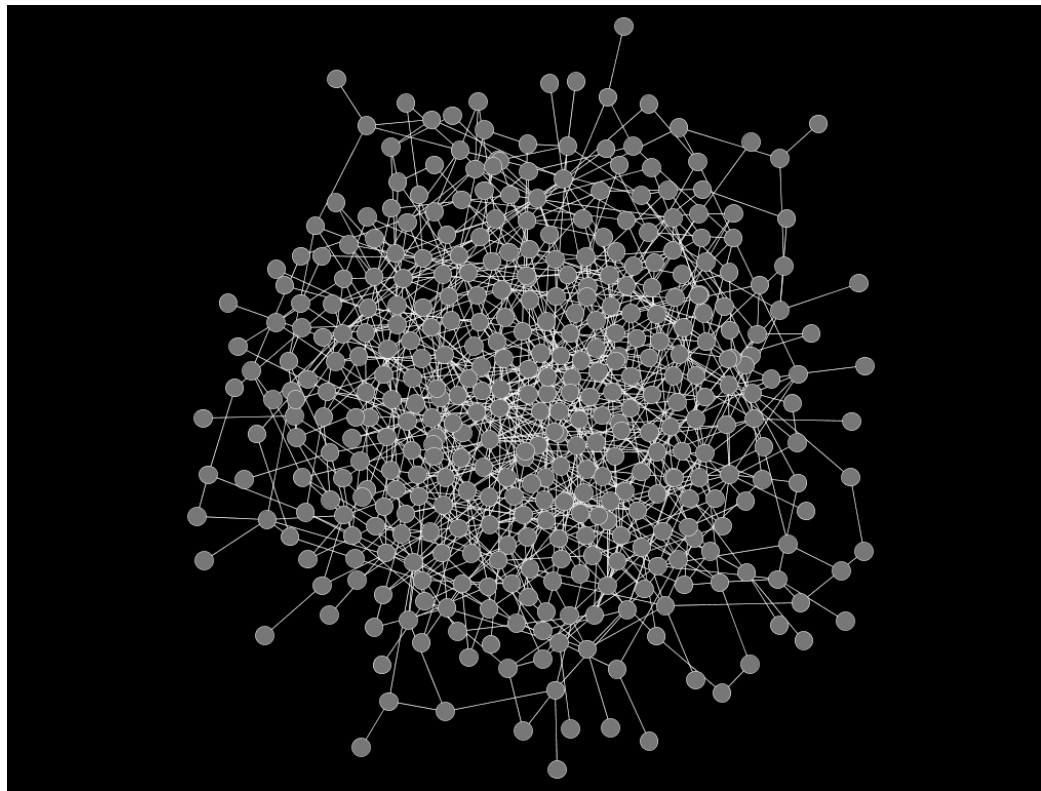
B



person to person: not just random, but friends, family, people with higher status



visually connect two people if they are somehow connected



Ties can be friendships, close contacts, digital ties (facebook friendships, twitter follower relationships), sexual etc.

If we know that it is a contagious process, we can try to connect individuals according to whether there is a chance of contagion

Selection vs. influence

show a graph with clustering (by pressing space twice)

Influence:

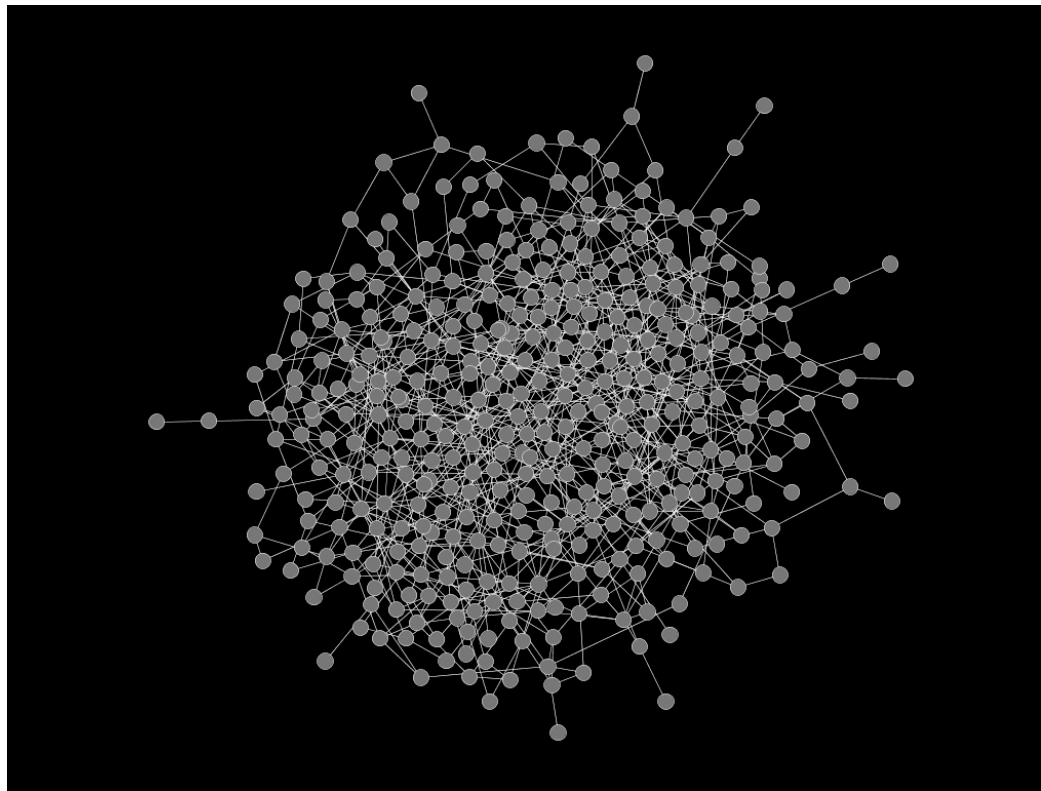


are two people alike BECAUSE they are connected

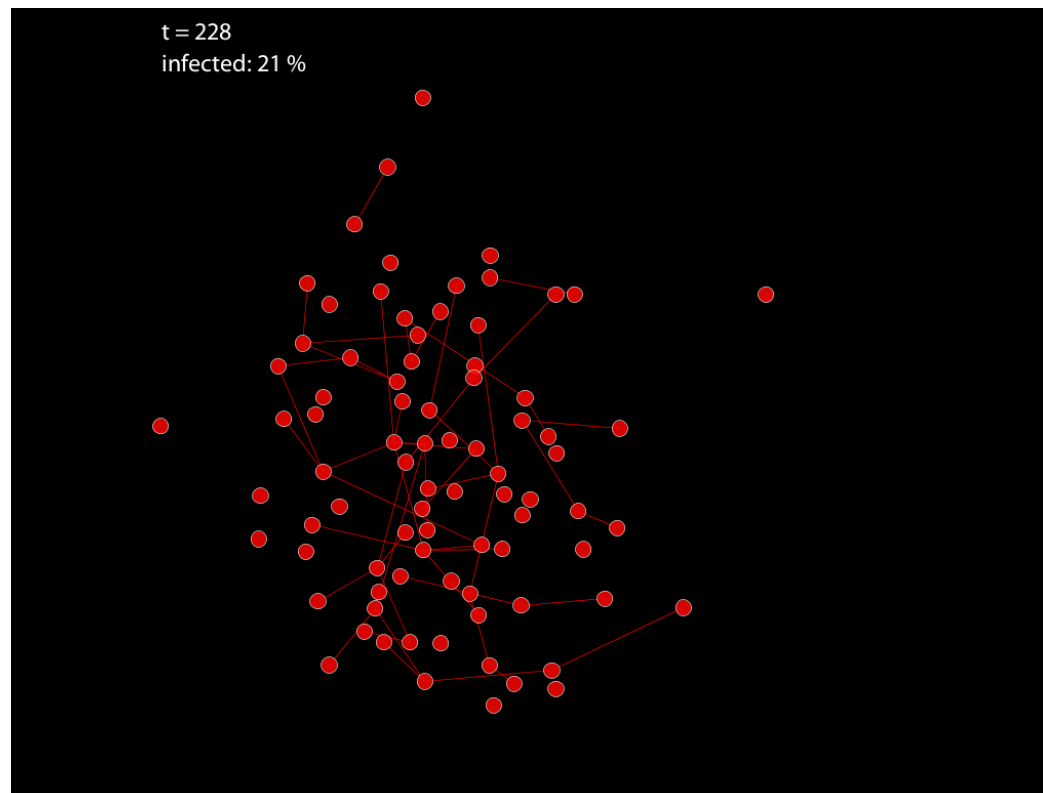
Selection:



