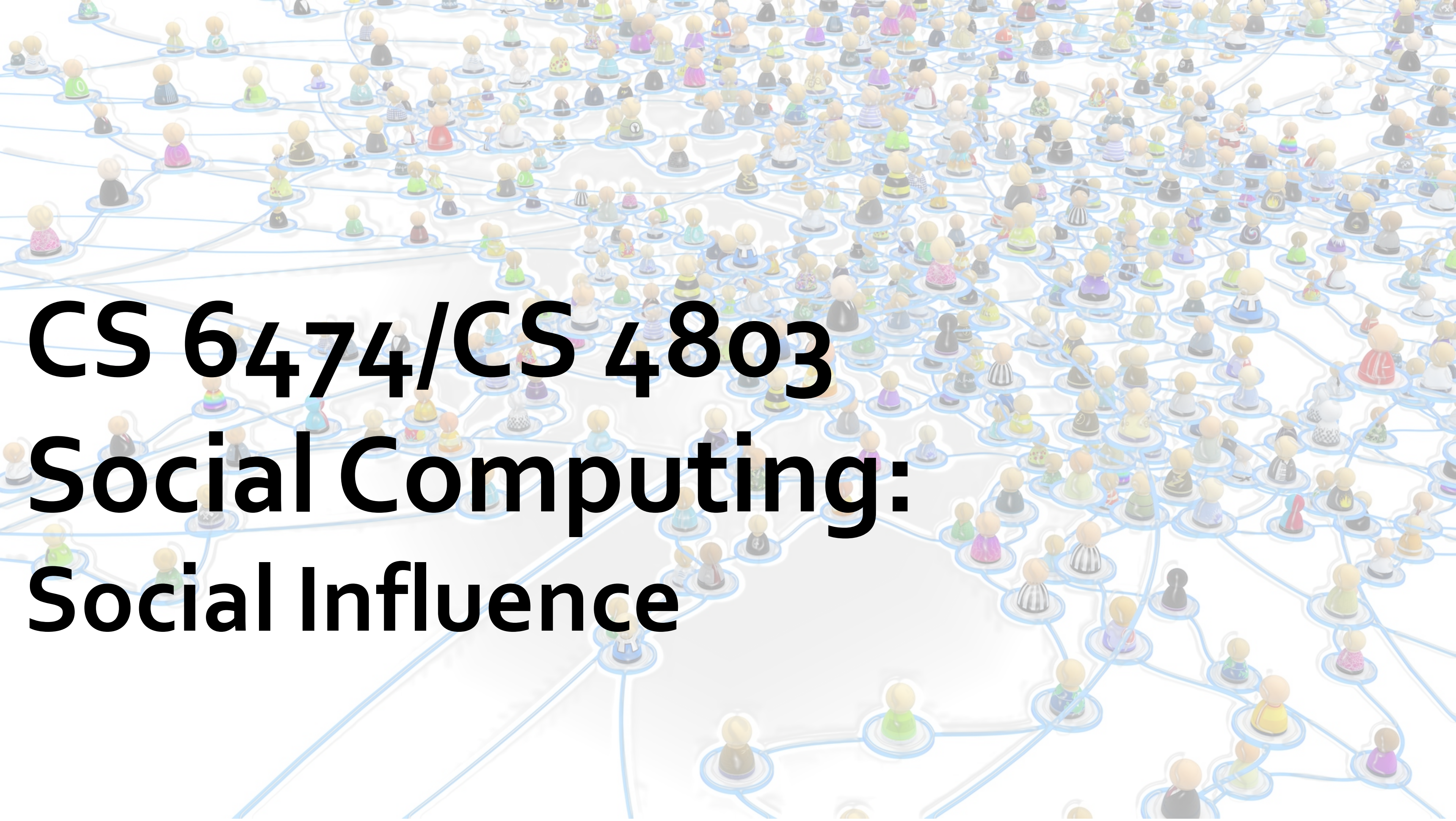


CS 6474/CS 4803 Social Computing:

Social Influence

Munmun De Choudhury

munmund@gatech.edu

Week 6 | September 27, 2017

Obesity contagion

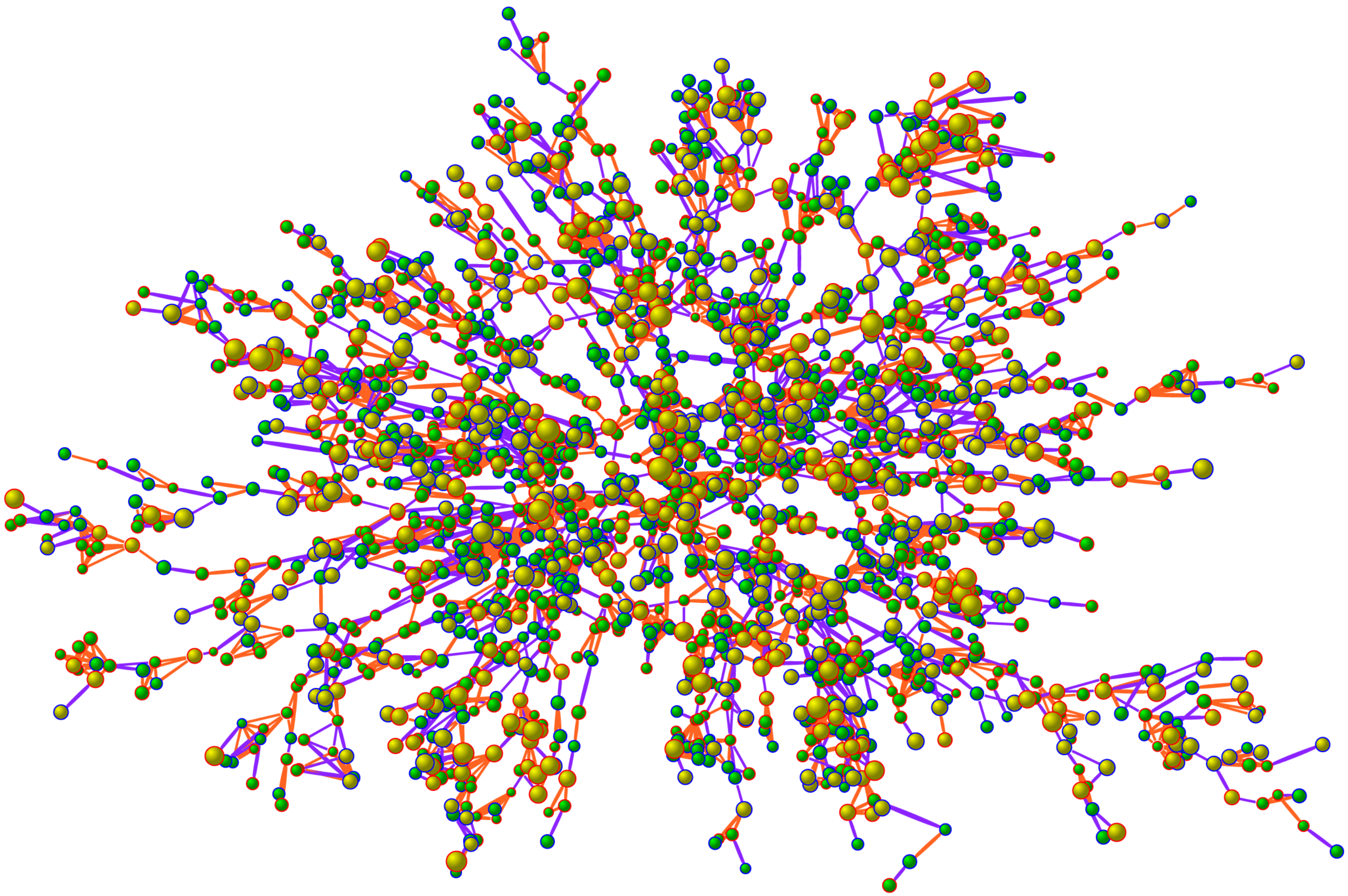


Figure: Network of 2,200 individuals from the Framingham Heart and Health Study. *Source*: Christakis and Fowler 2007.

A formal definition of social influence

Social influence occurs when one's emotions, opinions, or behaviors are affected by others.[1]

[1] http://qualities-of-a-leader.com/personal-mbti-type-analysis/ by way of Wikipedia.

Social influence – a critical construct of social networks

Why are certain things more contagious than others?
Why are certain things more popular or catchy?
How does the social network facilitate such contagion?

