Gesundheit!
Modeling Contagion through Facebook News Feed

Eric Sun, Stanford University
Itamar Rosenn, Facebook
Cameron A. Marlow, Facebook
Thomas M. Lento, Facebook

Motivation

- How do ideas diffuse through a large social network?

- Prior models start with an isolated event and explore conditions under which this event triggers a global cascade
  - Little empirical evidence to assess the validity of the models

- We present the first empirical analysis of repeated large-scale diffusion over a global social network.
Theory of the Influentials

- Popular Wisdom: it’s all about the “influentials” (Malcolm Gladwell, etc.)

- Idea: reach a tiny group of influential people, and you’ll reach everyone else through them for free

Accidental Influencers

- Duncan Watts: anyone can be an “influencer”.

- Ideas don’t spread via influentials
  - Ideas spread like viruses: either you’re susceptible, or you’re not

- Success depends not on how persuasive the early adopter(s) are, but whether everyone else is easily persuaded

- Watts simulations: influential nodes are no more likely to trigger cascades than average nodes.
Questions

- In most network models of diffusion, contagion is triggered by a fairly small number of sources: is this a good assumption?

- What are some characteristics of diffusion chains on Facebook?

- Can we use demographic or behavioral characteristics to predict the size of diffusion chains that a particular user will create?

Data and Methodology
Spreading Ideas on Facebook

Empirical Study

- Focus on one type of action for an empirical analysis: Page fanning
Sample News Feed story

Alice fans a Page

Bob sees Alice's action on his News Feed; Bob fans the Page as well

Charlie sees Alice's action on his News Feed; Charlie fans the Page as well

Aside: current model of Page fanning...
Data

- Data include all actor → follower connections for 262,985 Facebook Pages between 2/19/08 and 8/19/08

- Main dataset: Page-level data
- Second dataset: select 10 random, representative Pages and analyze the users that start chains
  - Pages were at least 40 days old and had at least 5,000 fans

Prediction Model

- Response: max_chain_length
- Predictors:
  - gender
  - log age
  - log Facebook age
  - log feed_exposure (# friends who saw News Feed story)
  - log friend_count
  - log activity_count (wall posts + messages sent + photos added)
  - log popularity (controls for News Feed exposure)
- Method: zero-inflated negative binomial regression
Results

Large Connected Trees of Diffusion

Diffusion chain for Stripy, a cartoon popular in Bosnia (blue), Slovenia (yellow), and Croatia (green).

Link drawn if the follower fans a Page within 24 hours of first seeing a News Feed story that a friend has fanned the same Page.
Large Clusters Not Started by “One Guy”

- Roughly 15% of fans in the biggest cluster of each Page are start points
  - The variability in this percentage becomes very small as # fans increases

- Clusters are formed when many short diffusion chains merge

- 86.4% of paths of Page diffusion involve at least four individuals
  - Compare to 38% in real-life study (Brown and Reingen 1987)

Results of Chains Regression

<table>
<thead>
<tr>
<th>Variable</th>
<th>Intuition</th>
<th>Fuddruckers</th>
<th>Cruise</th>
<th>Bolt</th>
<th>Zidane</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Intercept)</td>
<td>N/A</td>
<td>2.574</td>
<td>1.419</td>
<td>0.596</td>
<td>-0.105</td>
</tr>
<tr>
<td>age</td>
<td>?</td>
<td>-0.537</td>
<td>-1.006</td>
<td>-0.386</td>
<td>-0.981</td>
</tr>
<tr>
<td>gender==male</td>
<td>?</td>
<td>-0.011</td>
<td>-0.076</td>
<td>-0.144</td>
<td>0.116</td>
</tr>
<tr>
<td>Facebook_age</td>
<td>+</td>
<td>-0.522</td>
<td>-0.052</td>
<td>-0.415</td>
<td>0.153</td>
</tr>
<tr>
<td>activity_count</td>
<td>+</td>
<td>-0.102</td>
<td>-0.142</td>
<td>-0.064</td>
<td>-0.100</td>
</tr>
<tr>
<td>friend_count</td>
<td>+</td>
<td>-0.220</td>
<td>0.087</td>
<td>-0.023</td>
<td>0.009</td>
</tr>
<tr>
<td>feed_exposure</td>
<td>control</td>
<td>1.279</td>
<td>1.008</td>
<td>0.860</td>
<td>1.053</td>
</tr>
<tr>
<td>popularity</td>
<td>control</td>
<td>-0.014</td>
<td>-0.245</td>
<td>0.021</td>
<td>-0.120</td>
</tr>
</tbody>
</table>

- Recall: max_chain_length is the response variable
- Demographic characteristics not important
- Number of Facebook friends not important
Conclusions

▪ Facebook News Feed enables long-lasting chains of diffusion that may reach many more people than real-life diffusion chains.

▪ The Facebook network is very connected: ideas with good receptiveness will attract wide, long connected clusters.

▪ Long chains are not a function of Facebook age, activity, users’ demographics, or even # of friends: it’s only related to exposure.

Future Work

▪ Evaluate how accurately various theoretical models of diffusion account for the empirical phenomena uncovered

▪ Test experimental contagion events to better understand how different pieces of content and different start conditions determine the eventual structure of a diffusion cascade

▪ Test diffusion of other types of content
Gesundheit!
Modeling Contagion through Facebook News Feed

Eric Sun, Stanford University <esun@cs.stanford.edu>
Itamar Rosenn, Facebook <itamar@facebook.com>
Cameron A. Marlow, Facebook <cameron@facebook.com>
Thomas M. Lento, Facebook <dento@facebook.com>