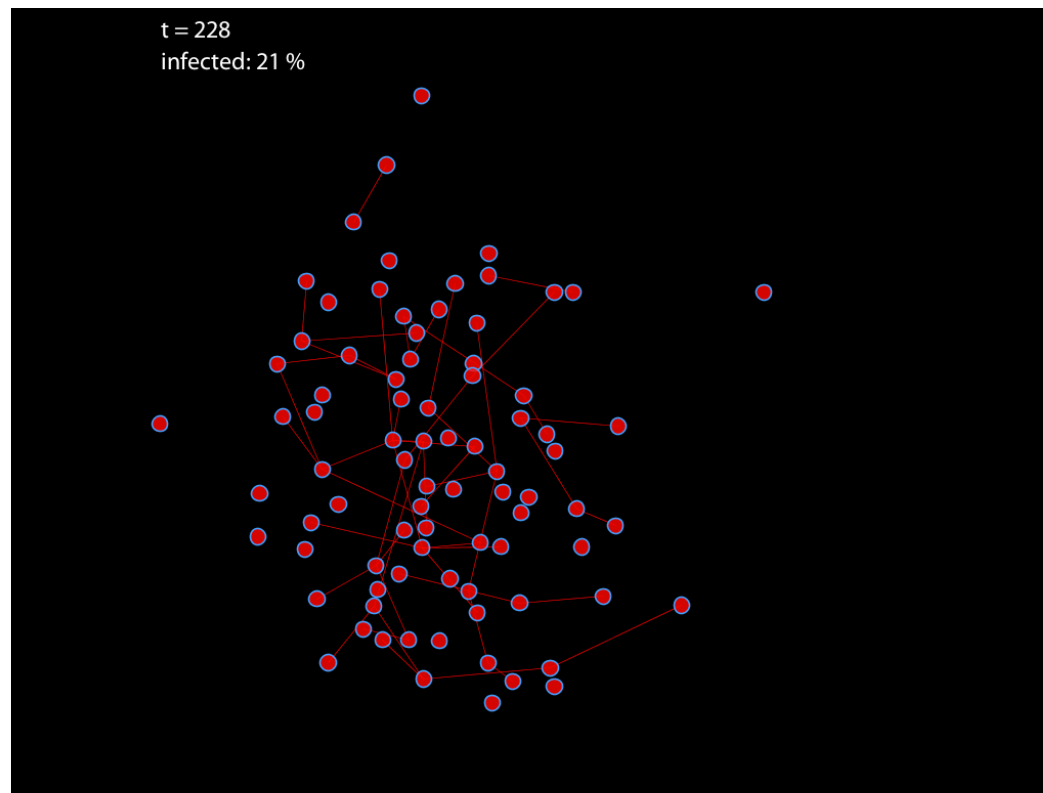
or are two people connected BECAUSE they are alike?



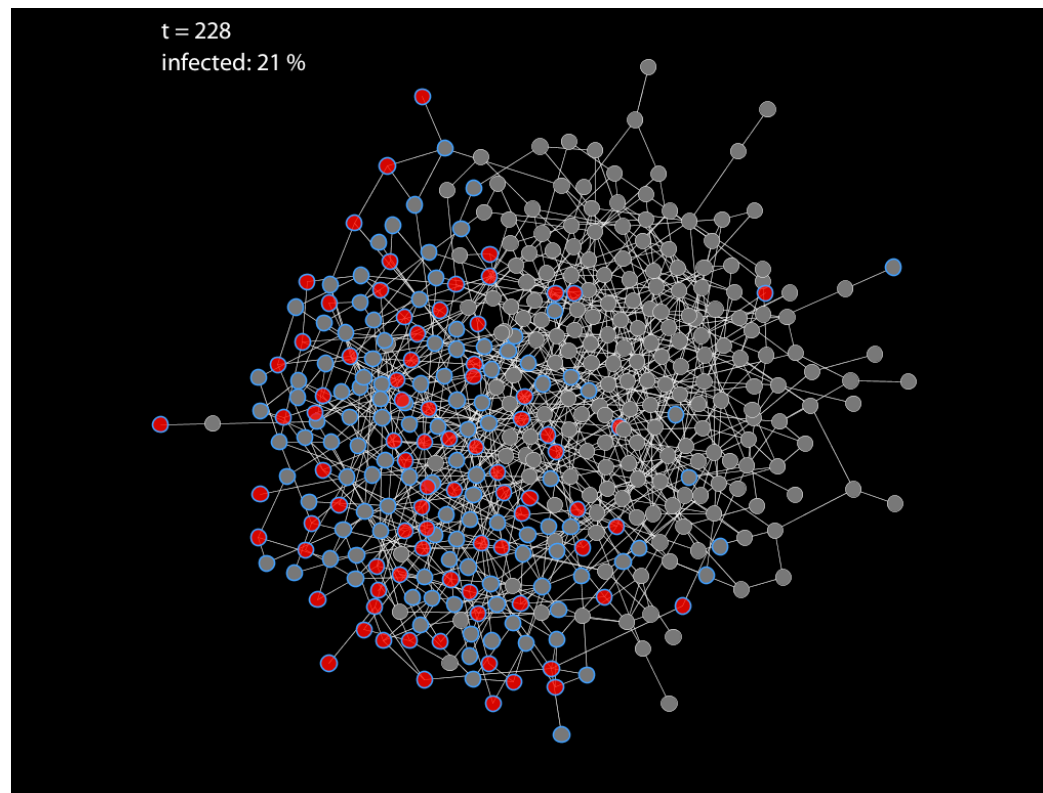
despite network clustering, the process here is random exposure with conditional adoption



despite network clustering, the process here is random exposure with conditional adoption



despite network clustering, the process here is random exposure with conditional adoption



despite network clustering, the process here is random exposure with conditional adoption

Complex Contagions and the Weakness of Long Ties¹

Damon Centola
Harvard University

Michael Macy
Cornell University

The strength of weak ties is that they tend to be long—they connect socially distant locations, allowing information to diffuse rapidly. The authors test whether this “strength of weak ties” generalizes from simple to complex contagions. Complex contagions require social affirmation from multiple sources. Examples include the spread of high-risk social movements, avant garde fashions, and unproven technologies. Results show that as adoption thresholds increase, long ties can impede diffusion. Complex contagions depend primarily on the width of the bridges across a network, not just their length. Wide bridges are a characteristic feature of many spatial networks, which may account in part for the widely observed tendency for social movements to diffuse spatially.

despite network clustering, the process here is random exposure with conditional adoption

The Spread of Behavior in an Online Social Network Experiment

Damon Centola

How do social networks affect the spread of behavior? A popular hypothesis states that networks with many clustered ties and a high degree of separation will be less effective for behavioral diffusion than networks in which locally redundant ties are rewired to provide shortcuts across the social space. A competing hypothesis argues that when behaviors require social reinforcement, a network with more clustering may be more advantageous, even if the network as a whole has a larger diameter. I investigated the effects of network structure on diffusion by studying the spread of health behavior through artificially structured online communities. Individual adoption was much more likely when participants received social reinforcement from multiple neighbors in the social network. The behavior spread farther and faster across clustered-lattice networks than across corresponding random networks.

Abstract

despite network clustering, the process here is random exposure with conditional adoption

Distinguishing influence-based contagion from homophily-driven diffusion in dynamic networks

Sinan Aral^{a,b,1}, Lev Muchnik^a, and Arun Sundararajan^a

^aInformation, Operations and Management Sciences Department, Stern School of Business, New York University, Kaufmann Management Center, 44 West 4th Street, New York, NY 10012; and ^bCenter for Digital Business, Sloan School of Management, Massachusetts Institute of Technology, 5 Cambridge Center-NE25, Cambridge, MA 02142

Edited by Matthew O. Jackson, Stanford University, Stanford, CA, and accepted by the Editorial Board October 6, 2009 (received for review August 4, 2009)

Node characteristics and behaviors are often correlated with the structure of social networks over time. While evidence of this type of assortative mixing and temporal clustering of behaviors among linked nodes is used to support claims of peer influence and social contagion in networks, homophily may also explain such evidence. Here we develop a dynamic matched sample estimation framework to distinguish influence and homophily effects in dynamic networks, and we apply this framework to a global instant messaging network of 27.4 million users, using data on the day-by-day adoption of a mobile service application and users' longitudinal behavioral, demographic, and geographic data. We find that previous methods overestimate peer influence in product adoption decisions in this network by 300–700%, and that homophily explains >50% of the perceived behavioral contagion. These findings and methods are essential to both our understanding of the mechanisms that drive contagions in networks and our knowledge of how to propagate or combat them in domains as diverse as epidemiology, marketing, development economics, and public health.

