Modeling and analysis of cascades and contagion

Jon Kleinberg

Cornell University

Including joint work with Lars Backstrom, Larry Blume, Justin Cheng, Cristian Danescu-Niculescu-Mizil, David Easley, Dan Huttenlocher, Bobby Kleinberg, Xiangyang Lan, Lillian Lee, Jure Leskovec, Cameron Marlow, Éva Tardos, and Johan Ugander
Diffusion and Contagion

The spread of information and behaviors from person to person through a network.

- Agricultural, medical innovations [Ryan-Gross 1943, Coleman et al 1966]
- Media influence and two-stage flow [Lazarsfeld et al 1944]
- Collective action, social movements [McAdam 1986, Chwe 1999]
- Viral marketing [Jurvetson 2000, Domingos-Richardson 2001]
- News, rumors, gossip, urban legends, ...
Diffusion and Contagion

The spread of information and behaviors from person to person through a network.

- Agricultural, medical innovations [Ryan-Gross 1943, Coleman et al 1966]
- Media influence and two-stage flow [Lazarsfeld et al 1944]
- Collective action, social movements [McAdam 1986, Chwe 1999]
- Viral marketing [Jurvetson 2000, Domingos-Richardson 2001]
- News, rumors, gossip, urban legends, ...
The spread of information and behaviors from person to person through a network.

- Agricultural, medical innovations [Ryan-Gross 1943, Coleman et al 1966]
- Media influence and two-stage flow [Lazarsfeld et al 1944]
- Collective action, social movements [McAdam 1986, Chwe 1999]
- Viral marketing [Jurvetson 2000, Domingos-Richardson 2001]
- News, rumors, gossip, urban legends, ...
Diffusion and Contagion

The spread of information and behaviors from person to person through a network.

- Models based on positive externalities and individual inference.
- Contagion with individual thresholds.
- Estimating thresholds in network data.
- Network neighborhood structure and diversity
- The effect of content
Long-standing framework: probability of adopting a behavior depends on number of network neighbors already adopting. [Bass 1969, Granovetter 1978, Schelling 1978]

Key issue: qualitative shape of the curves.
- Diminishing returns? Critical mass?

We still have very little understanding of simple threshold models.
Basic Threshold Model

Each node $v$ has $d$ neighbors, chooses threshold $f(v)$ at the start, from a distribution $\mu$ over $\{0, 1, 2, \ldots, d + 1\}$.

- $v$ will adopt as soon as it has $f(v)$ adopting neighbors.

Despite simple formulation, a challenging model to analyze.

- Special-case results for diminishing thresholds ($\mu(1) \geq \mu(2) \geq \cdots$) [Kempe-Kleinberg-Tardos 03, Mossel-Roch 07].
- Special-case results when graph $G$ is a tree [Dodds-Watts 04], lattice [Cox-Durrett 91], or clique [Granovetter 78, Schelling 78].
- General networks with $d$ neighbors per node [Blume-Easley-Kleinberg-Kleinberg-Tardos 11].
Cliques vs. Trees

Subtle trade-offs between cliques and trees in this model, based on “structural diversity” of neighbors.

- Trees can have high contagion probability due to large size.
- Cliques can have high contagion probability because of correlated outcomes among neighbors.

Compare cliques vs. trees on distributions

$$(\mu(0), \mu(1), \mu(2)) = (s, t, 1 - s - t)$$

where $s$ and $t$ are both small.

Lower contagion prob. for $(s, 1 - s, 0)$ and $(s, 0, 1 - s)$

Lower contagion prob. for $(s, \varepsilon, 1 - s - \varepsilon)$
Spread of Information

In on-line data:

- Can we use thresholds for prediction tasks?
- Does structural diversity of network neighbors play a role?
- Does the nature of the content being transmitted play a role?
Decisions in Social Media

Social networks where people make decisions about new behaviors.

- User-defined groups in on-line communities; participation in on-line collaborative projects; decision to use a hashtag on Twitter; ...

- Many instances in Facebook data: accepting an invitation to join the site; clicking on an ad; liking a page; commenting on a post.

Does set/structure of adopting neighbors help predict tendency to adopt?

Diffusion Curves

Long-standing framework: probability of adopting a behavior depends on number of network neighbors already adopting.  

![Graphs showing prob. of adoption vs. number of friends adopting]

Key issue: qualitative shape of the curves.
- Diminishing returns? Critical mass?
**Diffusion Curves**

(a) Joining a LiveJournal group  
   [Backstrom et al. 06]

(b) Editing a Wikipedia article  
   [Crandall et al. 08]

(c) Purchasing a product.  
   [Leskovec et al. 06]
You’re more likely to do something when more friends are doing it. Why is that?

The issue of homophily/selection vs. influence

[Cohen 77, Kandel 78, Manski 93, Aral et al. 09, Shalizi-Thomas 11]

An experiment to sort out these effects

[Bakshy-Eckles-Yan-Rosenn 2012]
Structural Diversity

Dependence on number of friends: a first step toward general prediction.

- Given the full pattern of connections among your friends, estimate probability of adopting a new behavior.

Structural diversity
[Ugander-Backstrom-Marlow-Kleinberg]
- Data from invitations to join Facebook.
Design questions: Many ways to show present someone with information; choices must now be made automatically billions of times per day.

- Incentives to propagate information: e.g. DARPA Network Challenge [Pickard et al 2011], Bitcoin [Babaioff et al 2012].

- Simultaneous evolution of network structure and behavior.
Reflections

- Many models deal with nodes, edges, time-stamps, discrete adoption.
- Textual content: Tracking quoted phrases through the system [Leskovec-Backstrom-Kleinberg 2009]
- Learning textual features in a corpus of viral text: movie quotes [DanescuNiculescuMizil-Cheng-Kleinberg-Lee 2012]