

Social and Information Network Analysis: Review of Key Concepts

CS224W: Social and Information Network Analysis
Jure Leskovec, Stanford University
<http://cs224w.stanford.edu>



Networks



How do we reason about networks?

Reasoning About Networks

How do we reason about networks?

- **Empirical:** Study network data to find organizational principles
 - **Mathematical models:** Probabilistic, graph theory
 - **Algorithms** for analyzing graphs
-
- **What do we hope to achieve from models of networks?**
 - Patterns and statistical **properties** of network data
 - **Design principles** and **models**
 - **Understand** why networks are organized the way they are (Predict behavior of networked systems)

Networks: Structure & Process

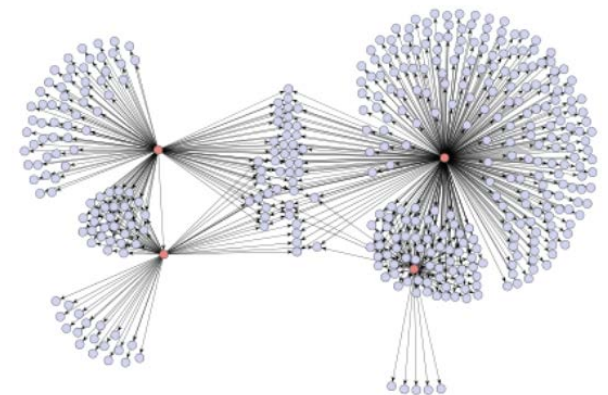
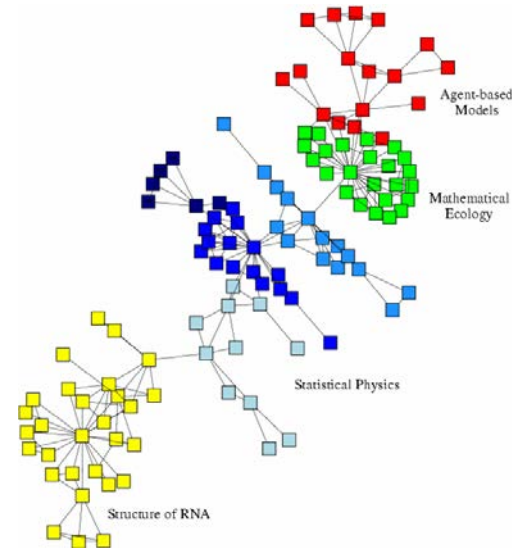
What do we study in networks?

■ Structure and evolution:

- What is the structure of a network?
- Why and how did it become to have such structure?

■ Processes and dynamics:

- Networks provide “skeleton” for spreading of information, behavior, diseases



What We Have Covered

- Network diameter
- Edge clustering
- Scale-free networks
- Strength of weak ties
- Core-periphery structure
- Densification power law
- Shrinking diameters
- Structural Balance
- Status Theory
- Memetracking
- Small-world model
- Erdős-Renyi model
- Preferential attachment
- Network cascades
- Independent cascade model
- Decentralized search
- PageRank
- Hubs and authorities
- Girvan-Newman
- Modularity
- Clique percolation
- Supervised random walks
- Influence maximization
- Outbreak detection
- Linear Influence Model
- Network Inference
- Kronecker Graphs
- Bow-tie structure

How It All Fits Together

Observations

Small diameter,
Edge clustering

Scale-free

Strength of weak ties,
Core-periphery

Densification power law,
Shrinking diameters

Patterns of signed edge
creation

Viral Marketing, Blogosphere,
Memetracking

Models

Small-world model,
Erdős-Renyi model

Preferential attachment,
Copying model

Kronecker Graphs

Microscopic model of
evolving networks

Structural balance,
Theory of status

Independent cascade model,
Game theoretic model

Algorithms

Decentralized search

PageRank, Hubs and
authorities

Community detection:
Girvan-Newman, Modularity

Link prediction,
Supervised random walks

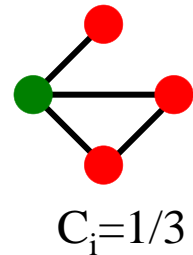
Models for predicting
edge signs

Influence maximization,
Outbreak detection, LIM

Small World Phenomena

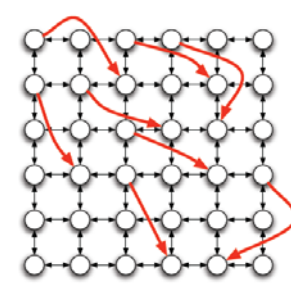
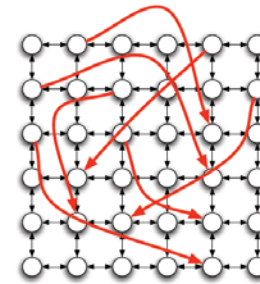
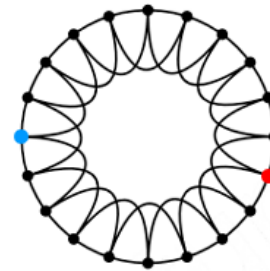
■ Observations:

- **Six degrees of separation**
 - Networks have small diameters
- **Edges in the networks cluster**
 - Clustering coefficient



■ Models:

- **Erdős-Renyi model**
 - Baseline model for networks
- **The Small-World model**
 - Small diameter and clustered edges



■ Algorithms:

- **Decentralized search in networks**
 - Kleinberg's model and algorithm

$$P(u \rightarrow v) \sim d(u, v)^{-\alpha}$$

Scale-Free Networks

■ Observations:

■ Power-law degrees

- Degrees are heavily skewed

■ Network resilience

- Networks are resilient to random attacks

■ Models:

■ Preferential attachment

- Rich get richer

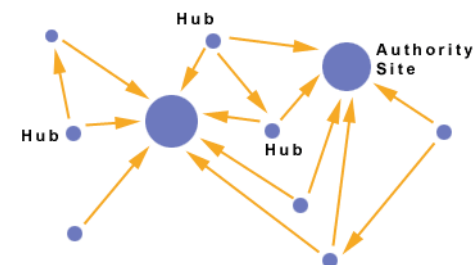
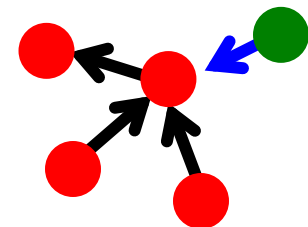
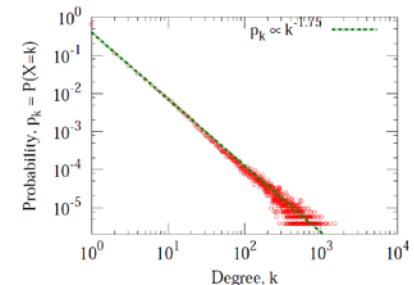
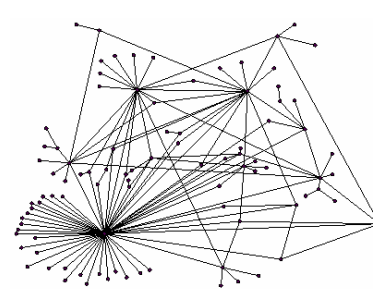
■ Algorithms:

■ Hubs and Authorities

- Recursive: $a_i = \sum_{j \rightarrow i} h_j$, $h_i = \sum_{i \rightarrow j} a_j$

■ PageRank

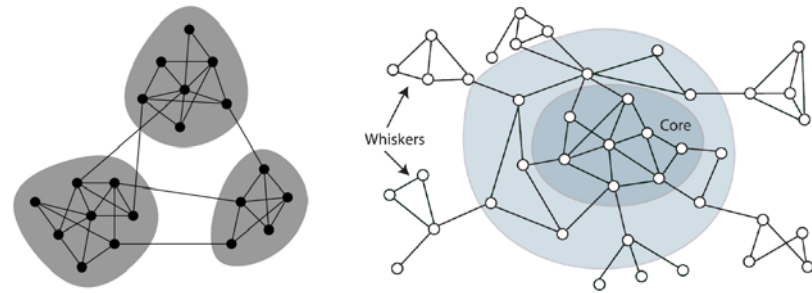
- Recursive formulation, Random jumps



Community Detection

■ Observations:

- Strength of weak ties
- Core-periphery structure



■ Models:

- Kronecker graphs model

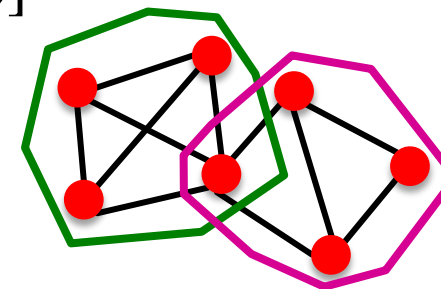
■ Algorithms:

- Girvan-Newman (Betweenness centrality)
- Modularity optimization
 - #edges within group – $E[\text{\#edges within group}]$
- Clique Percolation Method
 - Overlapping communities