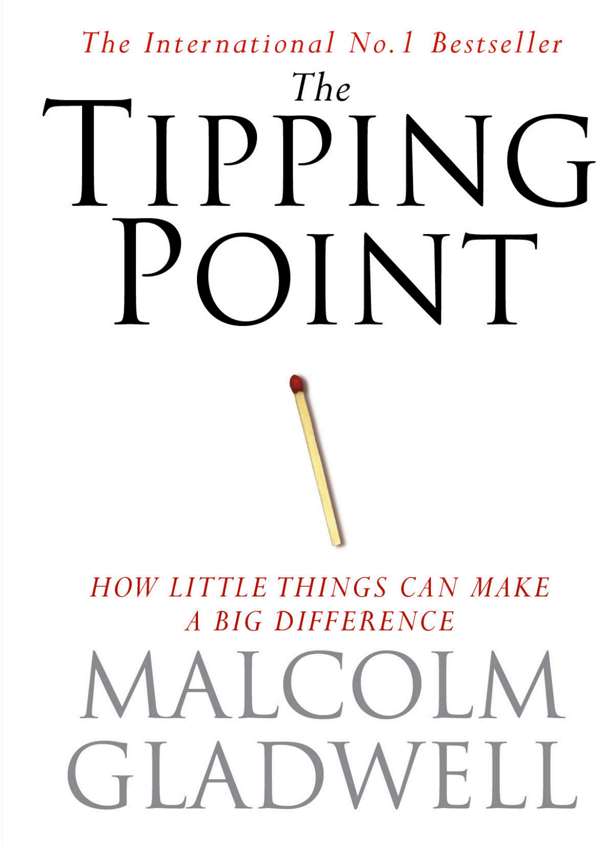
The International No.1 Bestseller

The Tipping Point

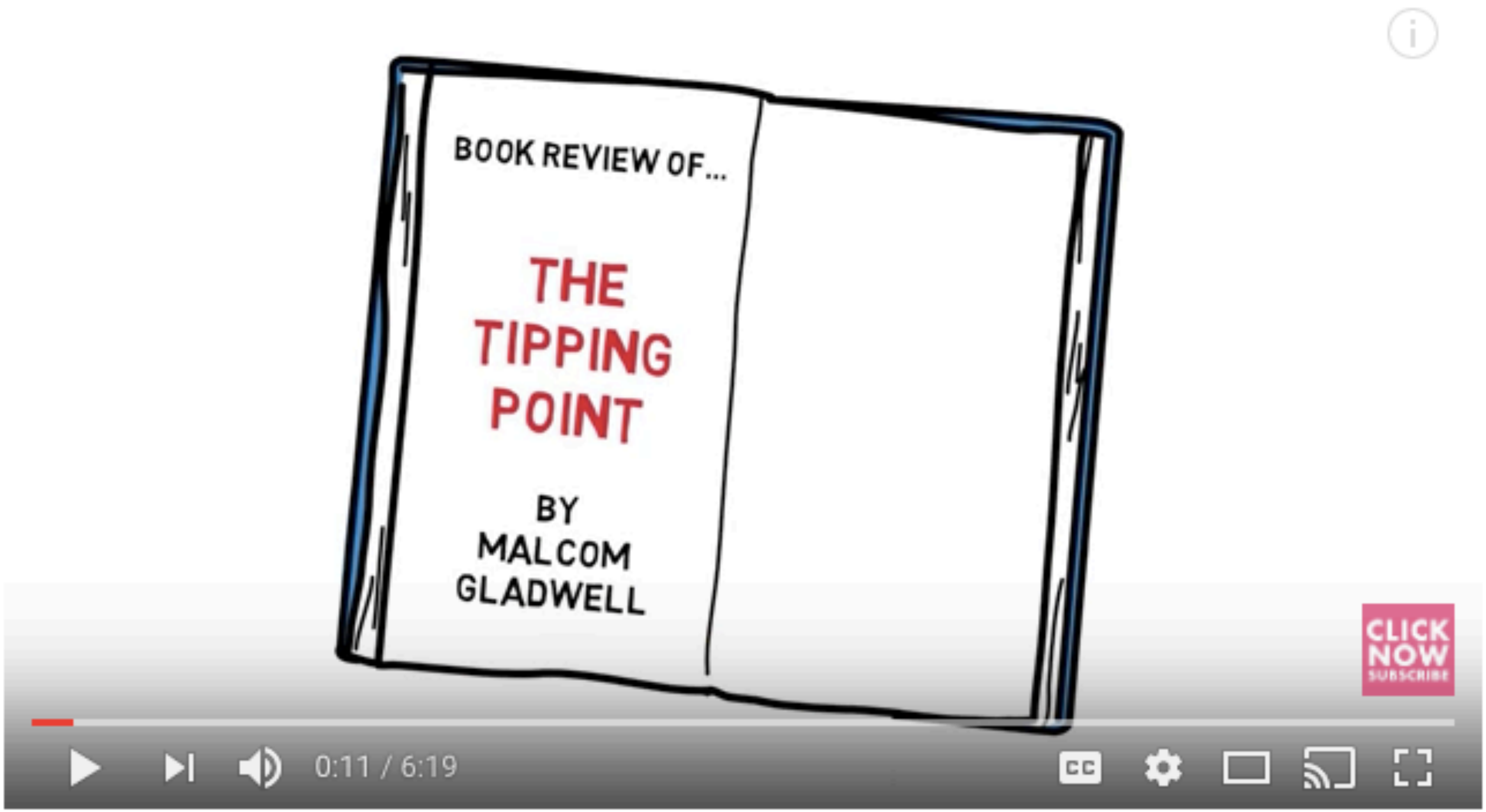
HOW LITTLE THINGS CAN MAKE A BIG DIFFERENCE

MALCOLM GLADWELL

What is the Tipping Point?



That magic moment when an idea, trend or social behavior crosses, tips and spreads like wildfire.



https://www.youtube.com/watch?v=FrpdxTGsjbE

Three Agents of Change

1. The Law of the Few
2. The Stickiness Factor
3. The Power of Context

These provide a direction for how to go about reaching a tipping point.

How Do Know if You're Going to Go Viral?

https://www.youtube.com/watch?v=wcpvKtmAKlo

Models of social influence

1. "*Influentials*"
   coterie of highly visible individuals or media sources (Rogers 1962).
   Ex: Oprah Winfrey, Natalie Portman, George Clooney
2. "*Accidental influentials*"
   social influence more dependent on the receptibility of public than the innate magnetism of an influential (Watts and Dodds 2007).
   Analogy: a forest fire
3. "*Social Networks*"
   close relationships determine the degree of influence

Structure and agency together matter…

Is she worth $10,000 per tweet? Kim Kardashian earns big money using her Twitter account to advertise to her 2million fans

By DAILY MAIL REPORTER
UPDATED: 11:16 EDT, 24 December 2009

Some might question her apparent celebrity status, however Kim Kardashian certainly seems to have acquired some pulling power.

