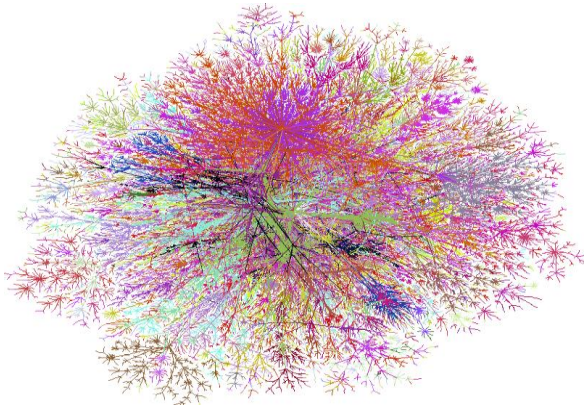
Information Networks

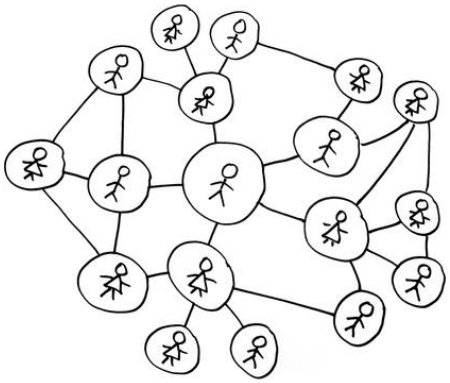
Review of Key Concepts

Thanks to Jure Leskovec, Stanford and Panayiotis Tsaparas, Univ. of Ioannina for slides

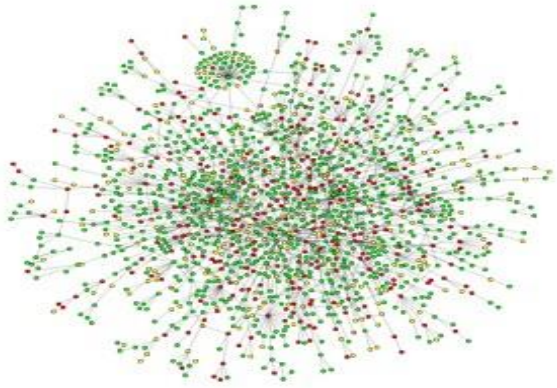
The "Age of Networks"



Technological



Social



Biological

why should we care about networks?

Why Networks? Why Now?

- **Universal language for describing complex data**
  - Networks from science, nature, and technology are more similar than one would expect
- **Shared vocabulary between fields**
  - Computer Science, Social science, Physics, Economics, Statistics, Biology
- **Data availability (/computational challenges)**
  - Web/mobile, bio, health, and medical
- **Impact!**
  - Social networking, Social media, Brain, Drug design
  - **We will never understand these systems unless we understand the networks behind them!**

how do we reason about networks?

Reasoning About Networks

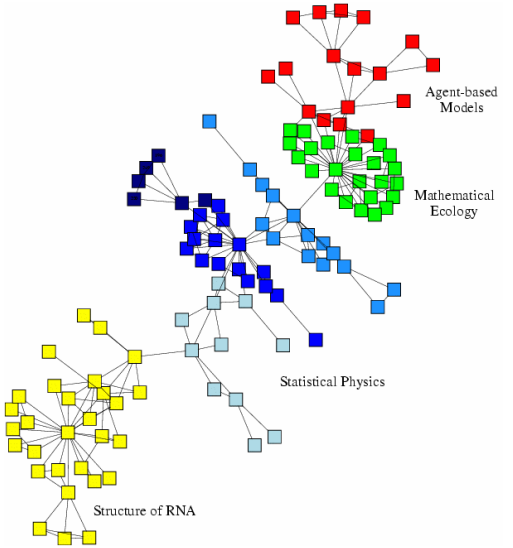
How do we reason about networks?