1	1	0
1	1	1
0	1	1



K_1	K_1	0
K_1	K_1	K_1
0	K_1	K_1



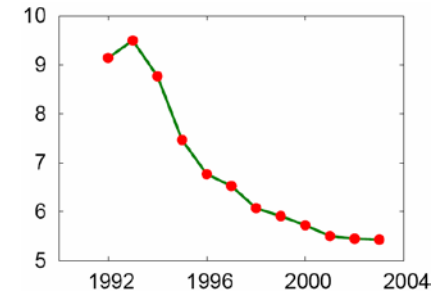
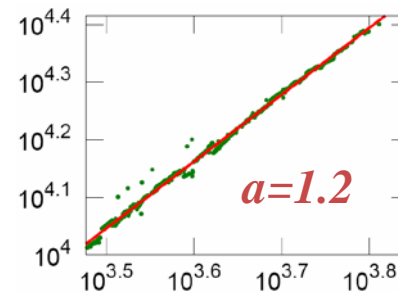
Network Evolution

■ Observations:

■ Densification Power Law

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

■ Shrinking Diameter



■ Models:

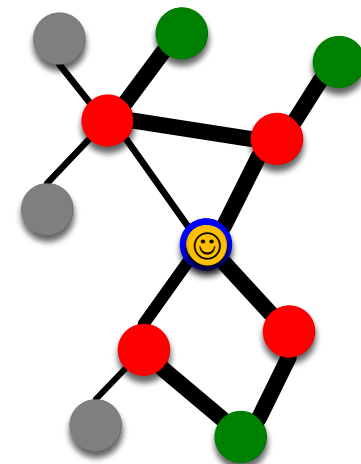
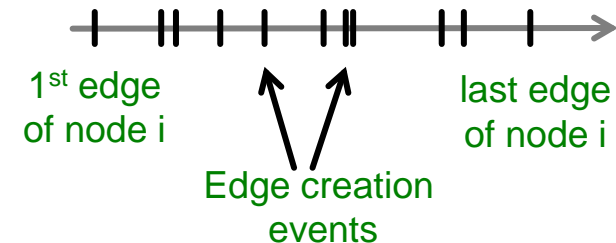
■ Microscopic Network Evolution

- Exponential life-times, Evolving sleeping times
- Random-Random edge attachment

■ Algorithms:

■ Link prediction

■ Supervised Random Walks



Signed Networks

■ Observations:

- Signed link creation
- +links are more embedded

■ Models:

■ Structural Balance

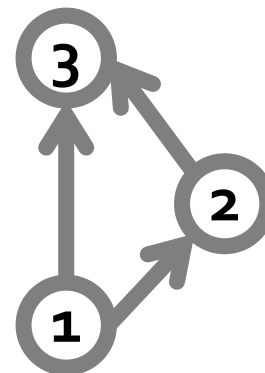
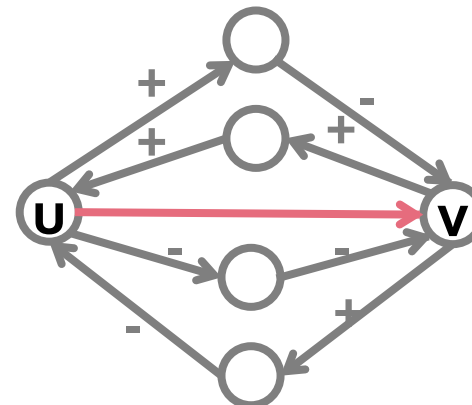
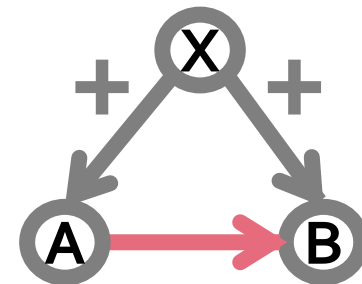
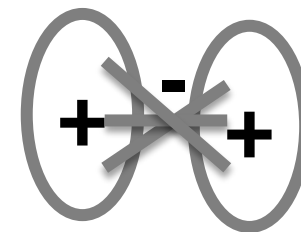
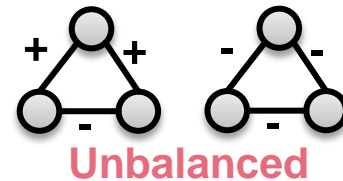
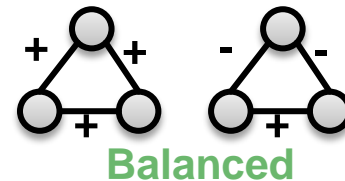
- Coalition structure of networks