According to a report out this week, the U.S. socialite allegedly commands up to $10,000 (£6,300) for every tweet she posts on her Twitter account as part of her contract with in-stream advertising company Ad.ly.

Kim, 29, is the highest earner on the company's books and the most popular on their roster of celebrity tweeters.

Measuring User Influence in Twitter: The Million Follower Fallacy

Summary

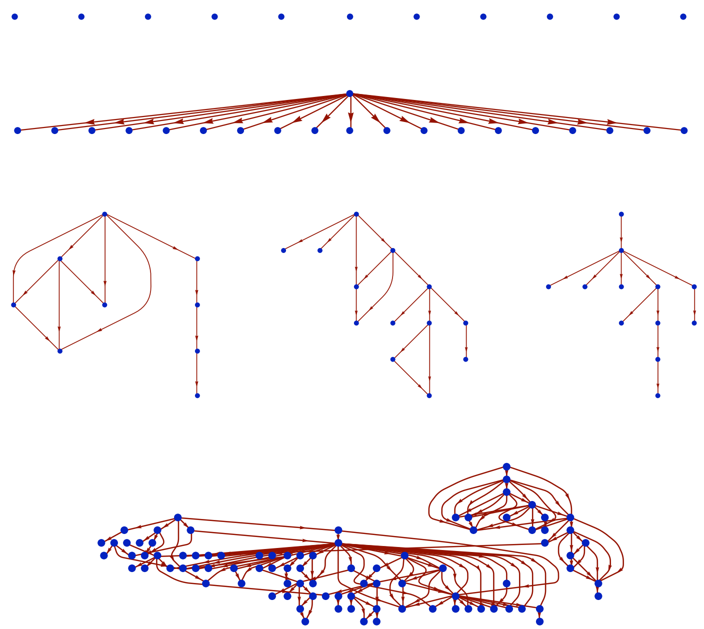
- An early paper examining spread of influence on Twitter and people's social network and interaction characteristics.
- Early uses of a very huge Twitter dataset
- Influence measured through three attributes: in-degree (number of followers), retweets, mentions
- Findings:
  - More in-degree doesn't imply more influence in terms of RTs or mentions
  - Influence users tend to share content on a number of topics
  - Influence generation is gradual, and happens through focus on specific topics

What intuition may explain why large number of followers does not necessarily imply greater influence?

Everyone’s an Influencer: Quantifying Influence on Twitter

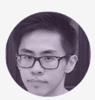
Summary

- Diffusion of URLs on Twitter
  - Easy to be traced back to the originating user through the follower graph
- "Influencers are identified only in retrospect, usually in the aftermath of some outcome of interest, such as the unexpected success of a previously unknown author or the sudden revival of a languishing brand"



Your Facebook Posts Will Probably Go Viral if You Follow These 5 Steps

FACEBOOK



ABOUT THE AUTHOR

Aaron Lee

Aaron Lee is the Founder of AskAaronLee.com & ShortofHeight.com, a fashion blog that shares style for short men. Follow him on Twitter at @askaaronlee

Do you want your Facebook posts to go viral?

Of course you do.

