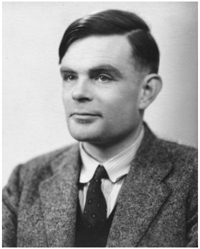
# Bursts, Cascades, and Hot Spots: Algorithmic Models of Social Phenomena

Jon Kleinberg

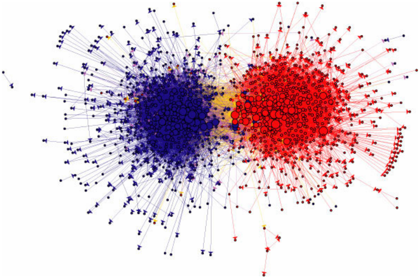
Cornell University

**Including joint work with**
**Lars Backstrom, Larry Blume, David Crandall, David Easley, Dan Huttenlocher, Bobby Kleinberg, Ravi Kumar, Cameron Marlow, Éva Tardos, Johan Ugander.**

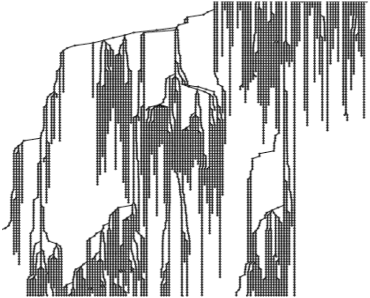
# The Many Faces of Computing




# The Terrain of the On-Line World



Political blogs
(Adamic-Glance, 2005)



Anti-war chain letter (LibenNowell-Kleinberg 2008)

The terrain of the on-line world is
geographic, social, and graph-theoretic.

- Computations on graphs form a central part of our understanding.

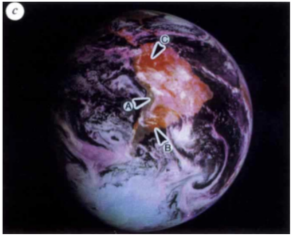
# Is There Life on Earth?

## A search for life on Earth from the Galileo spacecraft

**Carl Sagan[*], W. Reid Thompson[*], Robert Carlson[†], Donald Gurnett[‡] & Charles Hord[§]**

[*] Laboratory for Planetary Studies, Cornell University, Ithaca, New York 14853, USA
[†] Atmospheric and Cometary Sciences Section, Jet Propulsion Laboratory, Pasadena, California 91109, USA
[‡] Department of Physics and Astronomy, University of Iowa, Iowa City, Iowa 52242-1479, USA
[§] Laboratory for Atmospheric and Space Physics, University of Colorado, Boulder, Colorado 80309, USA

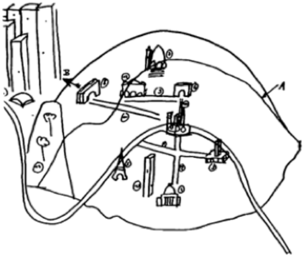
**In its December 1990 fly-by of Earth, the Galileo spacecraft found evidence of abundant gaseous oxygen, a widely distributed surface pigment with a sharp absorption edge in the red part of the visible spectrum, and atmospheric methane in extreme thermodynamic disequilibrium; together, these are strongly suggestive of life on Earth. Moreover, the presence of narrow-band, pulsed, amplitude-modulated radio transmission seems uniquely attributable to intelligence. These observations constitute a control experiment for the search for extraterrestrial life by modern interplanetary spacecraft.**

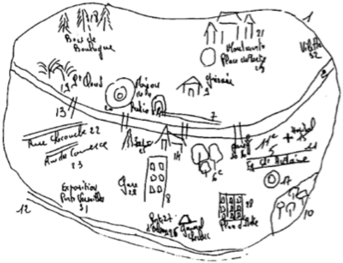



# A Portion of the Earth, as Seen from Flickr

# Maps as Social Objects

> A city is a social fact. We would all agree to that. But we need to add an important corollary: the perception of a city is also a social fact, and as such needs to be studied in its collective as well as its individual aspect. It is not only what *exists* but what is *highlighted* by the community that acquires salience in the mind of the person. A city is as much a collective representation as it is an assemblage of streets, squares, and buildings.