■ Status Theory

- Global node status ordering

■ Algorithms:

- Predicting edge signs



Network Diffusion (1)

■ Observations:

■ Tracking contagions

- Viral Marketing
- Hyperlinks

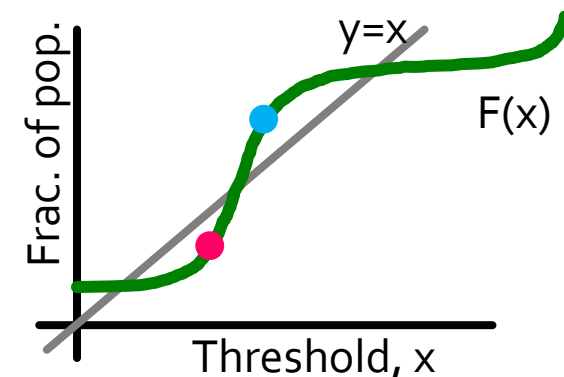
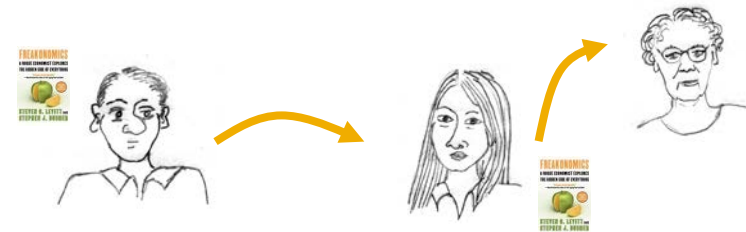
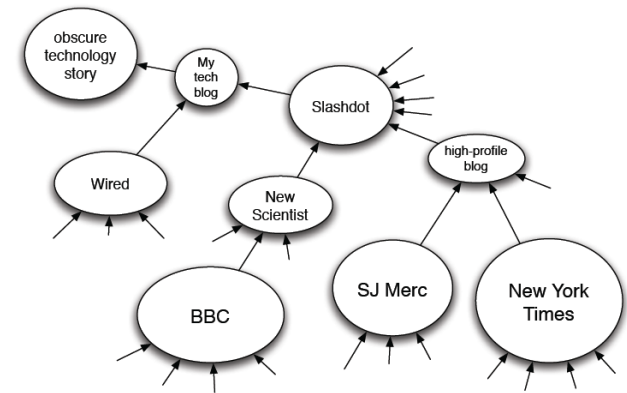
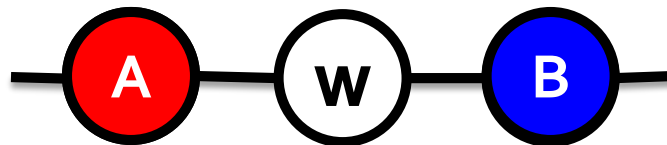
■ Models – Decision Based

■ Collective action:

- Node i will adopt the behavior iff at least t_i other people are adopters

■ Game theoretic model:

- Payoffs, Competing products



Network Diffusion (2)