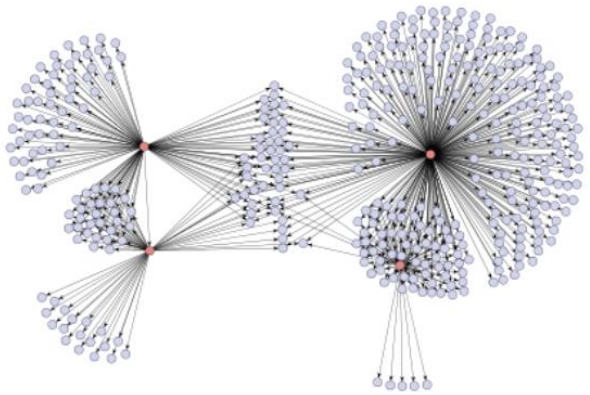
- **Empirical**: Study network data to find organizational principles
- **Mathematical models**: Probabilistic, graph theory
- **Algorithms**: Methods for analyzing graphs

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?

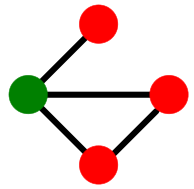
- **basic graph theory**
  - graphs, networks
  - bow-tie structure
- **network measurements**
  - degree distributions, power-laws
  - shortest paths, clustering coefficient
- **network models**
  - Erdos-Renyi model
  - small-world model
  - configuration model
  - scale-free networks
- **models of evolving graphs**
  - preferrential attachment model
  - microscopic/macroscopic evolution of networks
  - forest-fire model
- **community structure in networks**
  - Strength of weak ties, structural holes
  - community detection, Girvan-Newman algorithm
  - graph partitioning, graph cuts, conductance
  - spectral graph theory, spectral graph clustering
- **overlapping communities in networks**
  - cliques, clique percolation method
  - community-affiliation graph model
- **link analysis**
  - web search
  - hubs and authorities (HITS)
  - PageRank, topic-sensitive PageRank
- **link prediction**
  - neighborhood-based methods
  - node proximity based methods, supervised learning models, Facebook's "PYMK", Twitter's "WtF"
- **cascading behavior in networks**
  - Granovetter's model, threshold model
  - game theoretic model
  - epidemic model on trees
  - disease spreading models (SIR, SIS, SIRS)
  - independent cascade model
  - influence maximization
  - outbreak detection
- **recommender systems**
  - content-based
  - collaborative filtering based
  - latent factor models
  - the Netflix challenge

How It All Fits Together

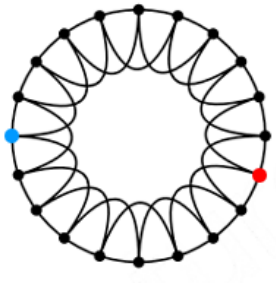
Properties	Models	Algorithms
Small diameter, Edge clustering	Small-world model, Erdös-Renyi model	Decentralized search
Scale-free	Preferential attachment, Copying model	PageRank, Hubs and authorities
Strength of weak ties, Core-periphery	Community-affiliation Graph Model	Community detection: Girvan-Newman, Modularity
Densification power law, Shrinking diameters	Microscopic model of evolving networks	Link prediction, Supervised random walks
Information virality, reproductive number	Independent cascade model, Game theoretic model, SIR	Influence maximization, Outbreak detection, LIM

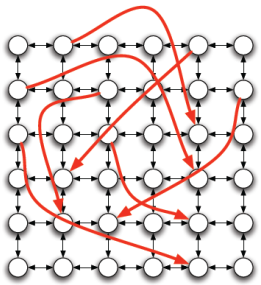
Small-World Phenomena

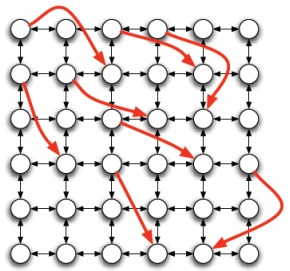
- **Properties:**
  - **Six degrees of separation**
    - Networks have small diameters
  - **Edges in the networks cluster**
    - Large clustering coefficient
- **Models:**
  - **Erdös-Renyi model**
    - Baseline model for networks
  - **The Small-World model**
    - Small diameter and clustered edges
- **Algorithms:**
  - **Decentralized search in networks**
    - Kleinberg's model and algorithm