[Jodelet-Milgram 1976]

# Organize Traces of Human Behavior Around Hot Spots

Organize activities around "hot-spots" in space and time.

Use geo-tagged data from millions of people,
via photos, search engine queries, mobile devices

- [Backstrom-Kleinberg-Kumar-Novak 2008, Kennedy-Naaman 2008, Crandall-Backstrom-Huttenlocher-Kleinberg 2009]

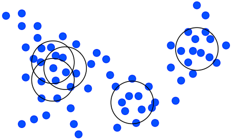
Hot-spot analysis

- Where are the hot-spots?
- How "intense" are they?
- What's distinctive in them?

Backstrom et al 2008, Kennedy-Naaman 2008, Crandall et al 2009

# How Do We Find and Describe Hot-Spots?

Start with a local-search heuristic to find hot-spots.



Identifying Distinctive Features of a Hot-Spot.

- First: textual tags.
- Significance of a tag $t$ at a hot-spot based on Bayes' Rule.
- Roughly: probability of seeing this density of photos w/ tag $t$, if tag were generated from world's background distribution?
- Next: identify distinctive photos at a location [Snavely-Seitz-Szeliski 2006, Kennedy-Naaman 2008]

Try this for the global Flickr dataset, using two mean-shift scales:

- 100 km: metropolitan scale
- 100 m: landmark scale

| | Top landmark | 2nd landmark | 3rd landmark | 4th landmark |
|---|---|---|---|---|
| 1. **manhattan** | empirestate | timessquare | grandcentral | applestore |
| 2. **london** | trafalgarsquare | tatemodern | eye | bigben |
| 3. **sanfrancisco** | coittower | sealions | unionsquare | lombardstreet |
| 4. **losangeles** | disneyland | hollywood | gettycenter | disneyhall |
| 5. **paris** | eiffel | cathedral | sacrecoeur | pyramid |
| 6. **washingtondc** | lincolnmemorial | monument | wwiimemorial | capitol |
| 7. **chicago** | cloudgate | michiganavenue | gehry | artinstitute |
| 8. **seattle** | spaceneedle | market | emp | library |
| 9. **boston** | fenwaypark | trinitychurch | faneuilhall | publicgarden |
| 10. **sandiego** | balboapark | sandiegozoo | seals | ussmidway |
| 11. **amsterdam** | dam | annefrank | nieuwmarkt | museumplein |
| 12. **rome** | colosseum | sanpietro | pantheon | fontanaditrevi |
| 13. **barcelona** | sagradafamilia | parcguell | boqueria | casamil |
| 14. **berlin** | brandenburggate | reichstag | potsdamerplatz | holocaustmemorial |
| 15. **monterey** | montereybay | downtown | canneryrow | boardwalk |
| 16. **lasvegas** | paris | bellagio | mgm | hooverdam |
| 17. **toronto** | cntower | phillipssquare | dundassquare | rom |
| 18. **vancouver** | granvilleisland | artgallery | aquarium | downtown |
| 19. **firenze** | cathedral | pontevecchio | firenze | piazzadelcampo |
| 20. **philadelphia** | libertybell | artmuseum | cityhall | jfkplaza |