As more of a perceived contagion is explained by homophily rather than peer influence, intervention strategies should shift from peer-to-peer methods based on network structure to outreach based on population segmentation across individuals' characteristics. Formal procedures for separating influence and homophily are therefore essential to support policies that encourage or discourage the spread of behaviors in networks, from health interventions to viral marketing campaigns.

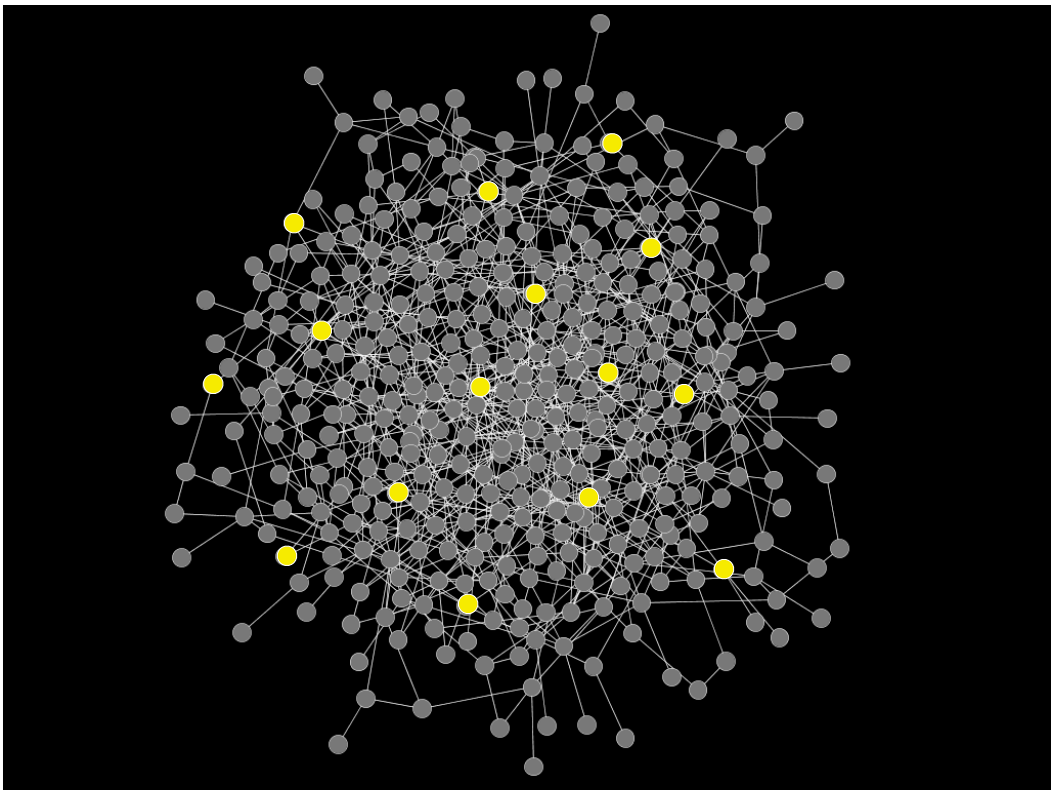
Contagions and homophilous diffusion are both typified by correlations between network structure and individual outcomes over time (1–3, 5–11, 17, 21, 22). Two empirical patterns have been used to substantiate claims of peer influence and contagion in networks (*i*) *assortative mixing*—correlations of behaviors among linked nodes (23, 24)—and (*ii*) *temporal clustering*—temporal interdependence of behaviors among linked nodes (12–14, 25–27). Because peer influence is likely to lead to assortative mixing, some studies claim assortative mixing is evidence of peer influence (12–14, 25–27). Evidence of temporal clustering is used to corroborate these claims because as Anastopoulos et al. (25) argue “if

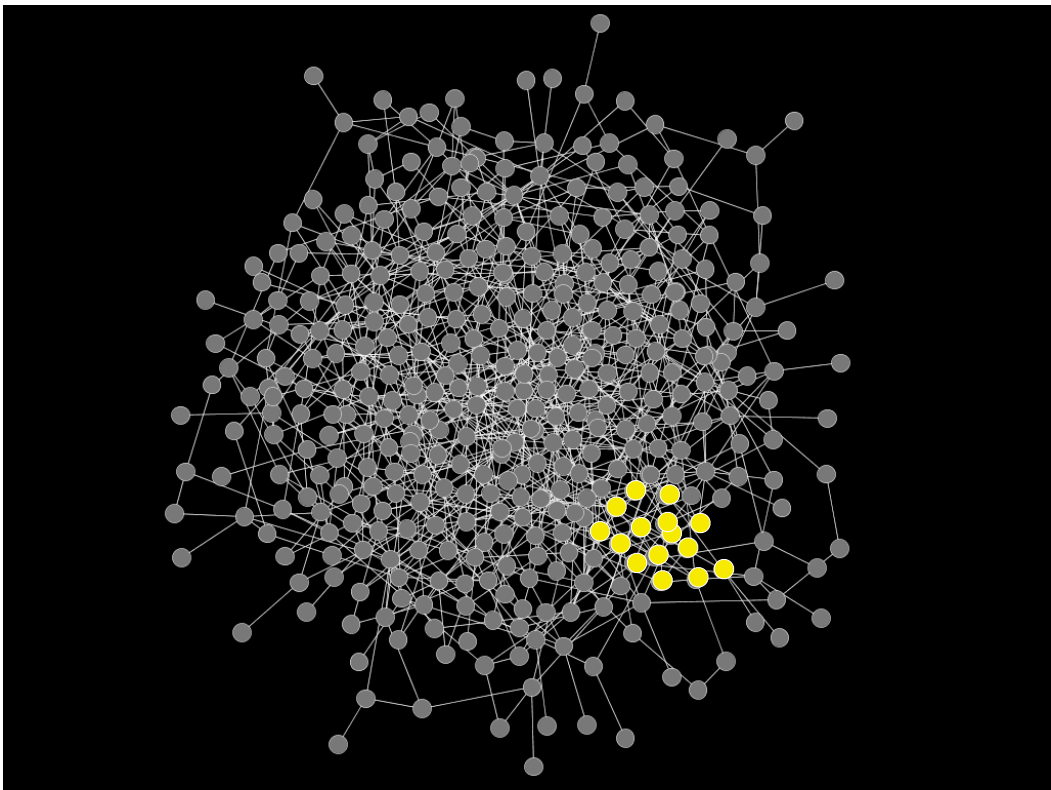
despite network clustering, the process here is random exposure with conditional adoption

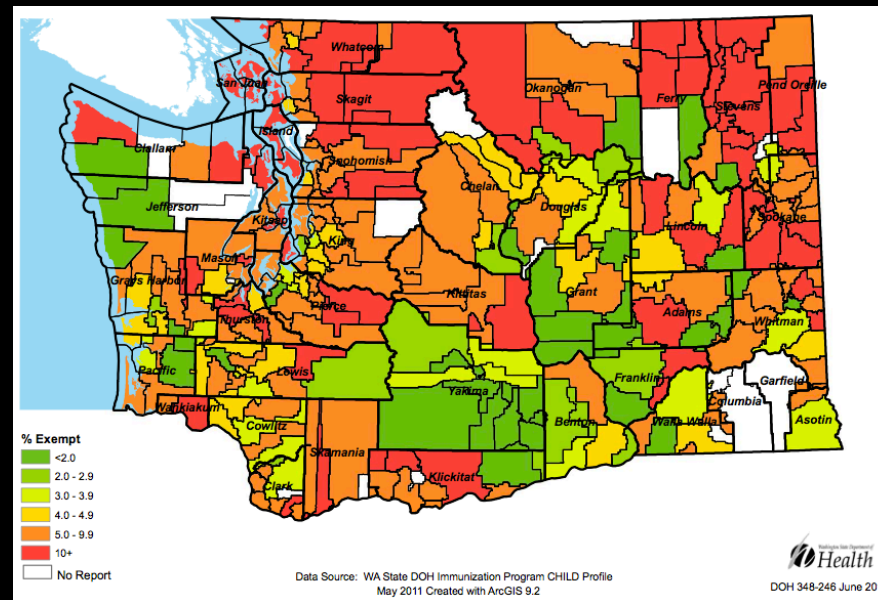
H1N1 vaccine Twitter study



<http://phil.cdc.gov>







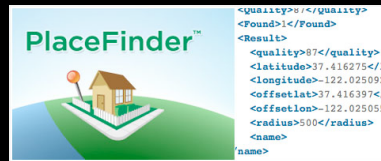
<http://www.doh.wa.gov/cfh/immunize/documents/k12exemptsd10-11.pdf>