■ Models – Probabilistic

■ Independent Cascade Model

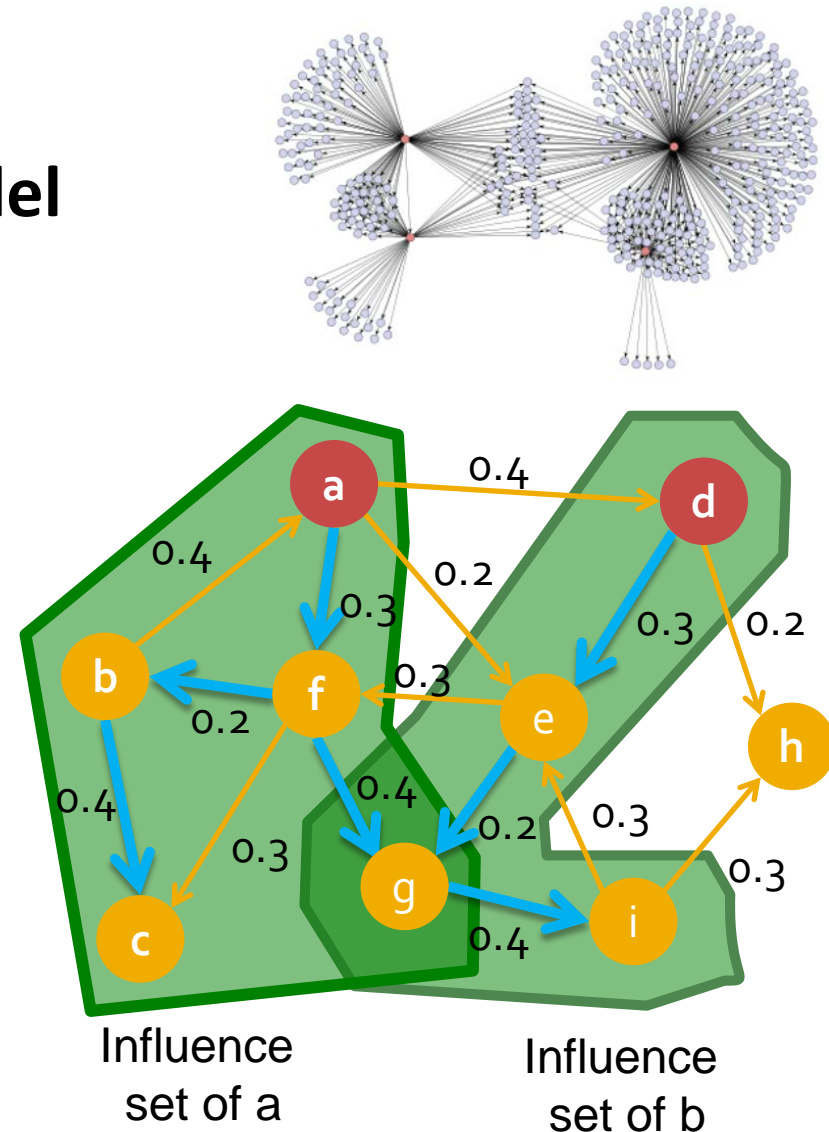
- Each node infects a neighbor with some probability

■ Algorithms:

■ Influence Maximization

- Set of k nodes producing **largest expected cascade size** if activated
- Submodularity
- Greedy hill-climbing

■ Outbreak Detection



Network Diffusion (3)

■ Observations:

■ Memetracking

- Blogs train mass media

■ Models:

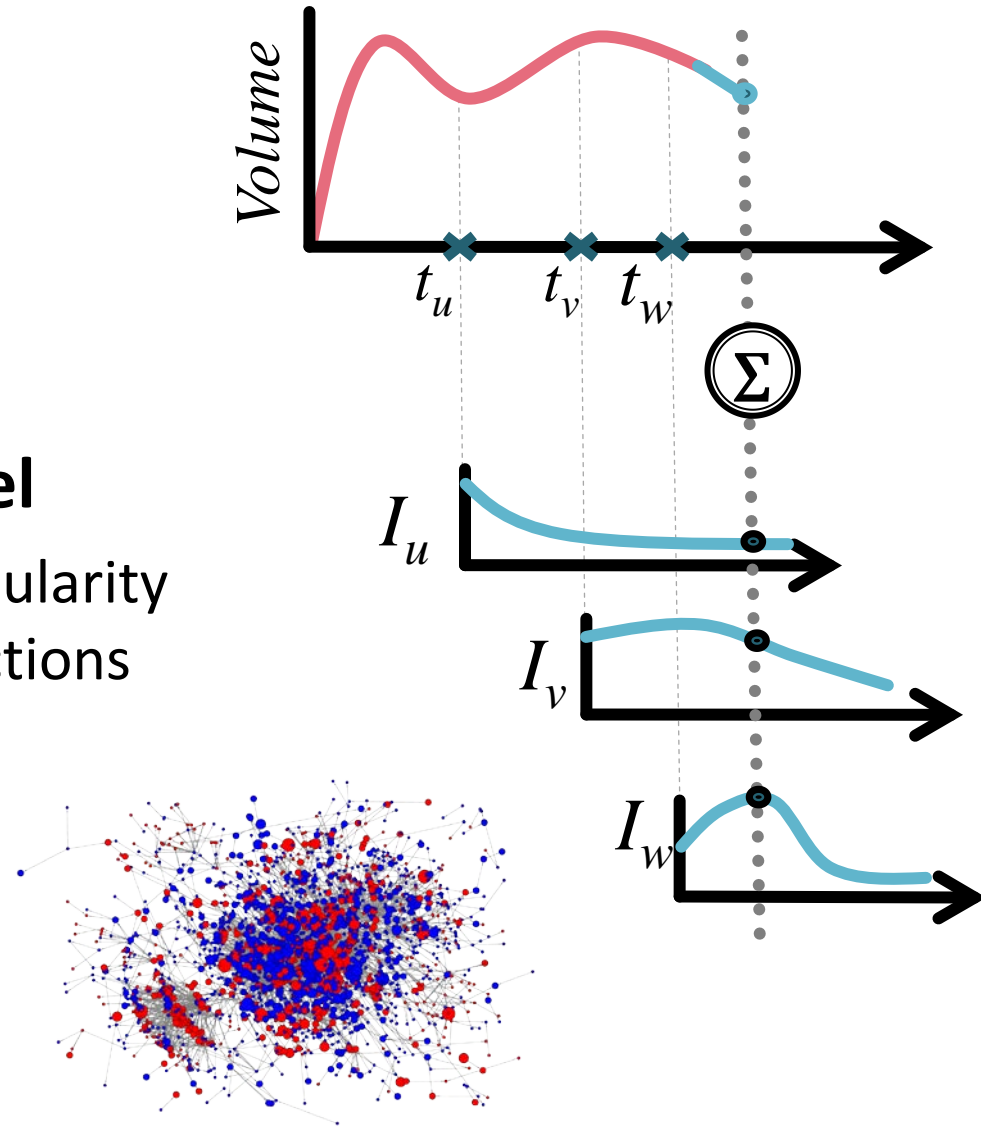
■ Linear influence model

- Predict information popularity based on influence functions

■ Algorithms:

■ Network Inference

- Given infection times
- Infer the network



Map of Superpowers

Observations

Small diameter,
Edge clustering

Scale-free

Strength of weak ties,
Core-periphery

Densification power law,
Shrinking diameters

Patterns of signed edge
creation

Viral Marketing, Blogosphere,
Memetracking

Models

Small-world model,
Erdős-Renyi model

Preferential attachment,
Copying model

Kronecker Graphs

Microscopic model of
evolving networks

Structural balance,
Theory of status

Independent cascade model,
Game theoretic model

Algorithms

Decentralized search

PageRank, Hubs and
authorities

Community detection:
Girvan-Newman, Modularity

Link prediction,
Supervised random walks

Models for predicting
edge signs

Influence maximization,
Outbreak detection, LIM

Applying Your Superpowers



Applying Your Superpowers

■ Link prediction

- Suggest friends in networks

■ Trust and distrust

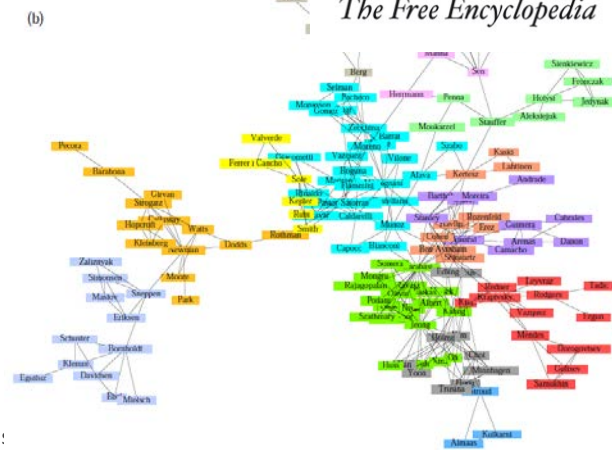
- Predict who are your friends/foes. Who you trust.