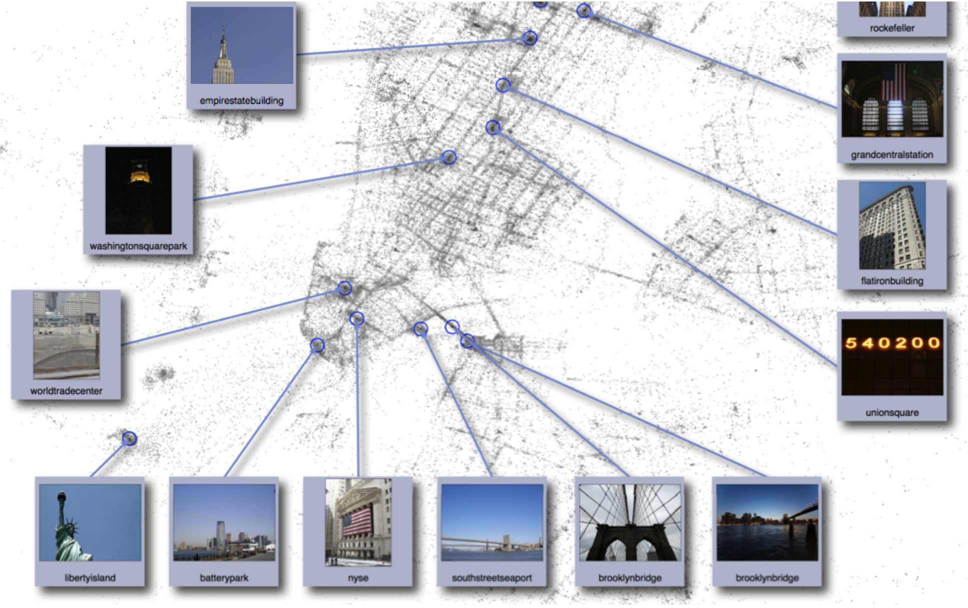
# The Earth, as Seen from Flickr

# U.S. and Canada (Crandall et al 2009)

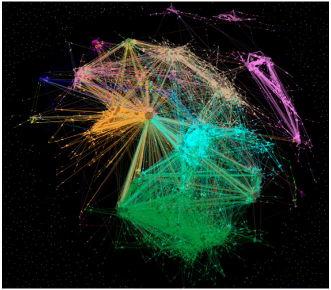
# Europe


edinburghcastle
edinburgh
exchangesquare
manchester
trafalgarsquare
london
damsquare
amsterdam
copenhagen
copenhagen
cathedral
köln
oconnellstreet
dublin
brandenburggate
berlin
bathabbey
bristol
europe
praha
eiffel
paris
rathaus
münchen
praçadocomércio
lisbon
sanmarco
venice
plazamayor
madrid
sagradafamilia
barcelona
casino
monaco
galleria
milano
coliseum
rome
pontevecchio
firenze

# Lower Manhattan



# Bringing Network Structure Into the Picture



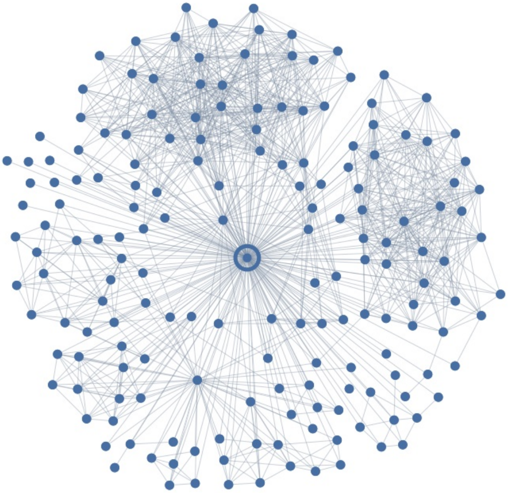

Wu et al 2012

Network structure as a crucial ingredient in new approaches to studying collective human behavior.

- Can we link individual-level models of decision-making to macroscopic models of large networks and populations?
- To what extent is collective human behavior predictable?
- Applications to political processes, economic outcomes, formation of public opinion, collective problem-solving, new kinds of social organization.

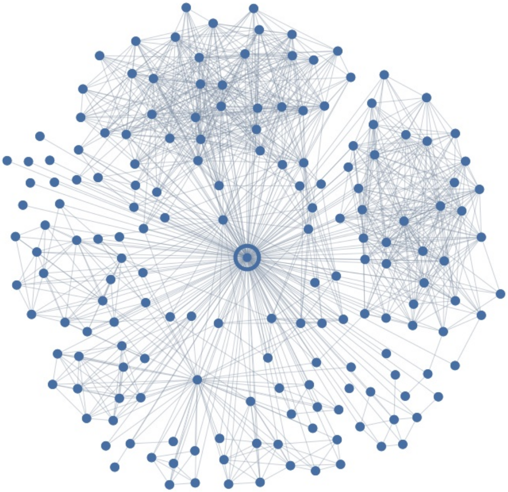
Understanding the strange geography of our collective social experience.