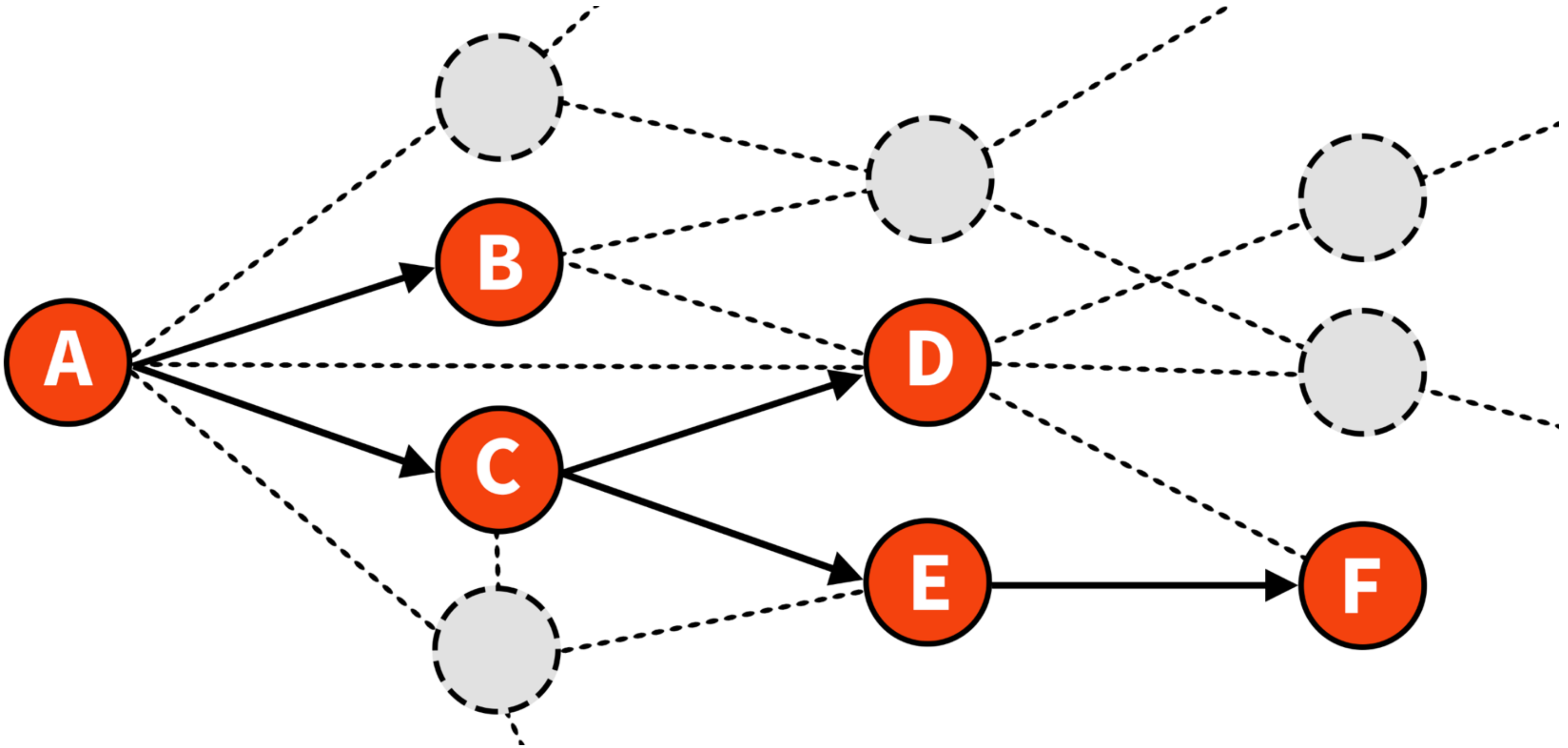
Who wouldn't want to dramatically increase the reach of their posts?!

Why do we care about social influence on social computing systems?

Why do we care about social influence on social computing systems?

Voting/elections, marketing, music/food/book recommendations, public health campaigns, shared interest groups, social movements

An information cascade



How have cascades been studied?

- Will information ever get shared?

  Petrovic, S., Osborne, M., & Lavrenko, V. (2011). RT to Win! Predicting Message Propagation in Twitter. ICWSM 2011.

- Will popular content remain popular?

  Ma, Z., Sun, A., & Cong, G. (2013). On predicting the popularity of newly emerging hashtags in Twitter. JASIST 2013.

- What do large cascades look like?

  Dow, P. A., Adamic, L. A., & Friggeri, A. (2013). The Anatomy of Large Facebook Cascades. ICWSM 2013.

Can cascades be predicted?

Justin Cheng
Stanford University
jcccf@cs.stanford.edu

Lada A. Adamic
Facebook
ladamic@fb.com

P. Alex Dow
Facebook
adow@fb.com

Jon Kleinberg
Cornell University
kleinber@cs.cornell.edu

Jure Leskovec
Stanford University
jure@cs.stanford.edu

ABSTRACT

On many social networking web sites such as Facebook and Twitter, resharing or reposting functionality allows users to share others' content with their own friends or followers. As content is reshared from user to user, large cascades of reshares can form. While a growing body of research has focused on analyzing and characterizing such cascades, a recent, parallel line of work has argued that the future trajectory of a cascade may be inherently unpredictable. In this work, we develop a framework for addressing cascade prediction problems. On a large sample of photo reshare cascades on Facebook, we find strong performance in predicting whether a cascade will continue to grow in the future. We find that the relative growth of a cascade becomes more predictable as we observe more of its reshares, that temporal and structural features are key predictors of cascade size, and that initially, breadth, rather than depth in a cascade is a better indicator of larger cascades. This prediction performance is robust in the sense that multiple distinct classes of features all achieve similar performance. We also discover that temporal features are predictive of a cascade's eventual shape. Observing independent cascades of the same content, we find that while these cascades differ greatly in size, we are still able to predict which ends up the largest.

Categories and Subject Descriptors: H.2.8 **[Database Management]**: Database applications—*Data mining*
General Terms: Experimentation, Measurement.
Keywords: Information diffusion, cascade prediction, contagion.

has focused on characterizing cascades in these domains, including their structural properties and their content.

In parallel to these investigations, there has been a recent line of work adding notes of caution to the study of cascades. These cautionary notes fall into two main genres: first, that large cascades are rare [11]; and second, that the eventual scope of a cascade may be an inherently unpredictable property [28, 31]. The first concern — that large cascades are rare — is a widespread property that has been observed quantitatively in many systems where information is shared. The second concern is arguably more striking, but also much harder to verify quantitatively: to what extent is the future trajectory of a cascade predictable; and which features, if any, are most useful for this prediction task?