■ Community detection

- Find clusters and communities in social networks

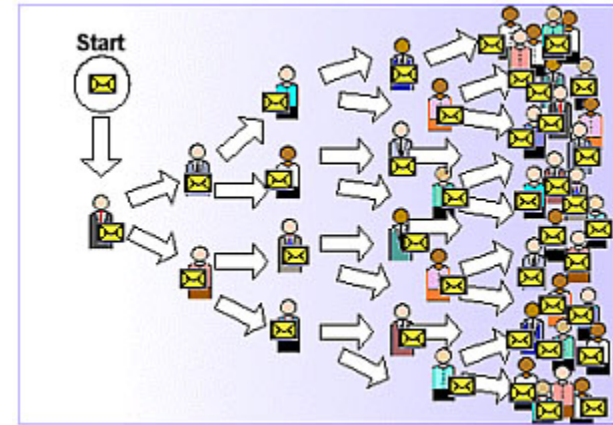


WIKIPEDIA
The Free Encyclopedia



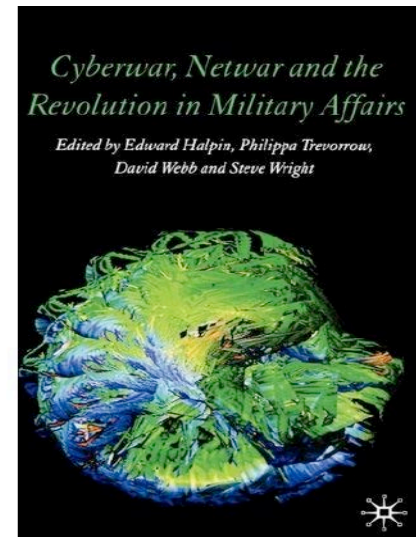
Applying Your Superpowers

- Marketing and advertising
 - Finding influencers
 - Tracing information flows
- Diffusion of information
 - How to trace information as it spreads
 - How to efficiently detect epidemics and information outbreaks