# Network structure via neighborhoods



Start with network neighborhoods

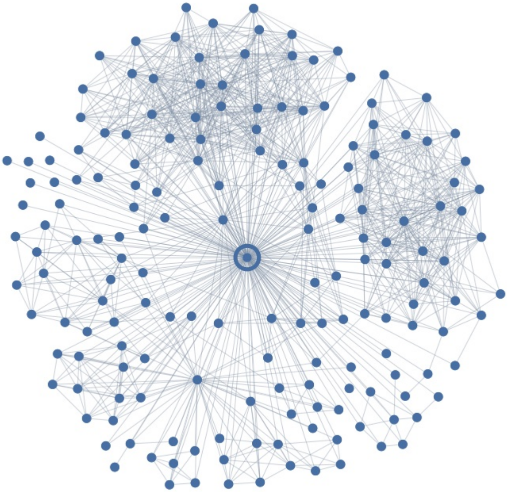
# Network structure via neighborhoods



Start with network neighborhoods

- Think of Facebook not as a billion-node network, but instead as a collection of a billion (relatively dense) small networks.

# Network structure via neighborhoods



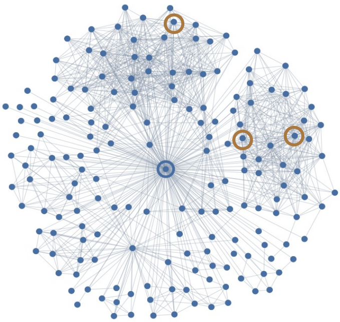
Start with network neighborhoods

- Describe neighborhood $G$ by vector of subgraph frequencies: For small $k$, and each $k$-node graph $H$, let $f_G(H)$ = frac. of $k$-node sets inducing $H$.

# Characterizing neighborhoods

$f_G(H) =$ frac. of $k$-node sets in $G$ that induce $H$.

- Triad census: Davis-Leinhardt 71
- Network motifs: Milo et al 02
- Frequent subgraph mining: Yan-Han 02, Kuramochi-Karypis 04
- Subgraph homomorphism density: Borgs et al 06
- Characterizing neighborhoods: Ugander et al 13