Twitter Project: Rate This Tweet

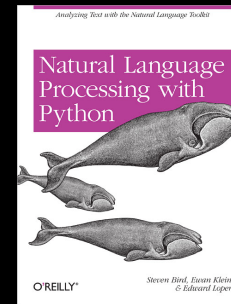
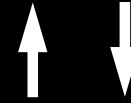
News Flu H1N1: Finland to vaccinate entire population against A/H1N1 flu - Xinhua
<http://ow.ly/164wfj>

[cc](#) [redo](#) [previous](#)



Result data set

sentiment score per user & location,
August 2009 – January 2010



name	age	message	history
 analysis	4 months ago	Initial public commit [salatheproduction]	
 student_ratings	4 months ago	Initial public commit [salatheproduction]	
 .gitignore	4 months ago	Initial public commit [salatheproduction]	
 README	4 months ago	Initial public commit [salatheproduction]	

README

This is all the source code used for the following paper:

Salathé M, Khandelwal S, 2011 Assessing Vaccination Sentiments with Online Social Media: Implications for Infectious Disease Dynamics and Control. PLoS Comput Biol 7(10): e1002199. doi:10.1371/journal.pcbi.1002199

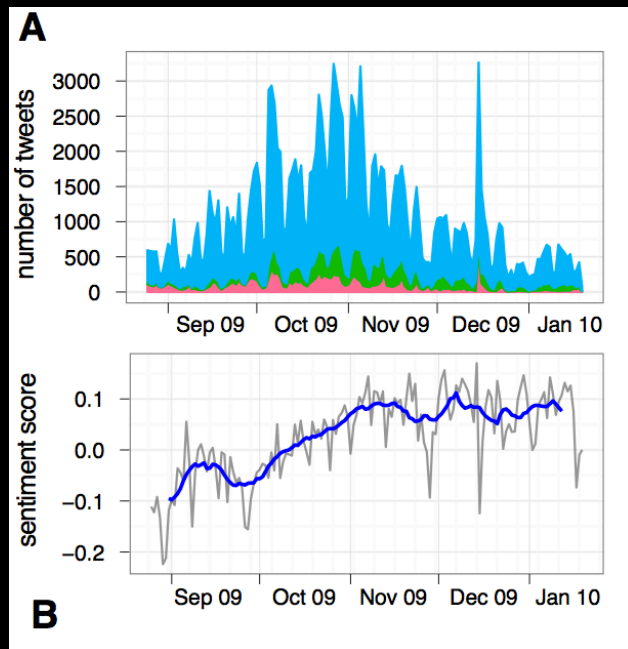
Please cite the paper above, if you use our code or create a derivative work.

The analysis/ directory contains all the code for the sentiment analysis (based on NLTK/MegaM). It also containst the geocoding code.

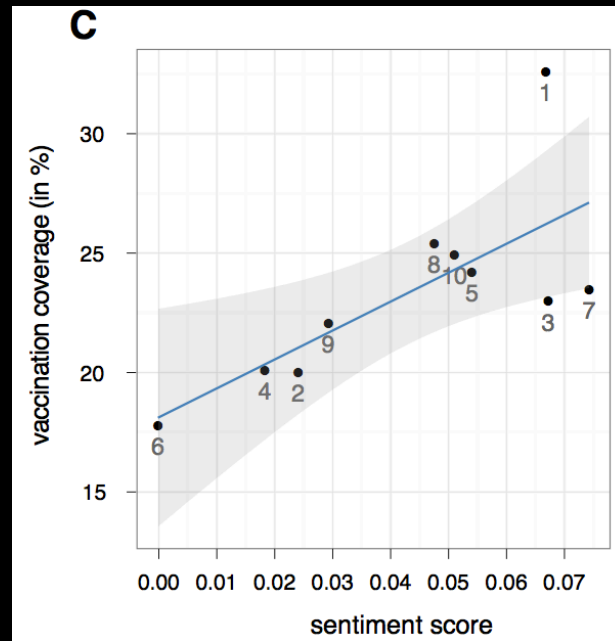
The student_ratings directory contains Python/Django code to let students rate tweets (as described in our paper). The Django code assumes a MySQL backend.

The code as written originally was designed to work with the original tweet ids, tweeter ids etc from Twitter. A lot of the scripts assume certain files exist (some of those have not been checked into this public repository). In those cases some changes will be necessary. If you need help navigating the code, let us know.