Part of the challenge in approaching this prediction question is that the most direct ways of formulating it do not fully address the two concerns above. Specifically, if we are presented with a short initial portion of a cascade and asked to estimate its final size, then we are faced with a pathological prediction task, since almost all cascades are small. Alternately, if we radically overrepresent large cascades in our sample, we end up studying an artificial setting that does not resemble how cascades are encountered in practice. A set of recent initial studies have undertaken versions of cascade prediction despite these difficulties [19, 23, 26, 29], but to some extent they are inherent in these problem formulations.

These challenges reinforce the fact that finding a robust way to formulate the problem of cascade prediction remains an open problem. And because it is open, we are missing a way to obtain a

Example Reshares on Facebook

Lada Adamic shared a link via Erik Johnston.
January 16, 2013

When life gives you an almost empty jar of nutella, add some ice cream...
(and other useful tips)

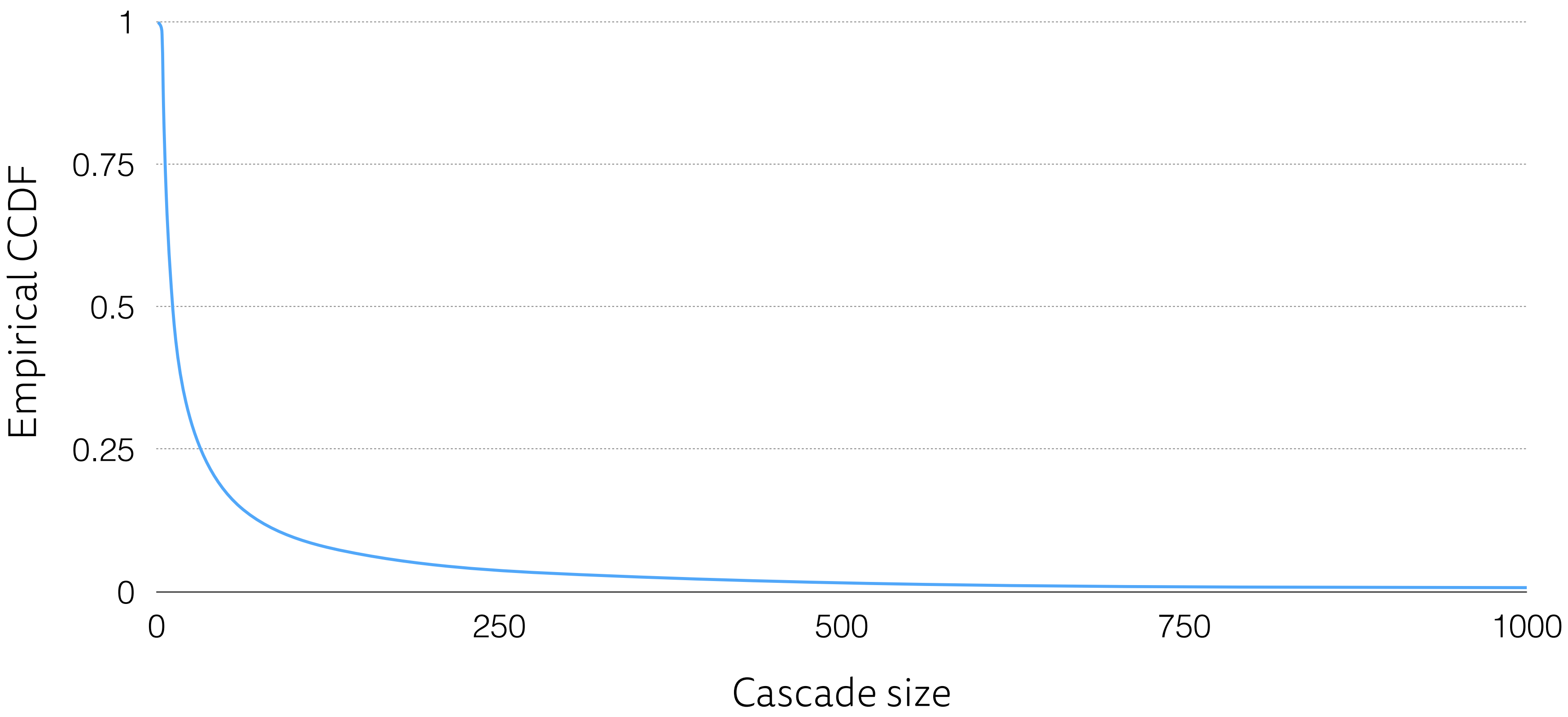
50 Life Hacks to Simplify your World
twistedsifter.com

Life hacks are little ways to make our lives easier. These low-budget tips and trick can help you organize and de-clutter space; prolong and preserve your products; or teach you...

