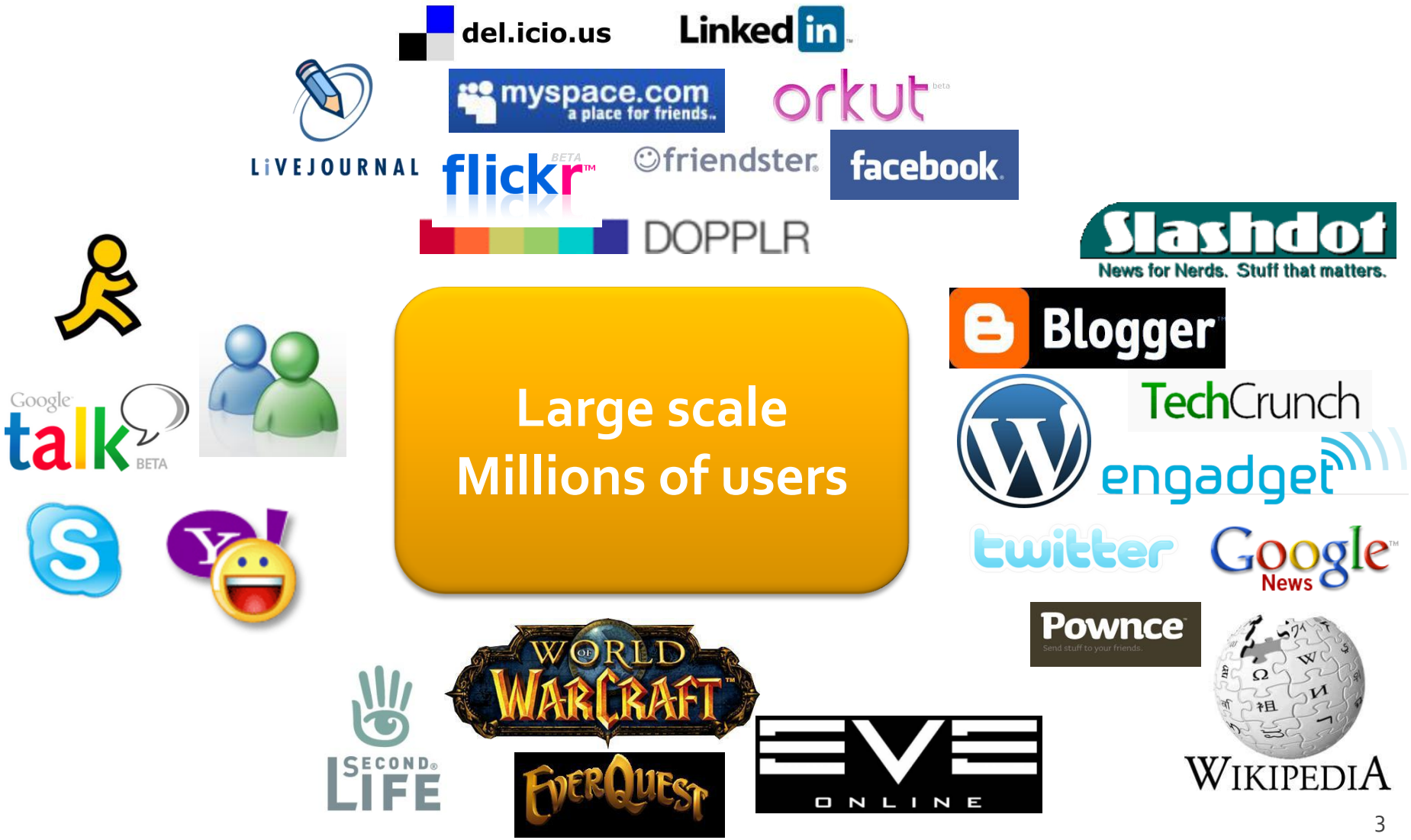
Jure Leskovec
PhD: Machine Learning Department, CMU
Now: Computer Science Department, Stanford University