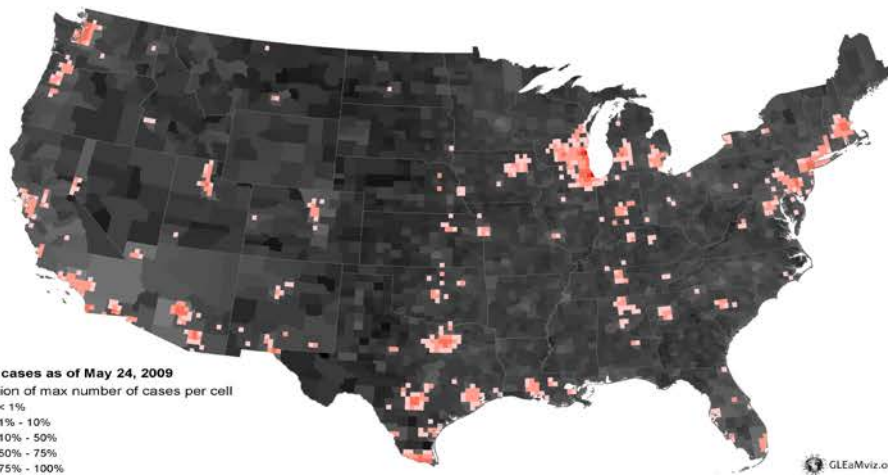
Applying Your Superpowers

- Intelligence and fighting (cyber) terrorism

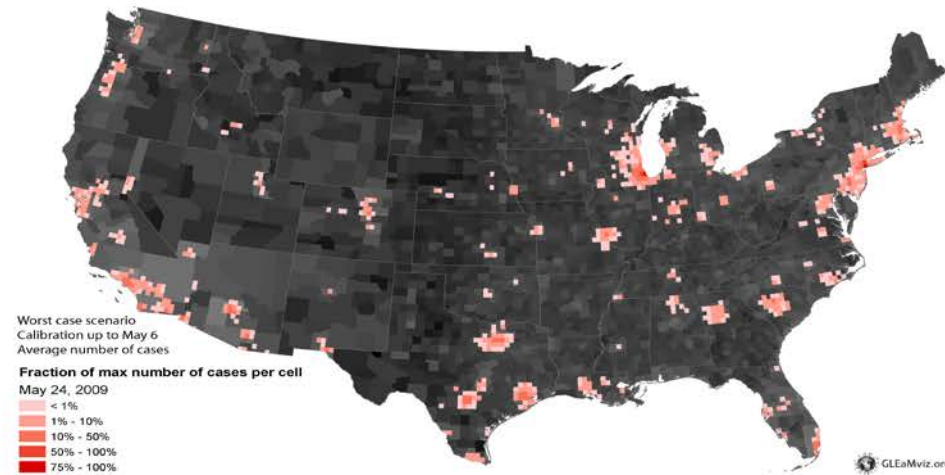


Applying Your Superpowers

■ Predicting epidemics



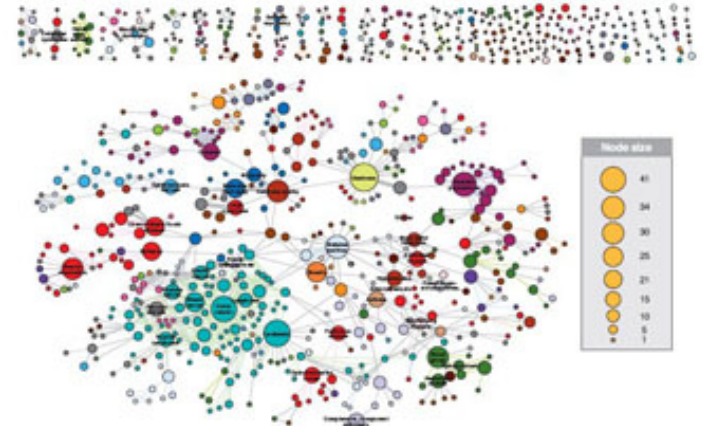
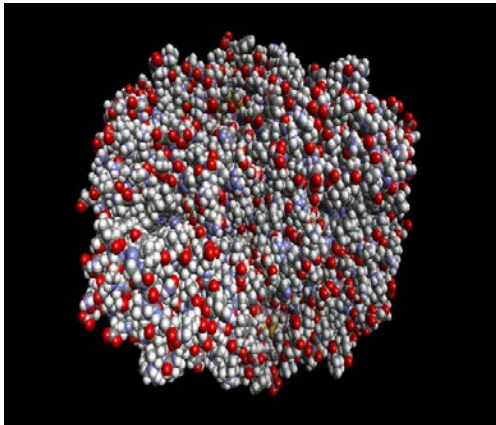
Real



Predicted

Applying Your Superpowers

- Interactions of human disease
- Drug design



Map of Superpowers

Observations

Small diameter,
Edge clustering

Scale-free

Strength of weak ties,
Core-periphery

Densification power
law,
Shrinking diameters

Patterns of signed edge
creation

Viral Marketing,
Blogsphere,
Memetracking

Models

Small-world model,
Erdős-Renyi model

Preferential
attachment, Copying
model

Kronecker Graphs

Microscopic model of
evolving networks

Structural balance,
Theory of status

Independent cascade
model, Game theoretic
model

Algorithms

Decentralized search

PageRank, Hubs and
authorities

Community detection:
Girvan-Newman,
Modularity

Link prediction,
Supervised random
walks

Models for predicting
edge signs

Influence maximization,
Outbreak detection,
LIM



Networks: BIG PICTURE

- **Availability of network data:**
 - Web & Social media: a “telescope” into humanity
- **Task:** find **patterns, rules, clusters, ...**
 - ... in large static and evolving graphs
 - ... in processes spreading over the networks
- **Goal:**
 - **Predict/anticipate future behaviors**
 - **Detect outliers**
 - **Design novel applications**

Networks: BIG PICTURE

- **Universal language for describing complex data**
 - **We are surrounded by hopelessly complex systems**
 - **Society** is a collection of six billion individuals
 - **Communication systems** link electronic devices
 - **Information** and **knowledge** is organized and linked
 - Networks from various domains of science, nature, and technology are more similar than expected
- **Shared vocabulary between fields**
 - Computer Science, Social science, Physics, Economics, Statistics, Biology

What Next?

- **Project writeups**

- Due Sunday Dec 11 at 11:59 pacific time

- **Poster session**

- Friday Dec 16 from 12:15 - 3:15 in Packard Atrium
- All groups which have at least one non-SCPD member are expected to present
- One group member should be at the poster at all times, **but** the goal of this is to give you a chance to see what your classmates have been working on, so make sure to explore
- There will be snacks!

What Next? Seminars

■ Seminars:

- RAIN Seminar: <http://rain.stanford.edu>
- InfoSeminar: <http://i.stanford.edu/infoseminar>

■ Conferences:

- **WWW**: ACM World Wide Web Conference
- **WSDM**: ACM Web search and Data Mining
- **ICWSM**: AAAI Int. Conference on Web-blogs and Social Media
- **KDD**: ACM Conference on Knowledge Discovery and Data Mining

What Next? Courses

- **CS246: Mining Massive Datasets (Winter 2012)**
 - Data Mining & Machine Learning for big data
 - (big=does' fit in memory/single machine)
 - MapReduce, Hadoop and similar
- **CS341: Project in Data Mining (Spring 2012)**
 - Do a research project on big data
 - Groups of 3 students
 - We provide interesting data, projects and unlimited access to the Amazon computing infrastructure
 - Nice way to finish up your class project & publish it!

What Next? Courses

- **Other relevant courses**
 - **CS276**: Information Retrieval and Web Search
 - **CS229**: Machine Learning
 - **CS245**: Database System Principles
 - **CS347**: Transaction Processing and Distributed Databases
 - **CS448g**: Interactive Data Analysis

In Closing

- You Have Done a Lot!!!
- And (hopefully) learned a lot!!!
 - Answered questions and proved many interesting results
 - Implemented a number of methods
 - And did excellently on the final project!

**Thank You for the
Hard Work!!!**