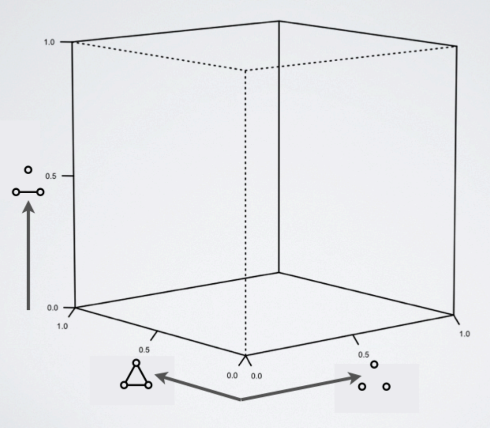


$(x_1, x_2, x_3, x_4) = (0.18, 0.37, 0.14, 0.31)$

$(y_1, y_2, y_3, y_4, y_5, y_6, y_7, y_8, y_9, y_{10}, y_{11})$

# The geography of Facebook neighborhoods

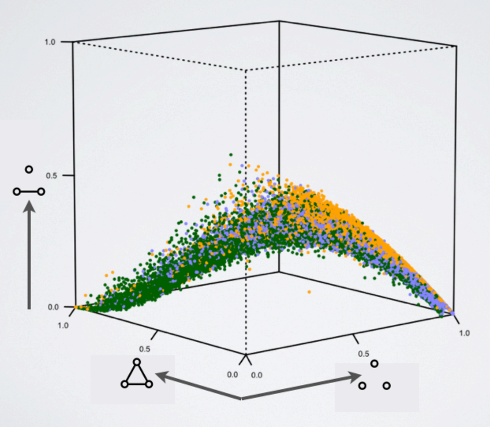
Axes: triad frequencies



# The geography of Facebook neighborhoods

Axes: triad frequencies

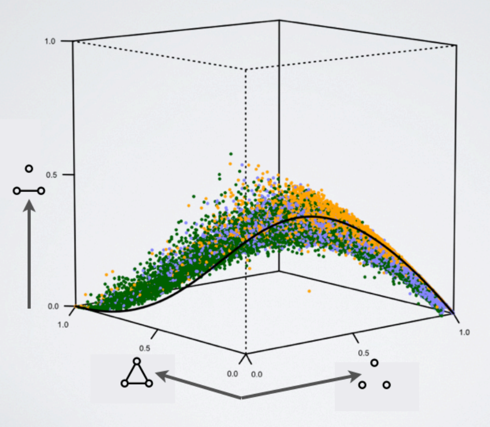
- "Coastlines:" freq of 1-edge triad is $\leq 3/4$.
- Unpopulated areas: freq of 2-edge triad never close to $3/4$ in real life.
- Full feasible region would imply [Razborov 2007].



# The geography of Facebook neighborhoods

Axes: triad frequencies

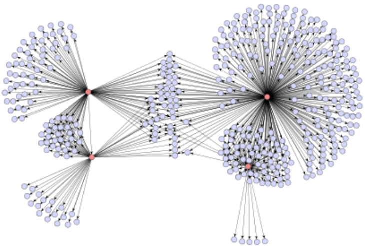
- "Coastlines:" freq of 1-edge triad is $\leq 3/4$.
- Unpopulated areas: freq of 2-edge triad never close to $3/4$ in real life.
- Full feasible region would imply [Razborov 2007].



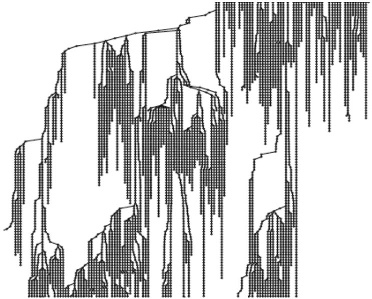

$G_{n,p}$ is the "river" that runs through the points.

- With deviations based on triadic closure and clustering.

# Diffusion and Contagion



Book recommendations (Leskovec et al 2006)



Anti-war chain letter (LibenNowell-Kleinberg 2008)

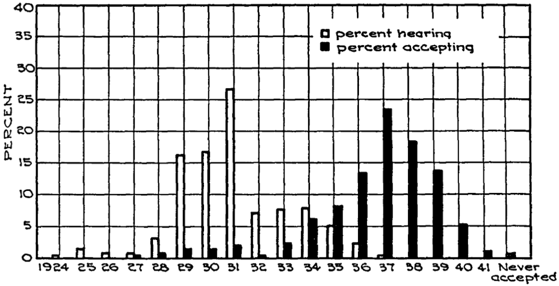
A basic "transport mechanism" for these systems:

- The movement of information through a social network.

Long history of research in diffusion:

- 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]
- "Virality" of news, rumors, marketing strategies, political messages, ...

# Diffusion and Contagion



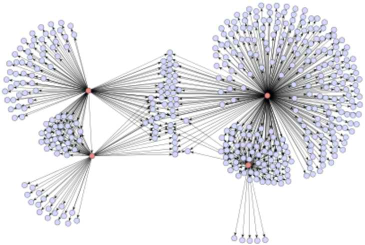

A basic "transport mechanism" for these systems:

- The movement of information through a social network.

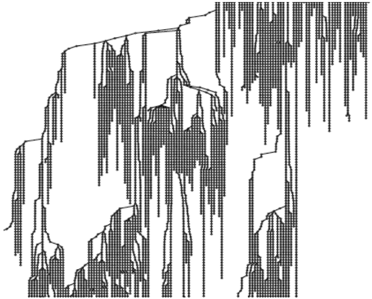
Long history of research in diffusion:

- 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]
- "Virality" of news, rumors, marketing strategies, political messages, ...