Dynamics of large networks

Web today – Diverse applications

Diverse on-line computing applications

Web today – Millions of users



Large scale
Millions of users

Web today – Rich content

Rich user generated content

Web today – Highly dynamic

Content is constantly being updated and changed

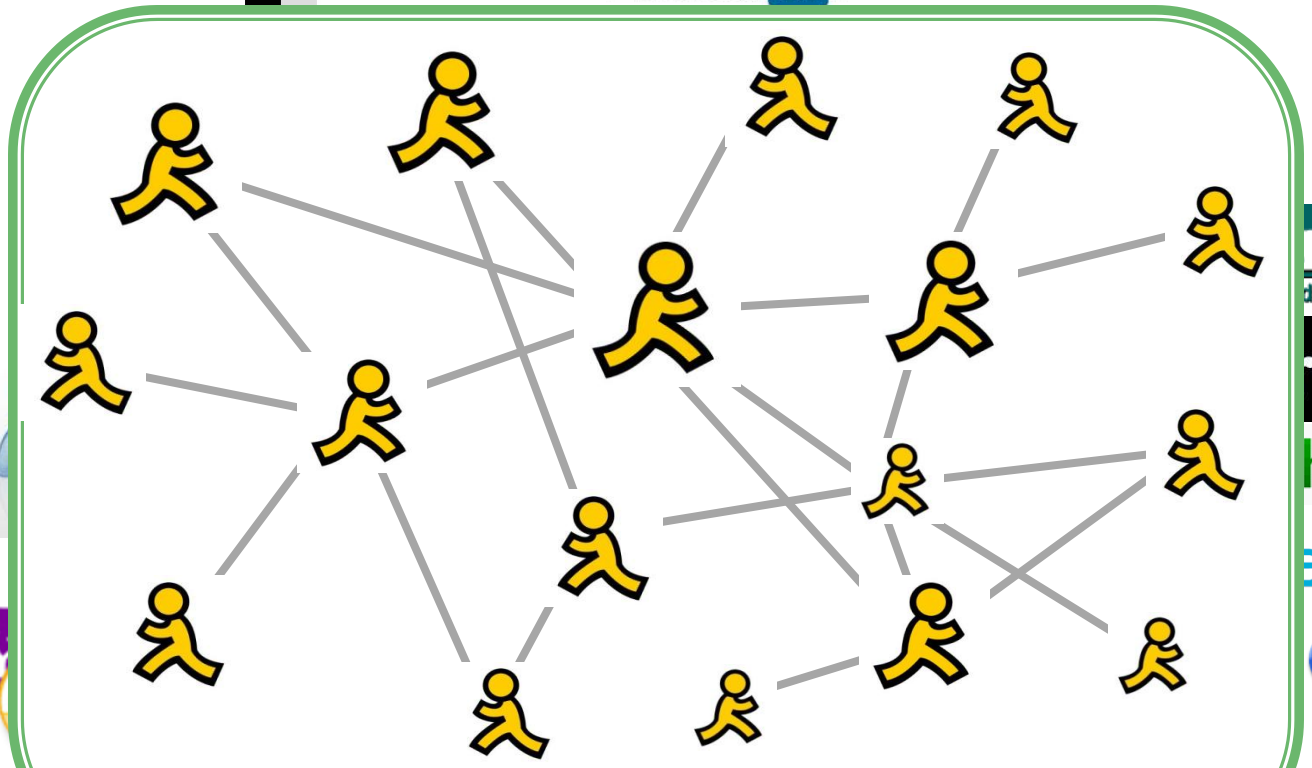
Web today – Traces of activity

Massive traces of human social activity are collected

Web today – Rich interactions

Rich interactions between users and content

Web today – Interaction networks



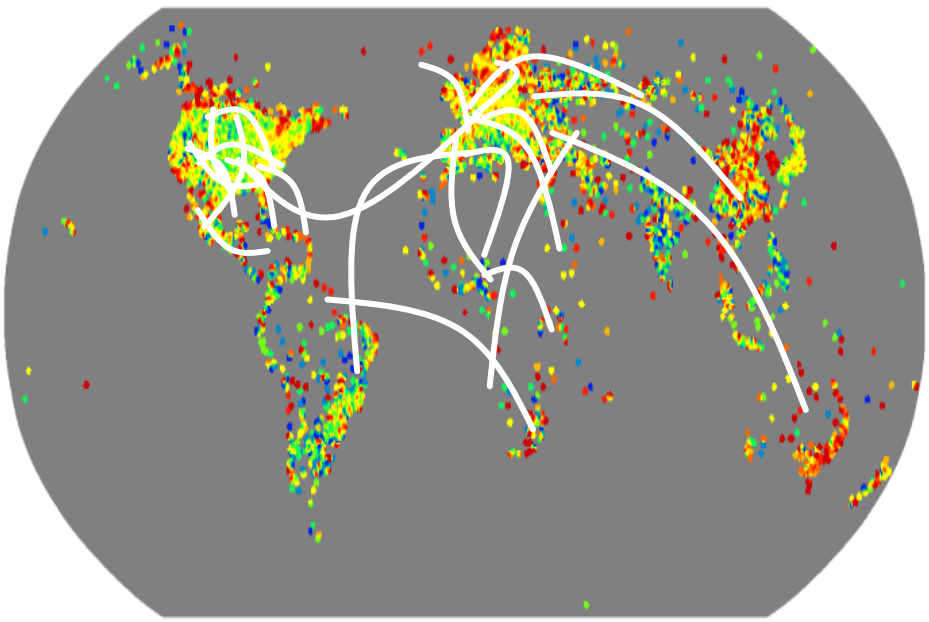