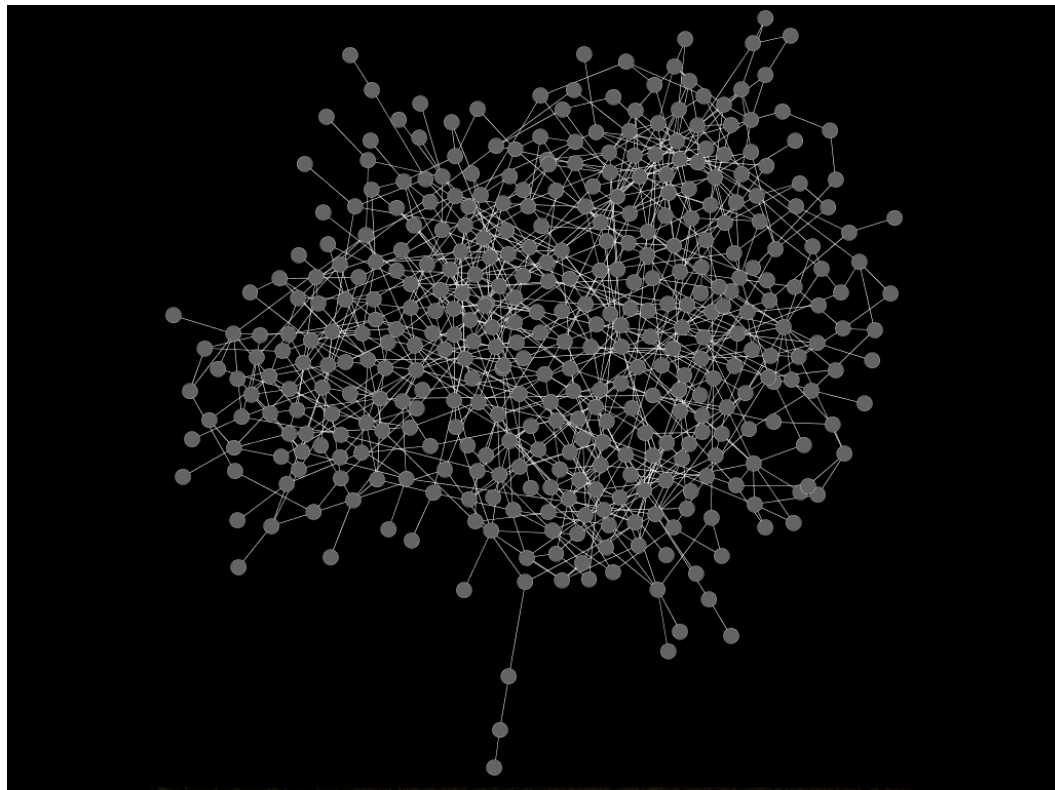
khandelwal@psu.edu, salathe@psu.edu

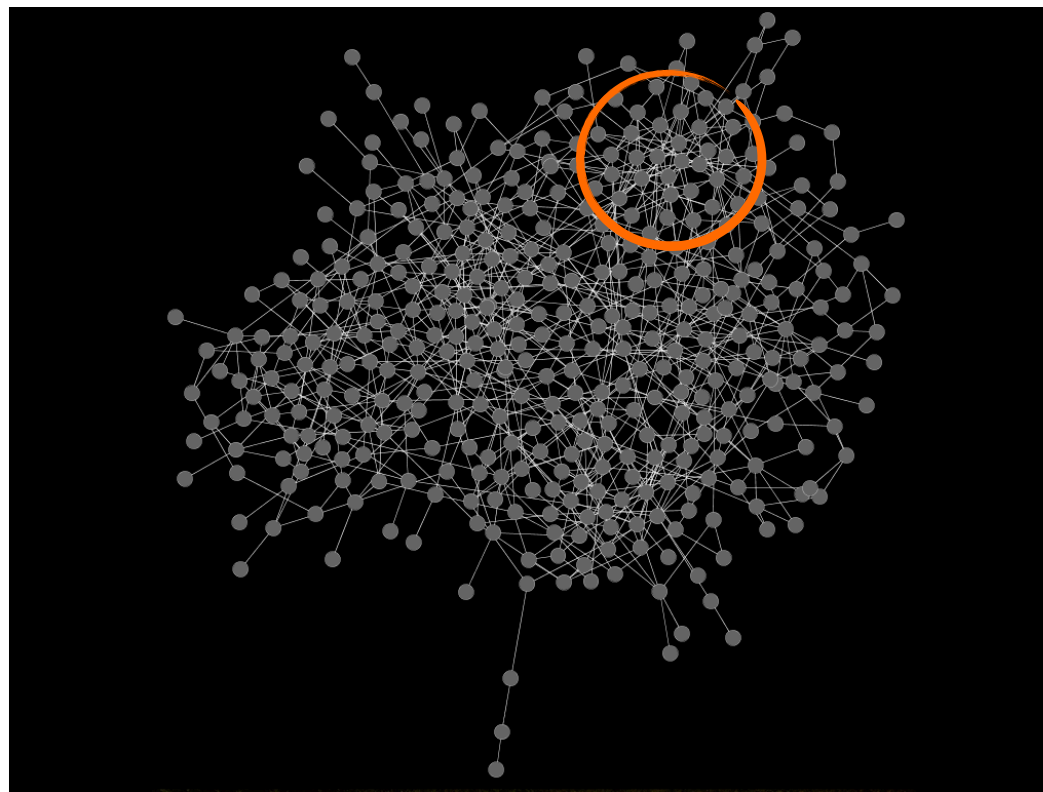


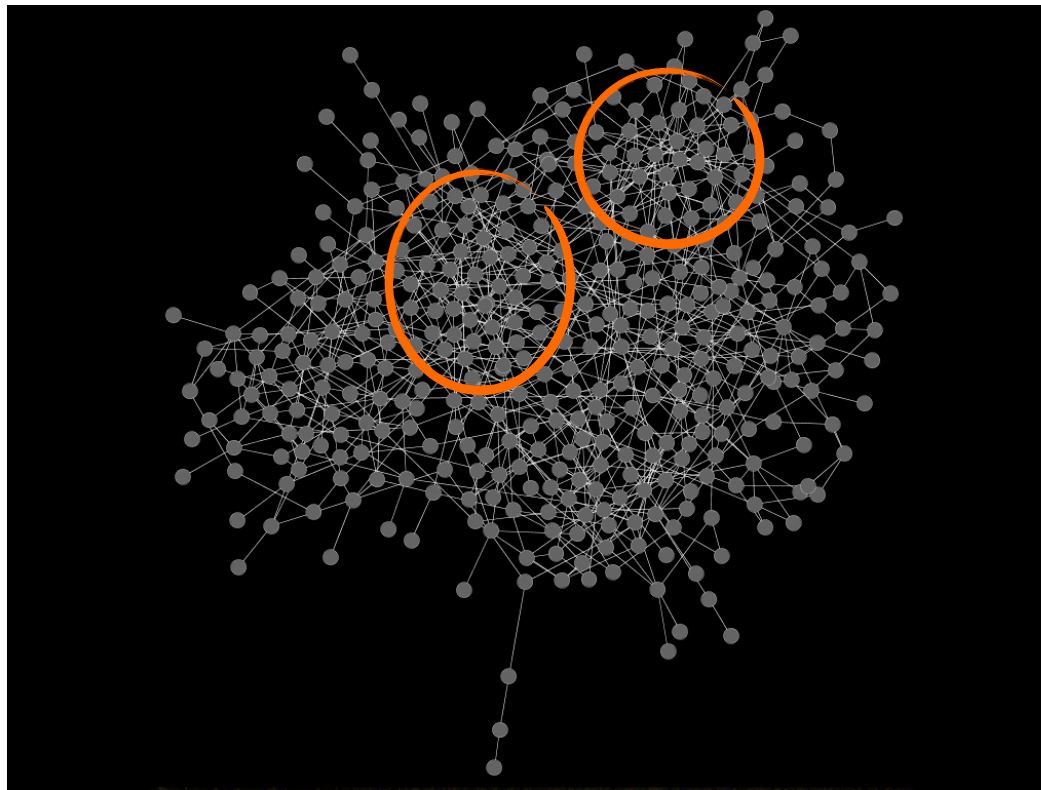
Salathé & Khandelwal, PLoS Computational Biology 2011

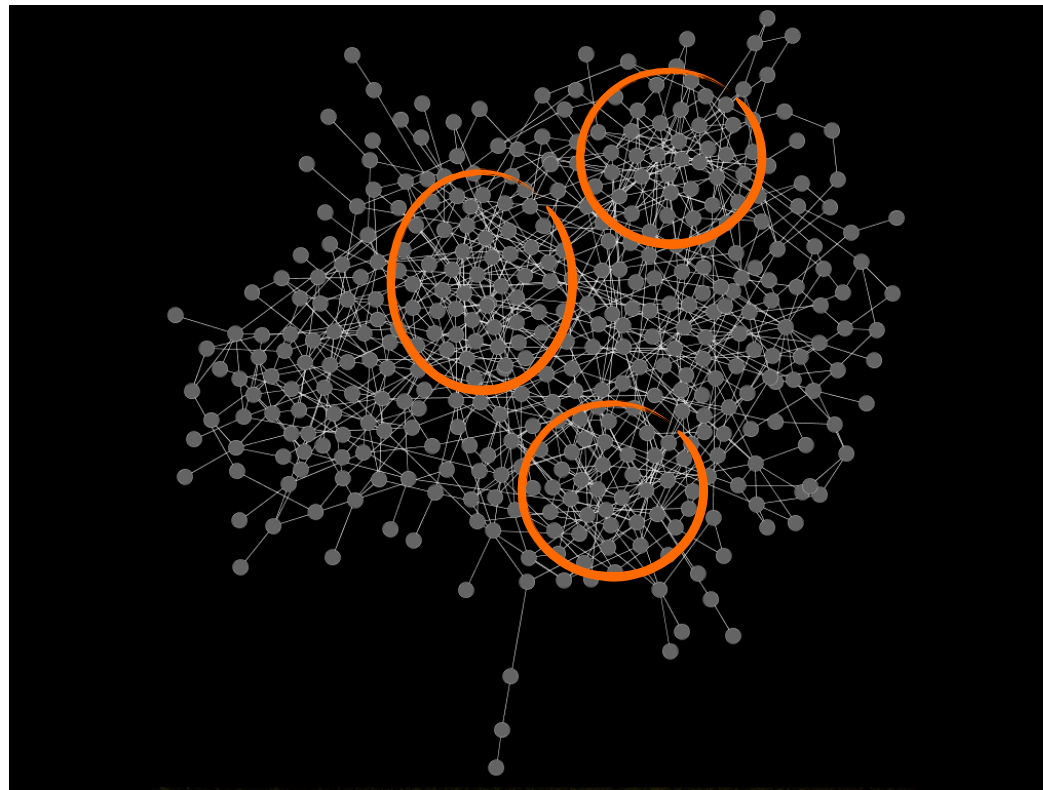


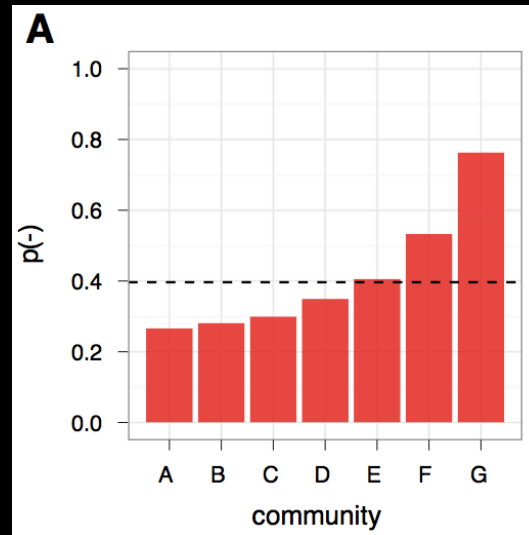
Salathé & Khandelwal, PLoS Computational Biology 2011