# Diffusion and Contagion



Book recommendations (Leskovec et al 2006)



Anti-war chain letter (LibenNowell-Kleinberg 2008)

A basic "transport mechanism" for these systems:

- The movement of information through a social network.

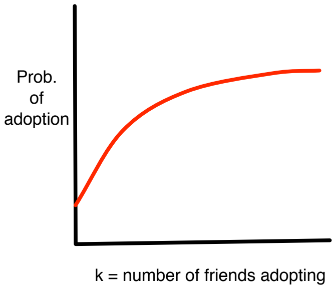
Long history of research in diffusion:

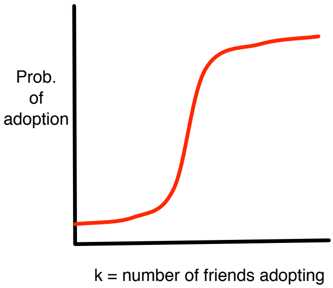
- Agricultural, medical innovations [Ryan-Gross 1943, Coleman et al 1966]
- Media influence and two-stage flow [Lazarsfeld et al 1944]
- Collective action, social movements [McAdam 1986, Chwe 1999]
- "Virality" of news, rumors, marketing strategies, political messages, ...

# Diffusion Curves

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.

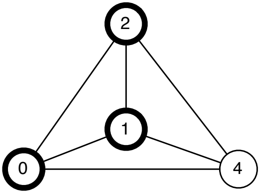
- Diminishing returns? Critical mass?

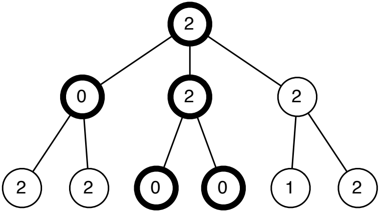
We still have very little understanding of simple threshold models.

# 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

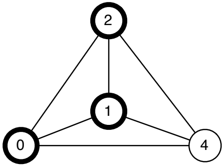
Subtle trade-offs in this model, based on "structural diversity" of neighborhoods.

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

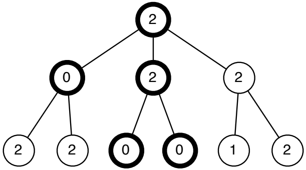
Compare cliques vs. trees on distributions

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

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

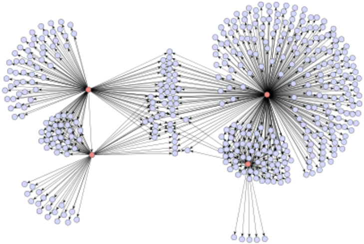


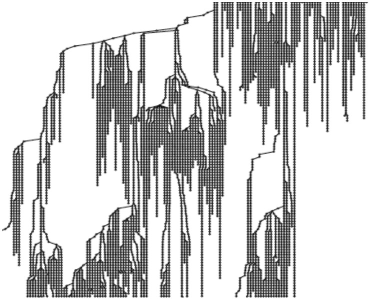

Lower contagion prob. for $(s, 1-s, 0)$ and $(s, 0, 1-s)$




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

# Spread of Information

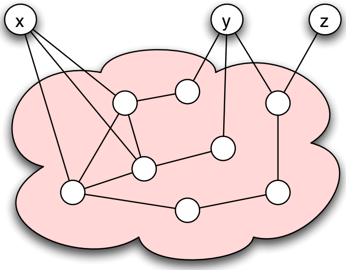

Book recommendations (Leskovec et al 2006)


Anti-war chain letter (LibenNowell-Kleinberg 2008)

In on-line data:

- Can we use thresholds for prediction tasks?
- Does structural diversity of network neighbors 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.