Modeled as interaction **network**

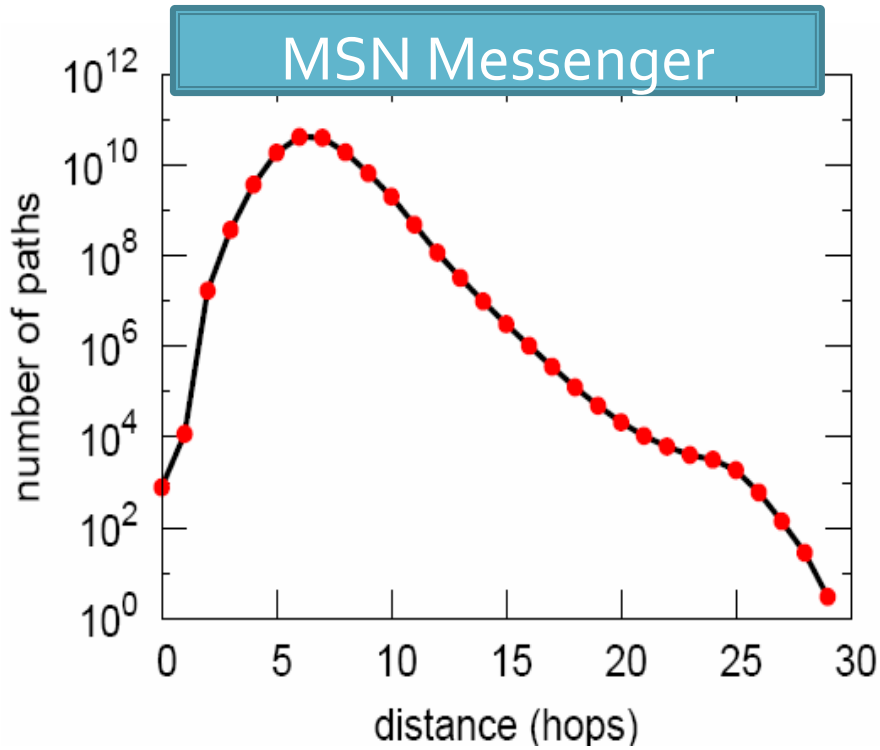
Web today – Immense posibilities

Web is like a "laboratory" for studying millions of humans

[w/ Horvitz WWW '08]

Testing the Small-world hypothesis





Network of who talks to whom on MSN Messenger: 240M nodes, 1.3 billion edges

Average path length is **6.6**
90% of nodes is reachable <8 steps

Why study web and networks?

- **Build understanding and theory:**
  - How users create content and interact with it and among themselves?
- **Build better on-line applications:**
  - How to design better services and algorithms?