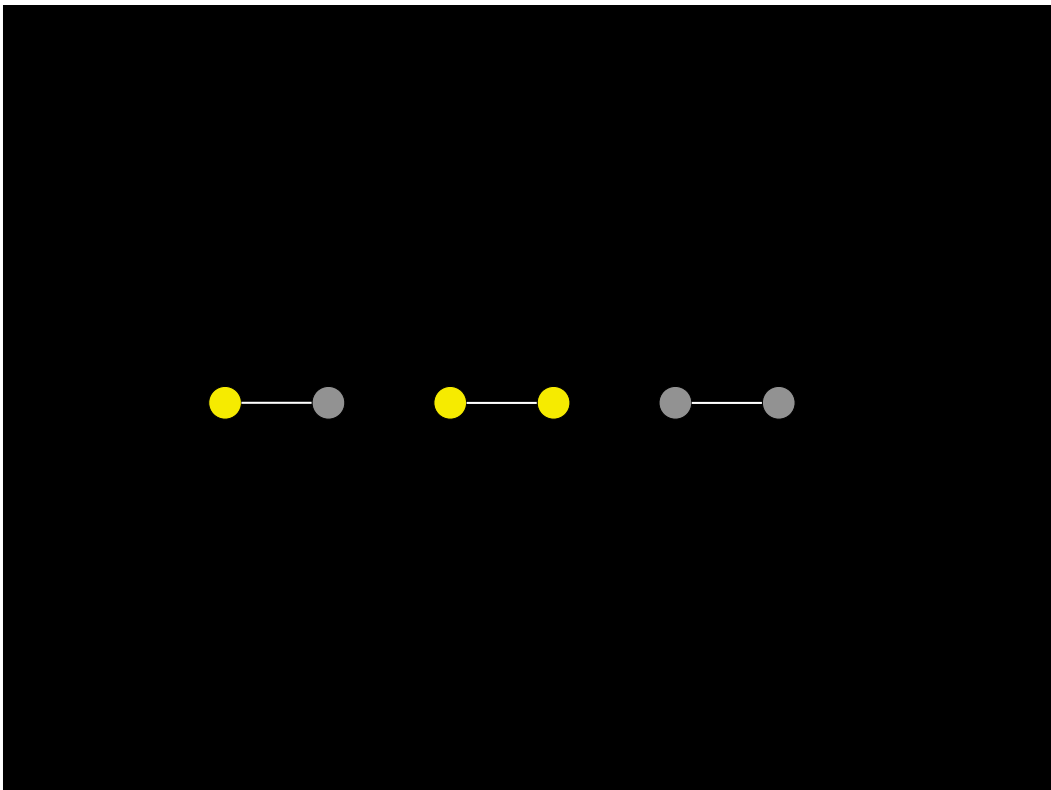


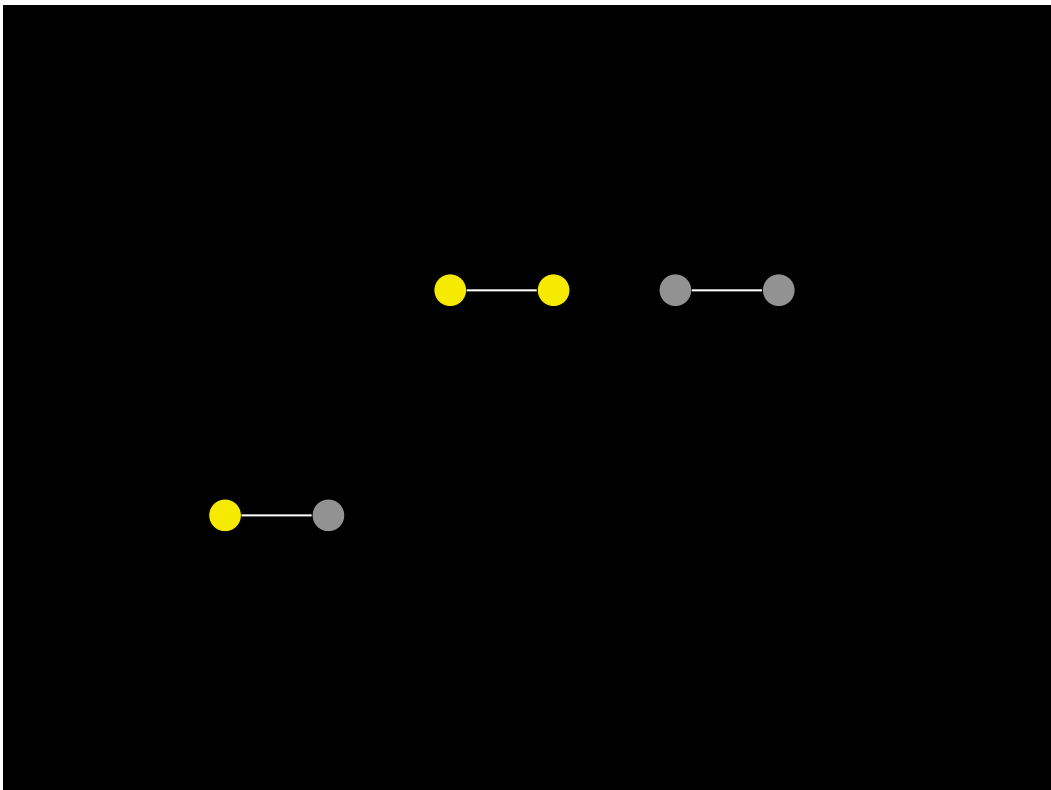


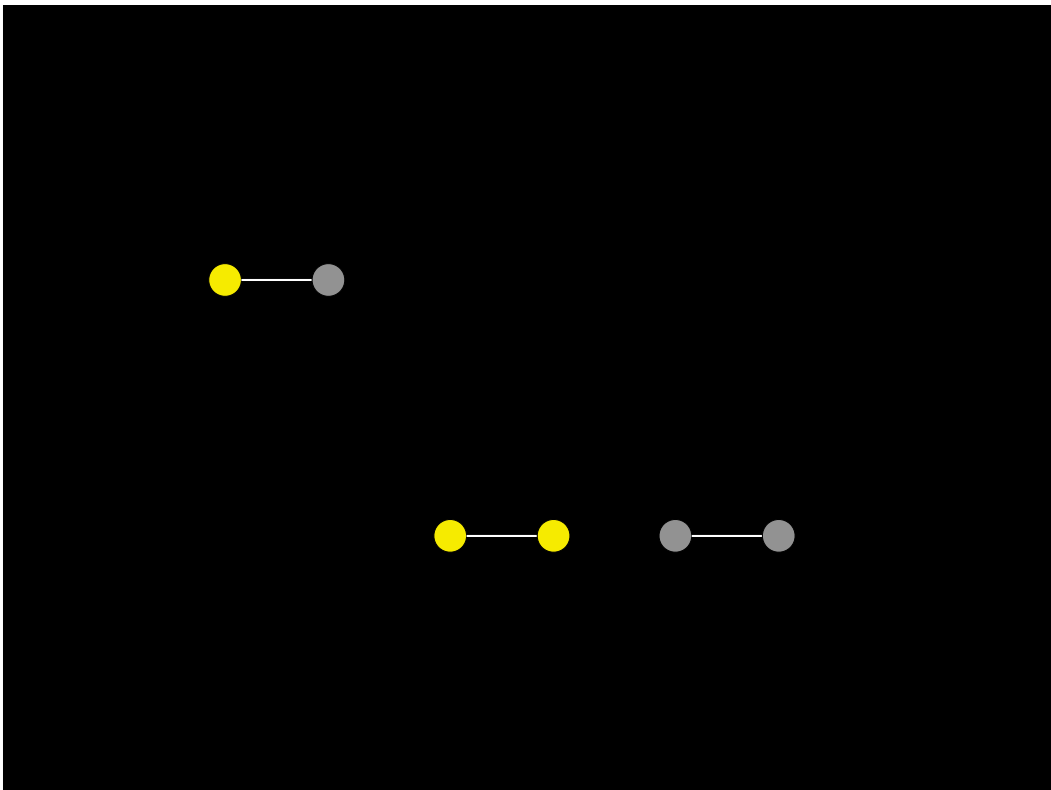


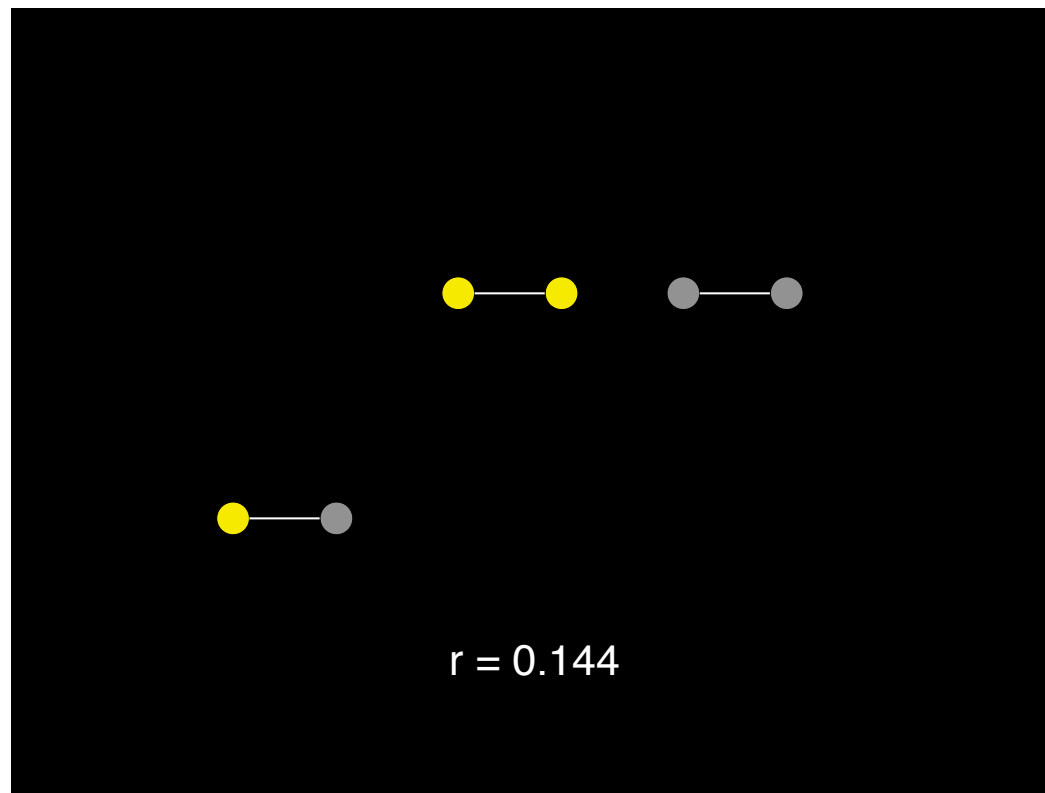


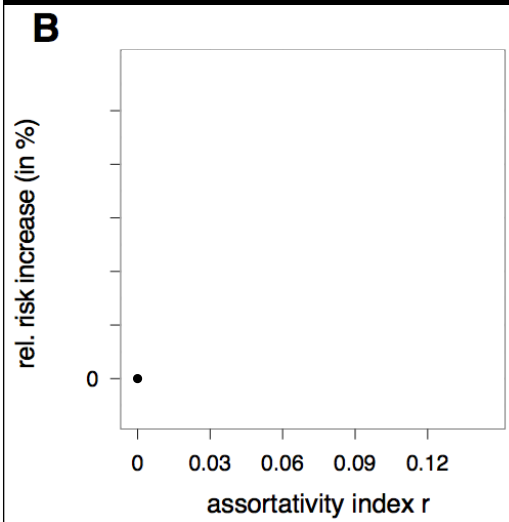
Salathé & Khandelwal, PLoS Computational Biology 2011



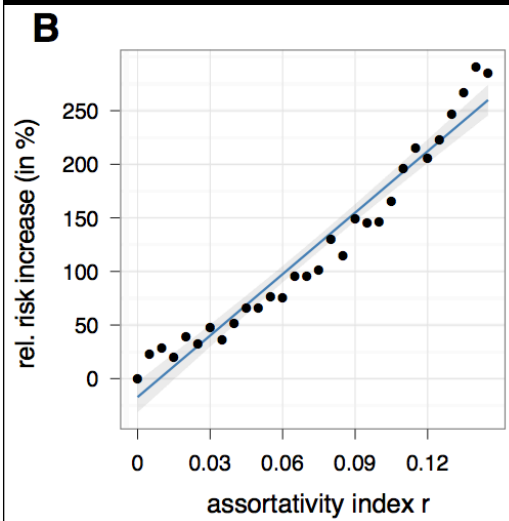




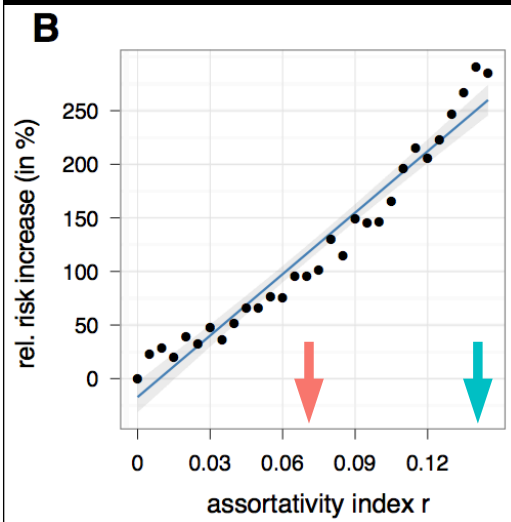




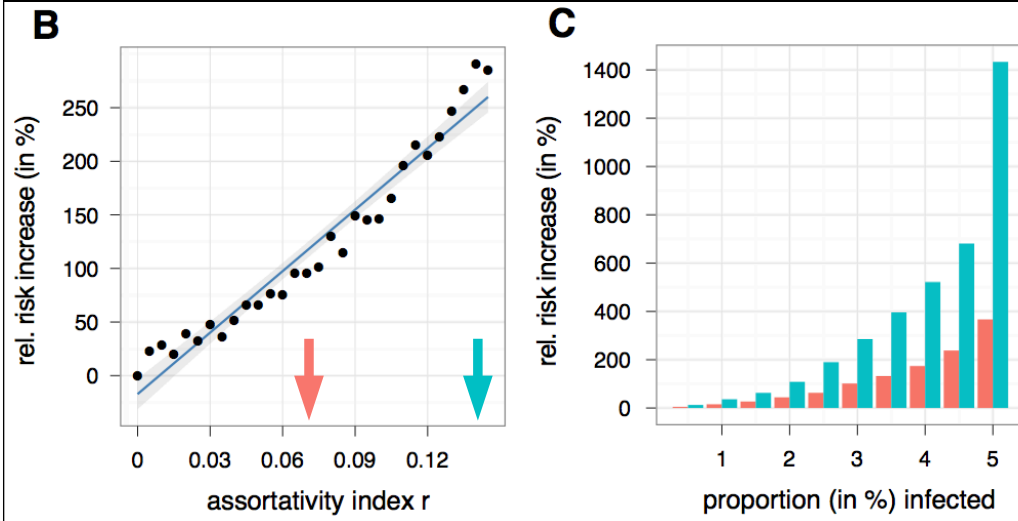
risk of outbreak size 3%

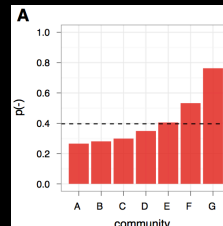
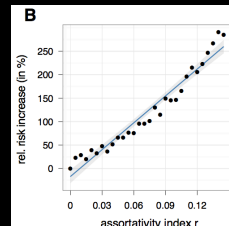
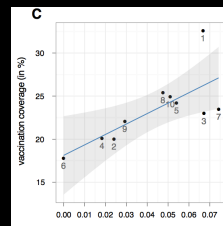
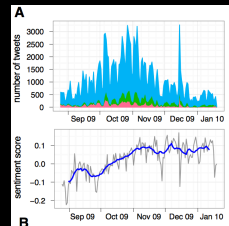