$C_i=1/3$



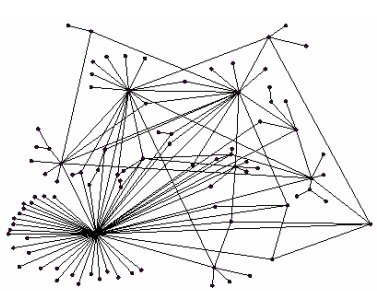


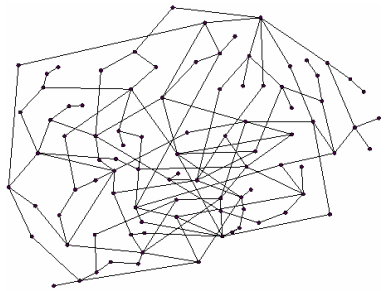


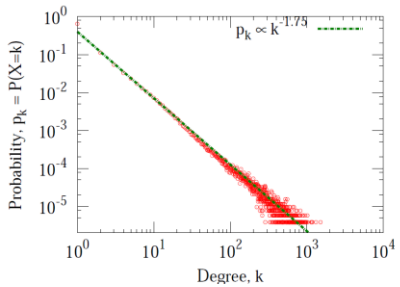
$$P\left(u \to v\right) \sim d(u,v)^{-\alpha}$$

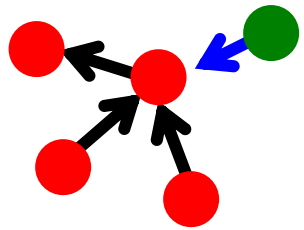
Scale-Free Networks

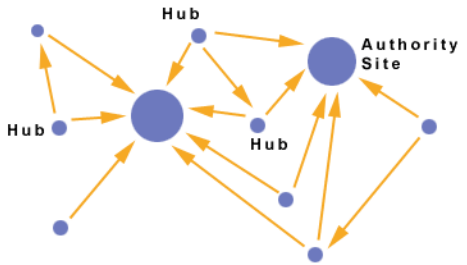
- **Properties:**
  - **Power-law degrees**
    - Degrees are heavily skewed
  - **Network resilience**
    - Networks are resilient to random attacks
- **Models:**
  - **Preferential attachment**
    - Rich get richer
- **Algorithms:**
  - **Hubs and Authorities**
    - Recursive: $a_i = \sum_{j \to i} h_j$, $h_i = \sum_{i \to j} a_j$
  - **PageRank**
    - Recursive formulation, Random jumps