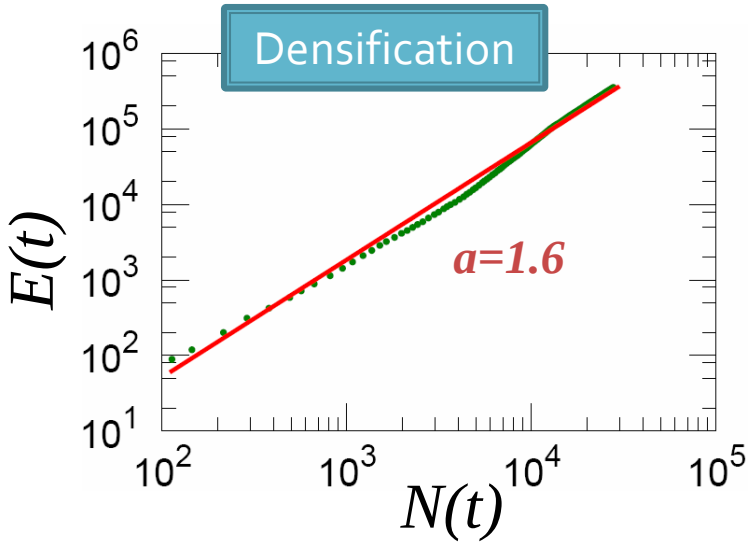
Thesis: Network dynamics

- Network evolution
  - How network structure changes as the network grows and evolves?
- Diffusion and cascading behavior
  - How does information and diseases spread over networks?

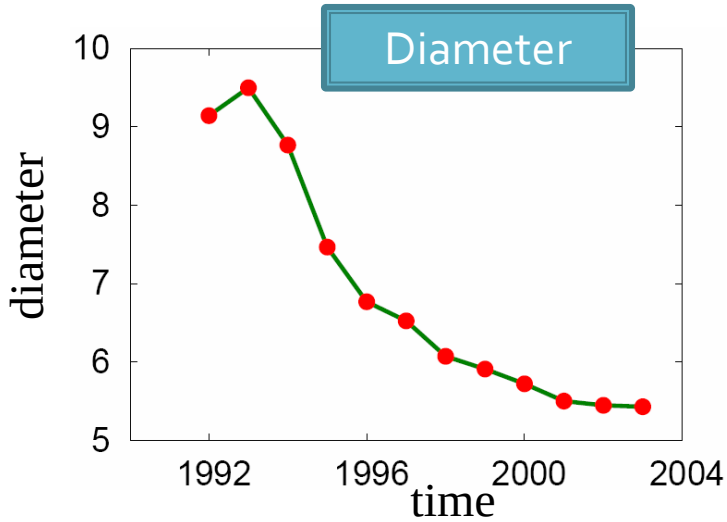
[w/ Kleinberg and Faloutsos KDD '05]

Thesis: Network evolution (1)

- **Observations:** measuring
  - **Q:** How do does network structure evolve?
  - **A:** Networks densify and diameter shrinks



$$E(t) \propto N(t)^a$$

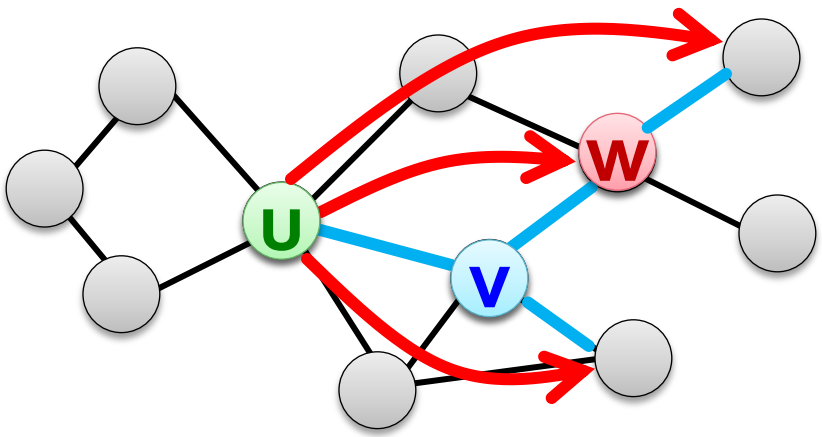


Network diameter shrinks over time

[w/ Kleinberg and Faloutsos KDD '05]