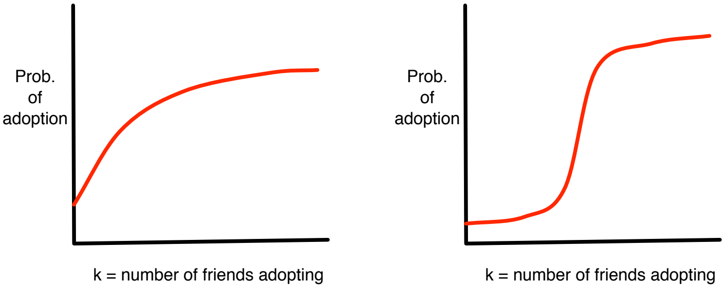
Does set/structure of adopting neighbors help predict tendency to adopt?

Backstrom et al. 2006, Crandall et al. 2008, Romero et al. 2011

# Diffusion Curves

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

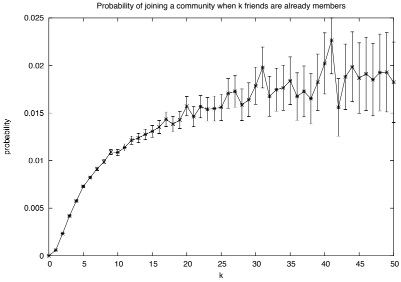
[Bass 1969, Granovetter 1978, Schelling 1978]

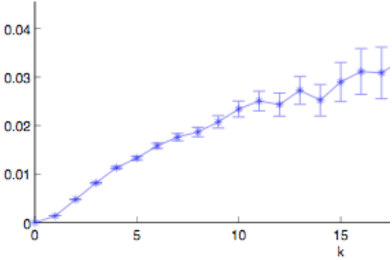


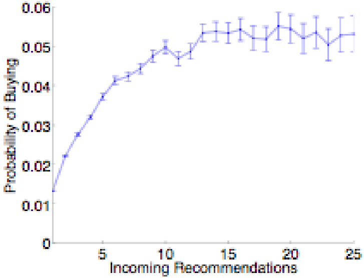

Key issue: qualitative shape of the curves.

- Diminishing returns? Critical mass?

# 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]

# Prediction and Potential Influence

You're more likely to do something when more friends are doing it. Why is that?

The issue of homophily/selection vs. influence

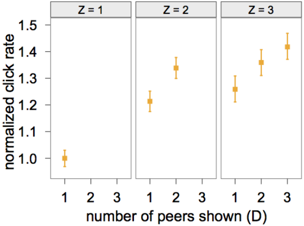
[Cohen 77, Kandel 78, Manski 93, Aral et al. 09, Shalizi-Thomas 11]

---


(a) (b) (c)

An experiment to sort out these effects

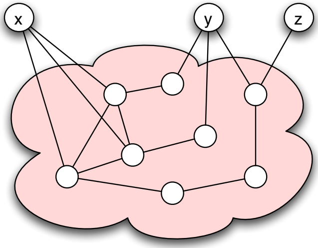
[Bakshy-Eckles-Yan-Rosenn 2012]



# Structural Diversity

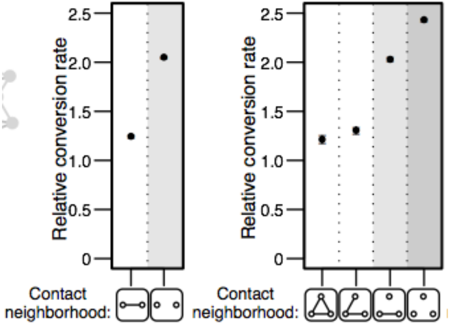
Dependence on number of friends:
a first step toward general prediction.