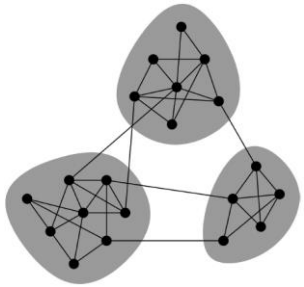


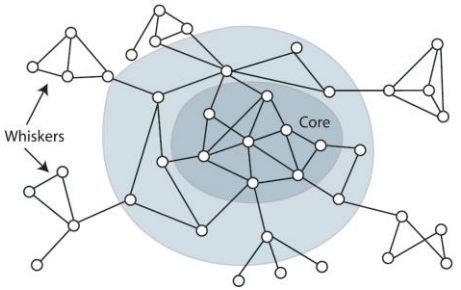




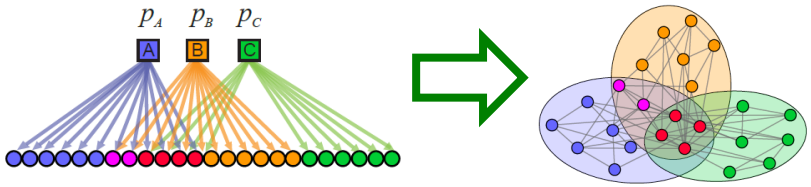
Community Detection

- **Properties:**
  - **Strength of weak ties**
  - **Core-periphery structure**

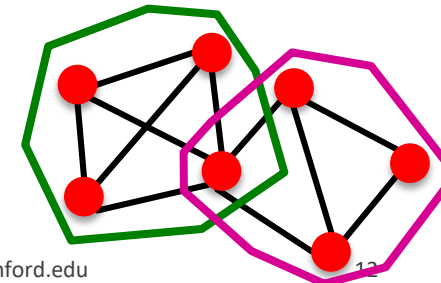




- **Models:**
  - **Community-affinity model**

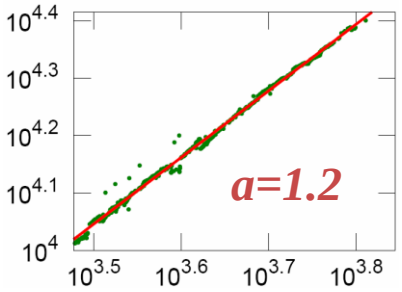


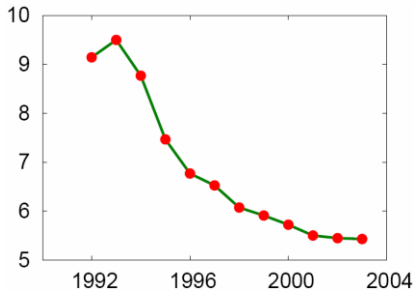
- **Algorithms:**
  - **Spectral Clustering**
  - **Girvan-Newman** (Betweeness centrality)
  - **Modularity:** #edges within group – E[#edges within group]
  - **Clique Percolation Method**
    - Overlapping communities



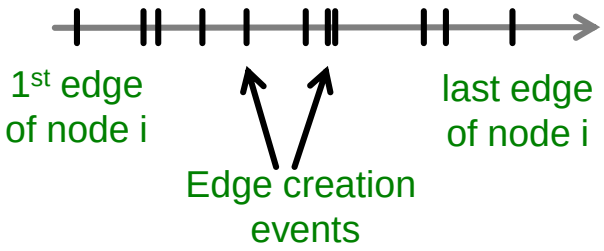
Network Evolution

- **Properties:**
  - **Densification Power Law**
    - $E(t) \propto N(t)^a$
  - **Shrinking Diameter**

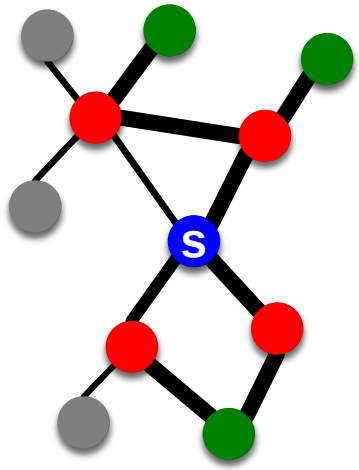




- **Models:**
  - **Microscopic Network Evolution**
    - Exponential life-times, Evolving sleeping times
    - Random-Random edge attachment