Thesis: Network evolution (2)

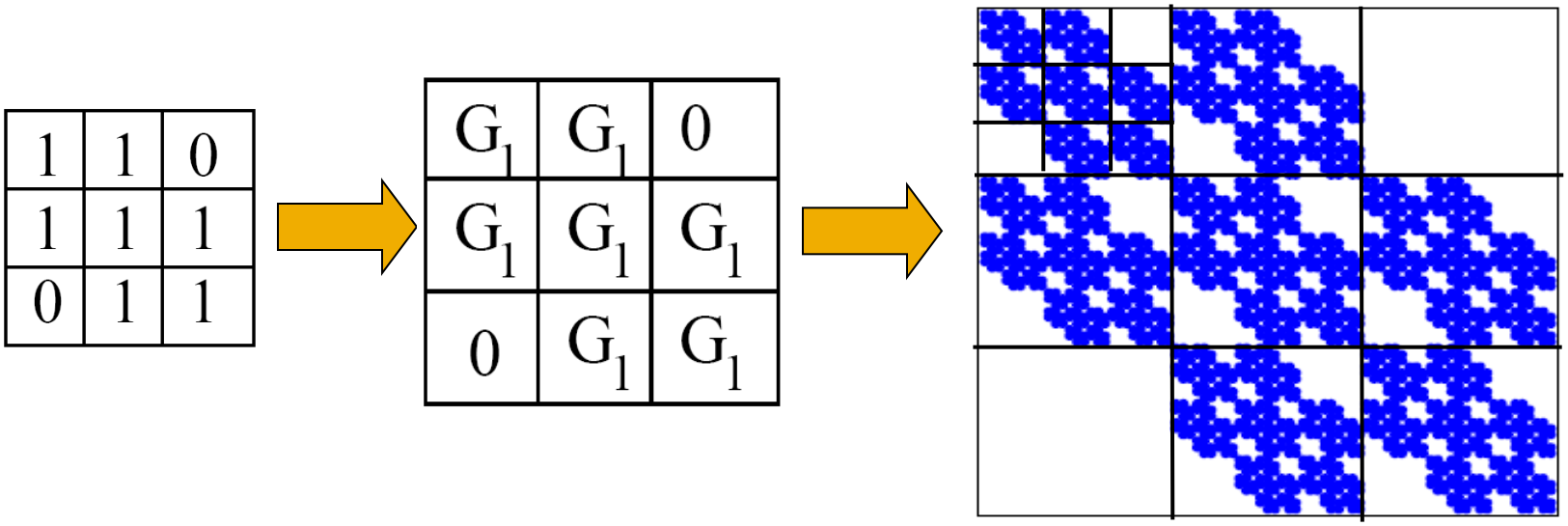
- **Models:** understanding
  - **Q:** What is a good model/explanation?
  - **A:** Forest Fire model



[w/ Faloutsos ICML '07]

Thesis: Network evolution (3)