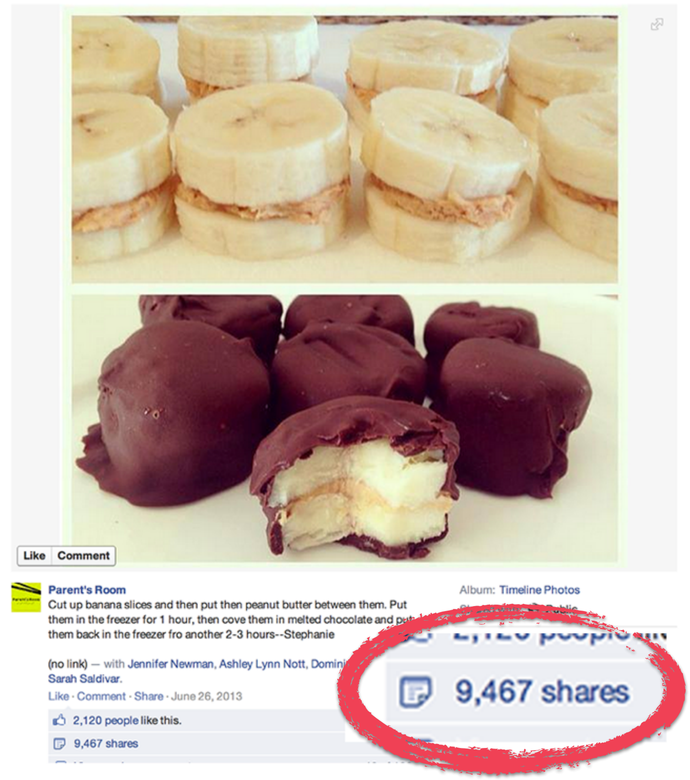
Like · Comment · Share 40 3 25

Which of these Facebook photos went viral?

Difficulty #1 Large cascades are rare



Difficulty #2 Same content, different popularity



Class Activity 1

What factors affect predictability?

Content
(e.g. has overlaid text)

User
(e.g. friend count)

Structural
(e.g. proximity to root in G)

Temporal
(e.g. time between reshares)

Content Features	
$score_{food/nature/...}$	The probability of the photo having a specific feature (food, overlaid text, landmark, nature, etc.)
is_en	Whether the photo was posted by an English-speaking user or page
$has\_caption$	Whether the photo was posted with a caption
$liwc_{pos/neg/soc}$	Proportion of words in the caption that expressed positive or negative emotion, or sociality, if English
Root (Original Poster) Features	
$views_{0,\ k}$	Number of users who saw the original photo until the kth reshare was posted
$orig\_is\_page$	Whether the original poster is a page
$outdeg(v_0)$	Friend, subscriber or fan count of the original poster
age_0	Age of the original poster, if a user
$gender_0$	Gender of the original poster, if a user
fb_age_0	Time since the original poster registered on Facebook, if a user
$activity_0$	Average number of days the original poster was active in the past month, if a user
Resharer Features	
$views_{1..k-1,\ k}$	Number of users who saw the first $k-1$ reshares until the kth reshare was posted
$pages_k$	Number of pages responsible for the first k reshares, including the root, or $\sum_{i=0}^{k} \mathbb{1}\{v_i \text{ is a page}\}$
$friends_k^{avg/90p}$	Average or 90th percentile friend count of the first k resharers, or $\frac{1}{k}\sum_{i=1}^{k} outdeg_{friends}(v_i)\mathbb{1}\{v_i \text{ is a user}\}$
$fans_k^{avg/90p}$	Average or 90th percentile fan count of the first k resharers, or $\frac{1}{k}\sum_{i=1}^{k} outdeg(v_i)\mathbb{1}\{v_i \text{ is a page}\}$
$subscribers_k^{avg/90p}$	Average or 90th percentile subscriber count of the first k resharers, or $\frac{1}{k}\sum_{i=1}^{k} outdeg_{subscriber}(v_i)\mathbb{1}\{v_i \text{ is a user}\}$
$fb\_ages_k^{avg/90p}$	Average or 90th percentile time since the first k resharers registered on Facebook, or $\frac{1}{k}\sum_{i=1}^{k} fb\_age_i$
$activities_k^{avg/90p}$	Average number of days the first k resharers were active in July, or $\frac{1}{k}\sum_{i=1}^{k} activity_i$
$ages_k^{avg/90p}$	Average age of the first k resharers, or $\frac{1}{k}\sum_{i=1}^{k} age_i$
$female_k$	Number of female users among the first k resharers, or $\sum_{i=1}^{k} \mathbb{1}\{gender_i \text{ is female}\}$
Structural Features	
$outdeg(v_i)$	Connection count (sum of friend, subscriber and fan counts) of the ith resharer (or out-degree of v_i on $G=(V,E)$)
$outdeg(v_i')$	Out-degree of the ith reshare on the induced subgraph $G'=(V',E')$ of the first k resharers and the root
$outdeg(\hat{v}_i)$	Out-degree of the ith reshare on the reshare graph $\hat{G}=(\hat{V},\hat{E})$ of the first k reshares
$orig\_connections_k$	Number of first k resharers who are friends with, or fans of the root, or $\lvert\{v_i \mid (v_0,v_i)\in E,\ 1\le i\le k\}\rvert$
$border\_nodes_k$	Total number of users or pages reachable from the first k resharers and the root, or $\lvert\{v_i \mid (v_i,v_j)\in E,\ 0\le i,j\le k\}\rvert$
$border\_edges_k$	Total number of first-degree connections of the first k resharers and the root, or $\lvert\{(v_i,v_j) \mid (v_i,v_j)\in E,\ 0\le i,j\le k\}\rvert$
$subgraph_k'$	Number of edges on the induced subgraph of the first k resharers and the root, or $\lvert\{(v_i,v_j) \mid (v_i,v_j)\in E',\ 0\le i,j\le k\}\rvert$
$depth_k'$	Change in tree depth of the first k reshares, or $\min_\beta \sum_{i=1}^{k}(depth_i-\beta i)^2$
$depths_k^{avg/90p}$	Average or 90th percentile tree depth of the first k reshares, or $\frac{1}{k}\sum_{i=1}^{k} depth_i$
did_leave	Whether any of the first k reshares are not first-degree connections of the root
Temporal Features	
$time_i$	Time elapsed between the original post and the ith reshare
$time_{1..k/2}'$	Average time between reshares, for the first $k/2$ reshares, or $\frac{1}{k/2-1}\sum_{i=1}^{k/2-1}(time_{i+1}-time_i)$
$time_{k/2..k}'$	Average time between reshares, for the last $k/2$ reshares, or $\frac{1}{k/2-1}\sum_{i=k/2}^{k-1}(time_{i+1}-time_i)$
$time_{1..k}''$	Change in the time between reshares of the first k reshares, or $\min_\beta \sum_{i=1}^{k-1}(time_{i+1}-time_i)-\beta i)^2$
$views_{0,k}'$	Number of users who saw the original photo, until the kth reshare was posted, per unit time, or $\frac{views_{0,\ k}}{time_k}$
$views_{1..k-1,\ k}'$	Number of users who saw the first $k-1$ reshares, until the kth reshare was posted, per unit time, or $\frac{views_{1..k-1,k}}{time_k}$