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



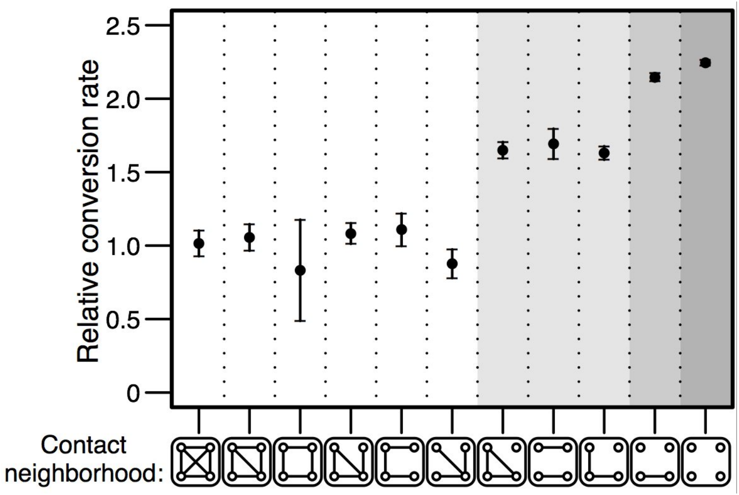

Structural diversity
[Ugander-Backstrom-Marlow-Kleinberg]

- Data from invitations to join Facebook.

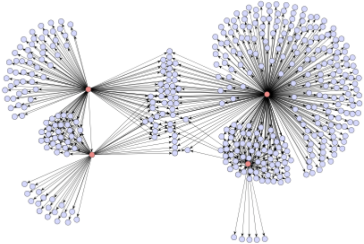


# Structural Diversity

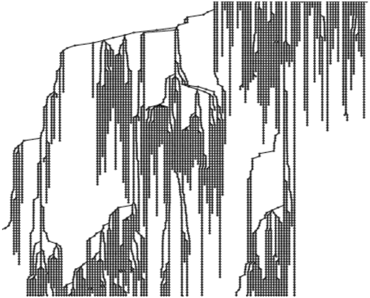
With four neighbors:



# The Flow of Information



Book recommendations (Leskovec et al 2006)



Anti-war chain letter (LibenNowell-Kleinberg 2008)

- Design questions: Many ways to show present someone with information; choices must now be made automatically billions of times per day.
- Graph structure of neighborhoods as a feature in conversational curation [Backstrom-Kleinberg-Lee-DanescuNiculescuMizil 2013]
- Incentives to propagate information: e.g. Query incentive networks [Kleinberg-Raghavan 2005], DARPA Network Challenge [Pickard et al 2011], Bitcoin [Babaioff et al 2012].
- Simultaneous evolution of network structure and behavior.

# Final Reflections

MySpace is doubly awkward because it makes public what should be private. It doesn't just create social networks, it anatomizes them. It spreads them out like a digestive tract on the autopsy table. You can see what's connected to what, who's connected to whom.

– Toronto Globe and Mail, June 2006.

- Social networks — implicit for millenia — are being recorded at high resolution.
- What is the right framework for capturing the structures and phenomena that we see?
- What are the dangers of stockpiling this much personal data?
- An opportunity for fundamental models in computing to inform the next steps on all these questions.

# Final Reflections: The Challenge of Prediction

With accurate models and enough data, can we predict outcomes?

MusicLab [Salganik, Dodds, and Watts 2006]:

Music download site: preview songs, then download.

A "leaderboard" showing the most downloaded songs.

Can we predict which songs will come out on top?

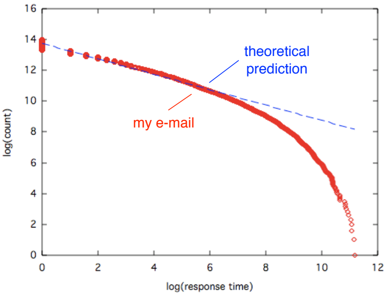
# Final Reflections: Toward a Model of You

Not just data on massive populations, but each of us individually.

Software that understands your behavior better than you do.

Example: How rapidly do you reply to e-mail?

What fraction of your messages do you answer on the day of arrival? 1 day later? 2 days? 3 days? ... 60 days? ...



# Final Reflections

MySpace is doubly awkward because it makes public what should be private. It doesn't just create social networks, it anatomizes them. It spreads them out like a digestive tract on the autopsy table. You can see what's connected to what, who's connected to whom.

– Toronto Globe and Mail, June 2006.

- Social networks — implicit for millenia — are being recorded at high resolution.
- What is the right framework for capturing the structures and phenomena that we see?
- What are the dangers of stockpiling this much personal data?
- An opportunity for fundamental models in computing to inform the next steps on all these questions.