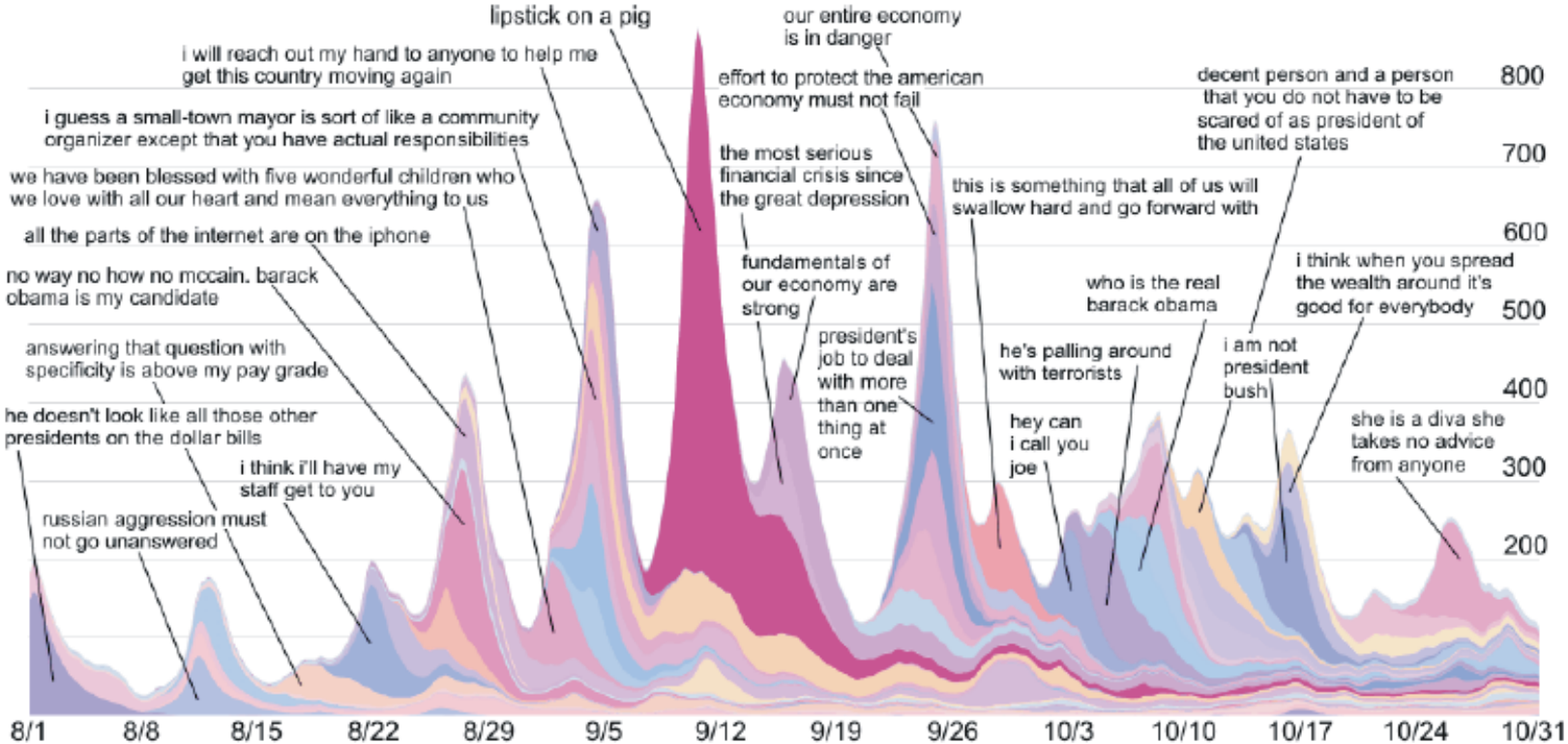
- **Algorithms:** doing things better
  - **Q:** How to generate realistic graphs?
  - **A:** Kronecker graphs



[w/ Backstrom and Kleinberg KDD '09]

Thesis: Diffusion and cascades (1)

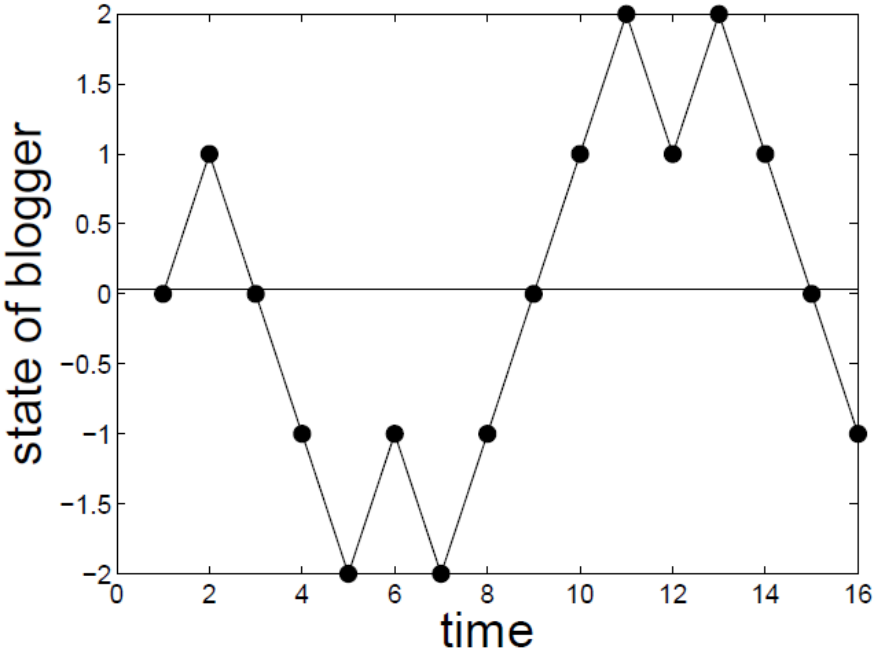
- **Observations:** measuring
  - **Q:** How does information propagate over the web?