risk of outbreak size 3%



Salathé & Khandelwal, PLoS Computational Biology 2011



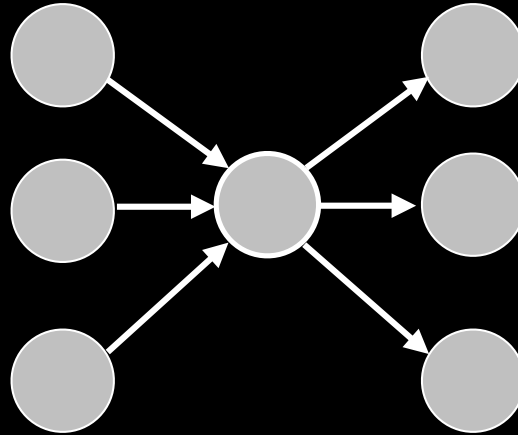


Using web data to measure health related behaviors (cheap, fast, real time)

Measure social network / contagion effects, potential for intervention

FRIENDS

FOLLOWERS

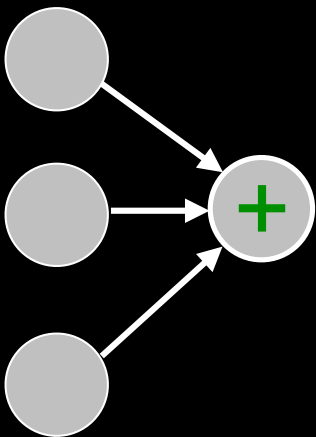


Network statistics:

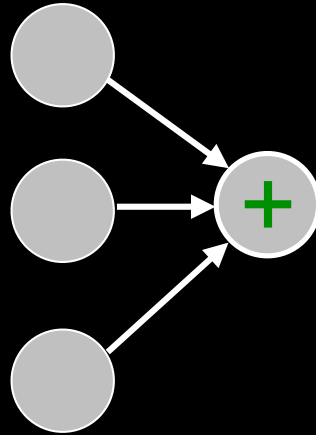
- > 98,235 nodes (max. 101,853)
- > 4,209,361 edges (max. 4,858,985)

we're looking at the final 45 days (Dec 2009, half of January 2010)

FRIENDS



FRIENDS



How many positive /
negative friends?