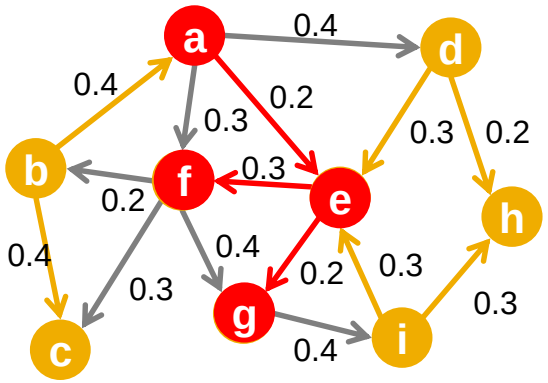


- **Algorithms:**
  - **Link prediction**



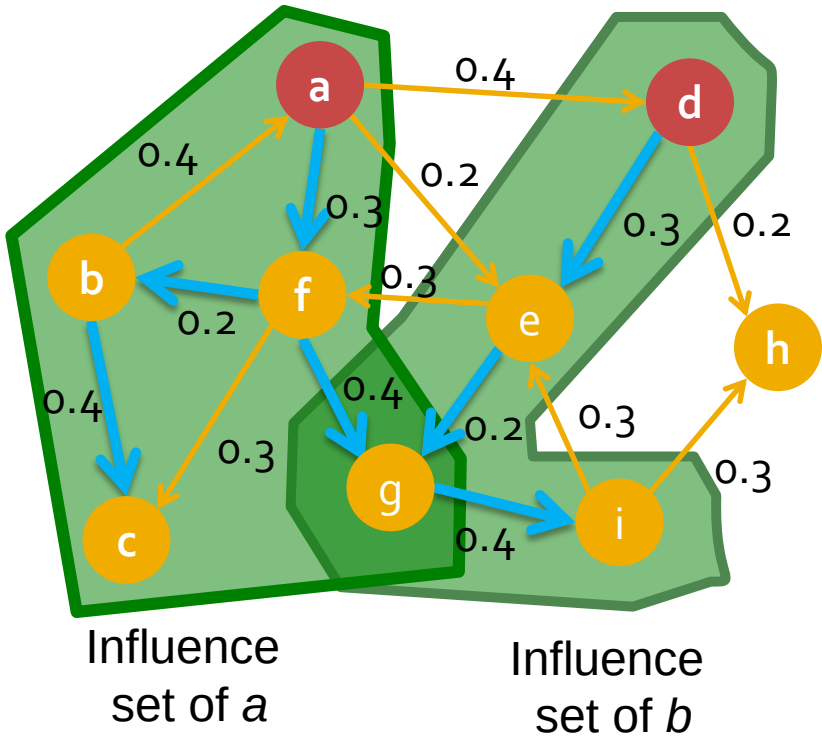
Network Diffusion (1)

- **Properties:**
  - **Node-to-node influence**
  - **Node threshold**
  - **Cascade spread**
- **Models:**
  - **Game theoretic model:**
    - Payoffs, Competing products
  - **Independent Cascade Model**
    - Each node infects a neighbor with some probability



Network Diffusion (2)

- **Algorithms:**
  - **Influence Maximization**
    - Set *of k* nodes producing **largest expected cascade size** if activated
    - Submodularity
    - Greedy hill-climbing
  - **Outbreak Detection**



Map of Superpowers

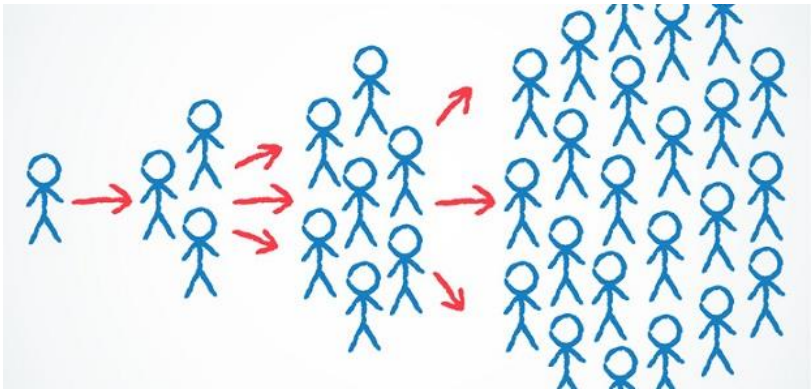
Properties	Models	Algorithms
Small diameter, Edge clustering	Small-world model, Erdös-Renyi model	Decentralized search
Scale-free	Preferential attachment, Copying model	PageRank, Hubs and authorities
Strength of weak ties, Core-periphery	Community-affiliation Graph Model	Community detection: Girvan-Newman, Modularity
Densification power law, Shrinking diameters	Microscopic model of evolving networks	Link prediction, Supervised random walks
Information virality, reproductive number	Independent cascade model, Game theoretic model, SIR	Influence maximization, Outbreak detection, LIM