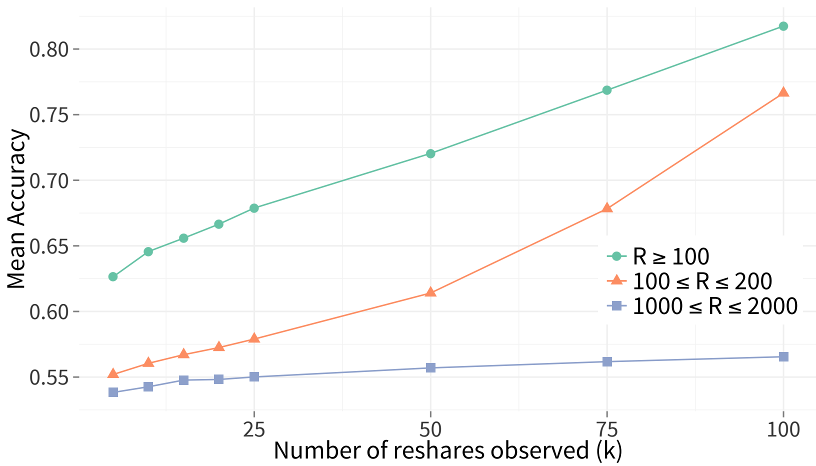


Figure 6: Knowing that a cascade obtains at least R reshares, prediction performance increases linearly with k, $k \leq R$. However, differentiating among cascades with large R also becomes more difficult.

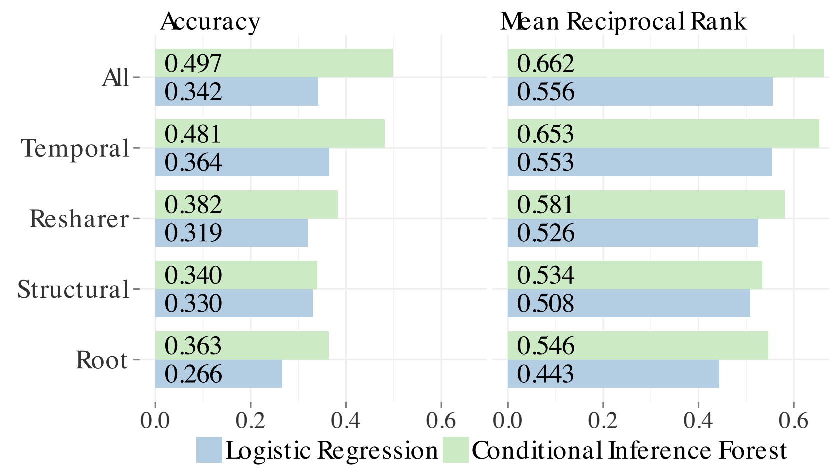


Figure 10: In predicting the largest cascade in clusters of 10 or more cascades of identical photos, we perform significantly above the baseline of 0.1.

Main findings

- The paper examines not only how the predictability of a cascade changes as more and more of the cascade is observed (it improves), but also how predictable large cascades are if we only observe them initially
- Larger cascades are more difficult to predict
- Some features, e.g., the average connection count of the first *k* re-sharers, have increasing predictive ability with increasing *k*
- Some features weaken in importance, e.g., the connectivity of the root node
- The importance of features depends on properties of the original upload as well: the topics present in the caption, the language of the root node, as well as the content of the photo

But the platform matters…

Class Activity 2

https://www.youtube.com/watch?v=AtnR5H6AVVU