[w/ Goetz, McGlohon and Faloutsos ICWSM '09]

Thesis: Diffusion and cascades (2)

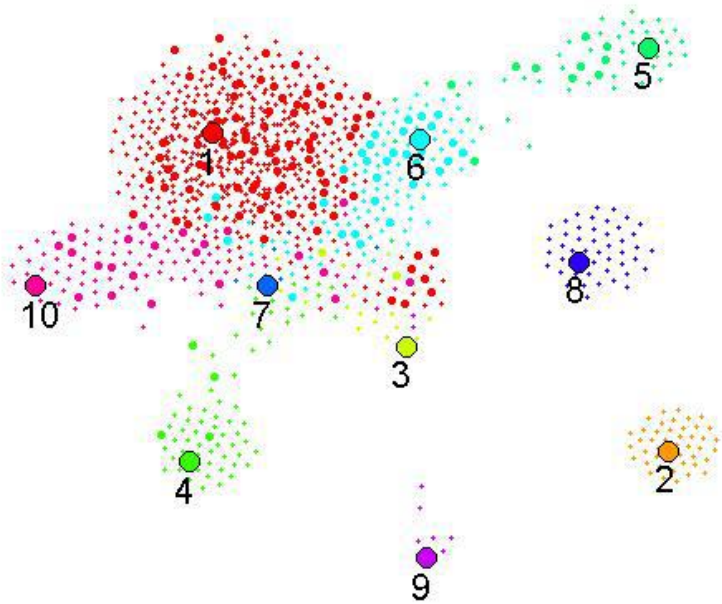
- **Models:** understanding
  - **Q:** How to information propagation
  - **A:** Zero-crossing model



[w/ Krause, Guestrin, Glance and Faloutsos KDD '07]

Thesis: Diffusion and cascades (3)

- **Algorithms:** doing things better
  - **Q:** How to identify influential nodes and epidemics?
  - **A: CELF** (cost-effective lazy forward-selection)



Thesis: Size matters

- **Massive data:**
  - MSN Messenger network [w/ Horvitz, WWW '08]
    - 240M people, 255B messages
  - Product recommendations [w/ Adamic, Huberman EC '06]
    - 4M people, 16M recommendations
  - Blogosphere [w/ Backstrom, Kleinberg KDD '09]
    - 164M posts, 127M links
- **Benefit:** Properties become "visible"
  - E.g.: In large networks only small clusters exist

    [w/ Dasgupta, Lang and Mahoney WWW '08]

Thesis: Reflections

- Why are networks the way they are?
- Only recently have basic properties been observed on a large scale
  - Confirms social science intuitions; calls others into question
- What are good tractable network models?
  - Builds intuition and understanding
- Benefits of working with large data
  - Observe structures not visible at smaller scales