How many positive /
negative tweets seen?

How many positive /
negative tweets tweeted?

etc.

Predictor	Outcome: neg.	Outcome: pos.
# followers		
# friends		
# neg. friends		
# pos. friends		
# neg. tweets		
# pos. tweets		

Predictor	Outcome: neg.	Outcome: pos.
# followers	↘	-
# friends		
# neg. friends		
# pos. friends		
# neg. tweets		
# pos. tweets		

Predictor	Outcome: neg.	Outcome: pos.
# followers	↘	-
# friends	↗	↗
# neg. friends		
# pos. friends		
# neg. tweets		
# pos. tweets		

Predictor	Outcome: neg.	Outcome: pos.
# followers	↘	-
# friends	↗	↗
# neg. friends	↘	↘
# pos. friends		
# neg. tweets		
# pos. tweets		

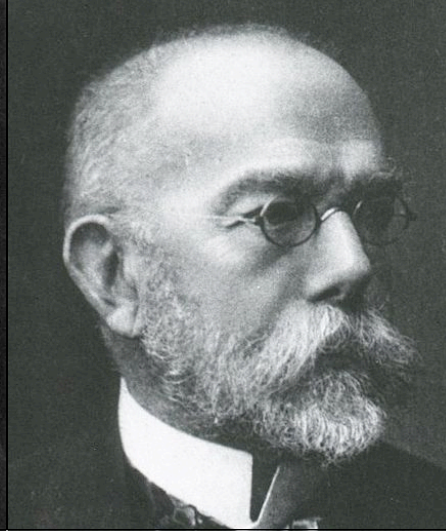
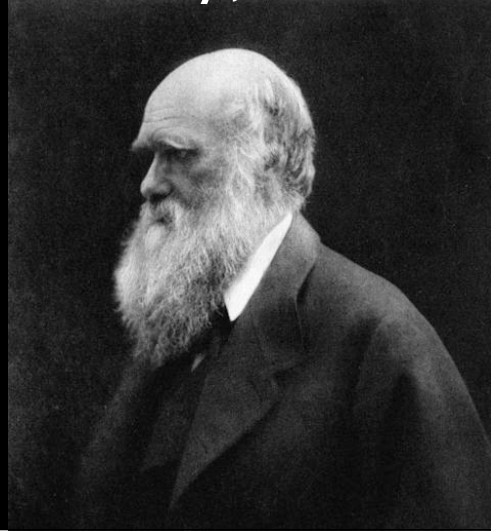
Predictor	Outcome: neg.	Outcome: pos.
# followers	↘	-
# friends	↗	↗
# neg. friends	↘	↘
# pos. friends	↘	-
# neg. tweets		
# pos. tweets		

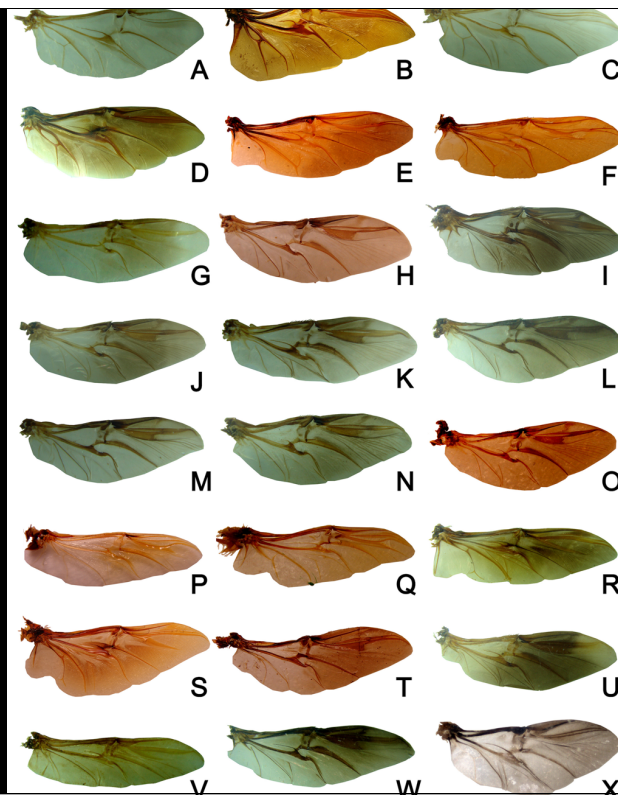
Predictor	Outcome: neg.	Outcome: pos.
# followers	↘	-
# friends	↗	↗
# neg. friends	↘	↘
# pos. friends	↘	-
# neg. tweets	↗	↘
# pos. tweets		

Predictor	Outcome: neg.	Outcome: pos.
# followers	↘	-
# friends	↗	↗
# neg. friends	↘	↘
# pos. friends	↘	-
# neg. tweets	↗	↘
# pos. tweets	↗	-

Predictor	Outcome: neg.	Outcome: pos.
# followers	↘	-
# friends	↗	↗
# neg. friends	↘	↘
# pos. friends	↘	-
# neg. tweets	↗	↘
# pos. tweets	↗	-

Someone needs to make
sense of all those data ->
theory, models





Bai M, McCullough E,
Song K-Q, Liu W-G,
Yang X-K (2011)
Evolutionary
Constraints in Hind
Wing Shape in Chinese
Dung Beetles
(Coleoptera:
Scarabaeinae). PLoS
ONE 6(6): e21600.

linkage disequilibrium. We focus on the contribution of additive genetic variance (V_A) to genotypic variance (V_G).

Genotypic Values

Single Locus with Arbitrary Dominance. Consider a single biallelic locus with genotypic values for CC, Cc and cc of + a , d and $-a$, respectively (notation of [3]). Then, from [3]

$$V_A = 2p(1-p)[a + d(1-2p)]^2, \quad V_D = 4p^2(1-p)^2d^2 \text{ and } V_G = V_A + V_D.$$

For the *uniform* distribution of p

$$E(V_A) = \int_0^1 2p(1-p)[a + d(1-2p)]^2 dp$$

Hence $E(V_A) = a^2/3 + d^2/15$ and $E(V_D) = 2d^2/15$, giving $E(V_A)/E(V_G) = 1 - 2d^2/(5a^2 + 3d^2)$.

For the ‘ U ’-distribution, assuming N is large, and ignoring terms of $O(1/N)$, the integrals simplify to $E(V_A) \sim (a^2 + d^2/3)/K$, $E(V_D) \sim d^2/(3K)$ and $E(V_A)/E(V_G) = 1 - d^2/(3a^2 + 2d^2)$.

Additive \times Additive Model without Dominance or Interactions Including Dominance. A simple additive \times additive epistatic model has these genotypic values:

	CC	Cc	cc
BB	2a	a	0

is complete dominance.

$$M = a[1 - \Pi_i(1 - p_i)^2] \quad \text{and} \quad V_G = M(a - M)$$

$$\alpha_i = a(1 - p_i)\Pi_{j \neq i}(1 - p_j)^2 \quad \text{and} \quad V_A = 2a^2 \sum_i [p_i(1 - p_i)^3$$

For $p_i = 0.5$ at all loci: $V_G = a^2[(\frac{1}{2})^{2L} - (\frac{1}{4})^{2L}]$, $V_A = a^2$ and $V_A/V_G = 2L/(2^{2L} - 1)$. For two loci, $V_A/V_G = 4/15$

Uniform: $E(V_G) = a^2 \left[\left(\frac{1}{3}\right)^L - \left(\frac{1}{3}\right)^L \right]$ and $E(V_A) = \frac{a^2}{3}$

For two loci, $E(V_A)/E(V_G) = 9/16$ and declines as L increases.

‘ U ’: $E(V_G) = \frac{a^2}{2^L} \left[\left(1 - \frac{1}{K}\right)^L - \left(1 - \frac{11}{6K}\right)^L \right]$ and $E(V_A) = \frac{a^2}{2^L} \left(1 - \frac{11}{6K}\right)^{L-1}$

For two loci $E(V_G) = \frac{5a^2}{12K} \left(1 - \frac{17}{12K}\right)$, $E(V_A) = \frac{a^2}{6K}$ and $\frac{E(V_A)}{E(V_G)} = \frac{4(1 - \frac{11}{6K})}{5(1 - \frac{17}{12K})}$

For large N , with two loci $E(V_A)/E(V_G) \rightarrow 4/5$ and for many loci $E(V_A)/E(V_G) \rightarrow 0$

Complementary Model. Another simple epistatic model involving all components is the following:

	CC	Cc	cc
BB	a	a	0
Bb	a	a	0
bb	0	0	0

which can also be defined for multiple loci. For two

Thank you.

salathegroup.com