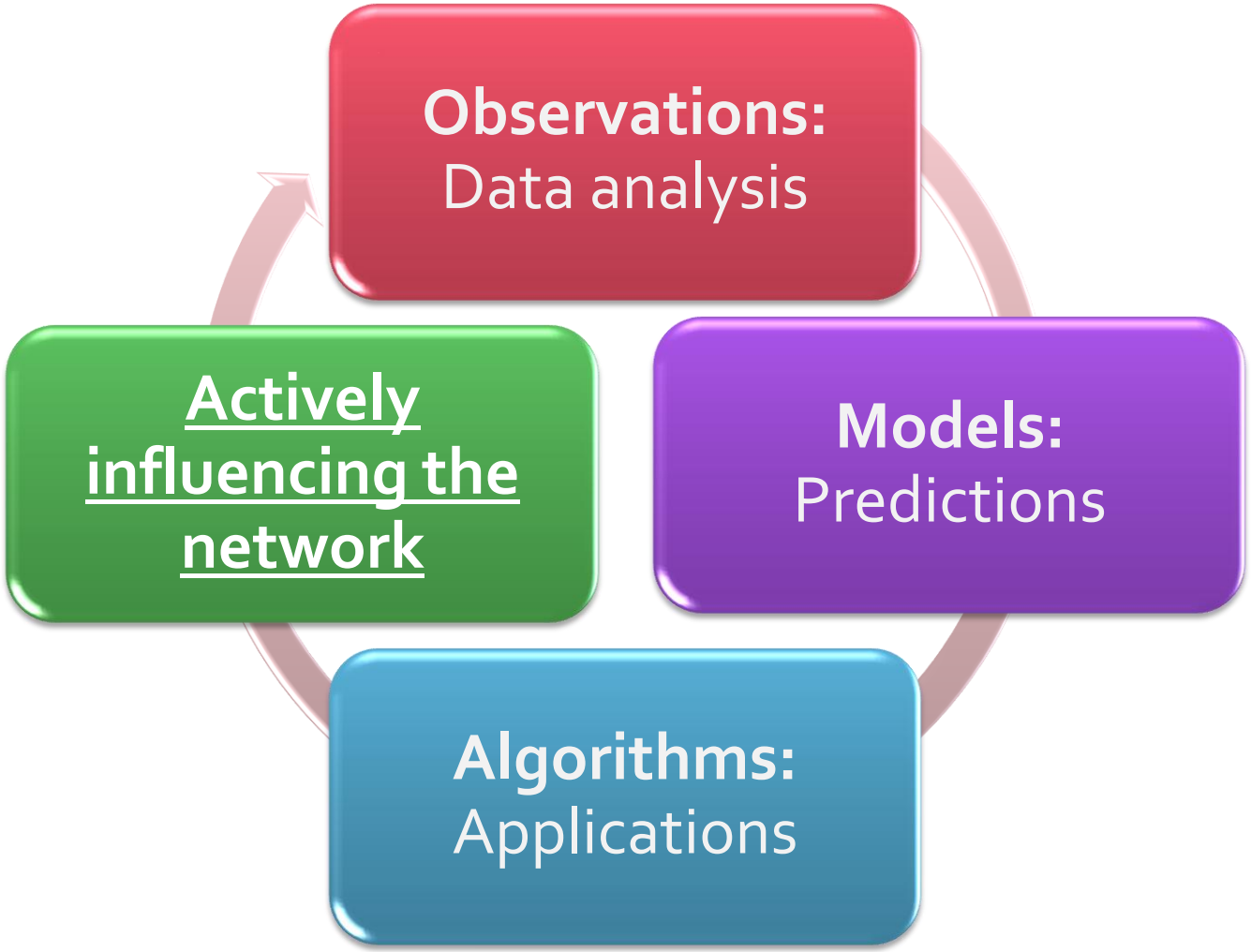
Challenges

- Why are networks the way they are?
- Richer networks, richer more detailed data
  - New findings and observations
- More accurate models
  - Predictive modeling
- Large scale
  - Will find phenomena and emergent patterns not visible at small scales

Future directions

- Global predictive models
  - Online massively multi-player games
- Information diffusion
  - When, where and what post will create a cascade?
- Where to tap the network to get right effects?
  - Social Media Marketing
- Steering the evolution of the network
  - Cultivating the social network

What's next?


Observations:
Data analysis
Models:
Predictions
Algorithms:
Applications
Actively influencing the network

Acks

Susan Dumais
Andrew Tomkins
Ravi Kumar
Lada Adamic
Bernardo Huberman
Kevin Lang
Michael Manohey
Zoubin Gharamani
Anirban Dasgupta
Tom Mitchell
John Lafferty
Larry Wasserman
Avrim Blum
Sam Madden
Michalis Faloutsos
Ajit Singh

THANKS!

Questions:

jure@cs.stanford.edu

Thesis:

Google: Jure thesis

Data + Code:

http://snap.stanford.edu