Applying Your Superpowers

Applying Your Superpowers

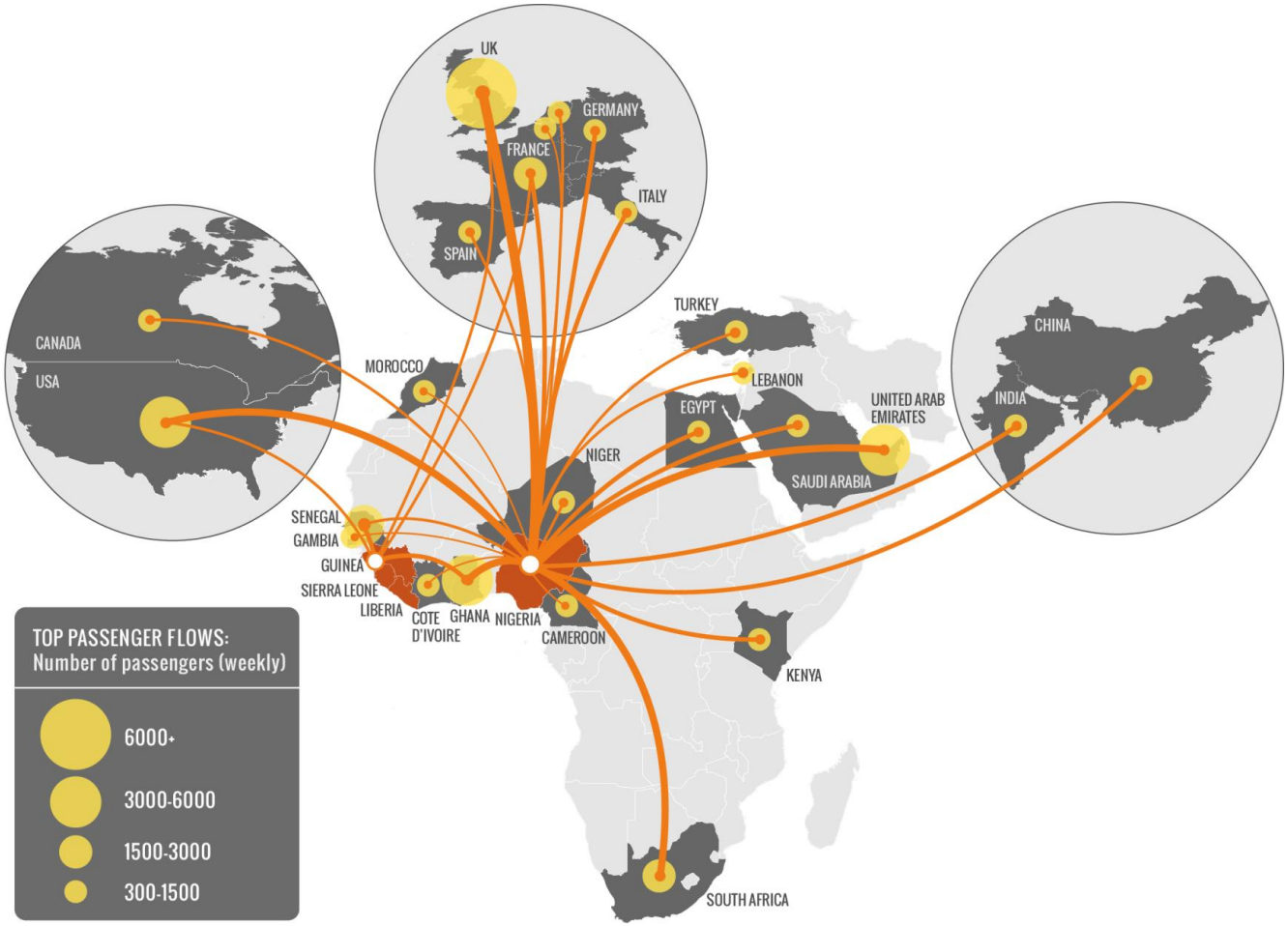
- **Social media analytics**

- **Viral marketing**



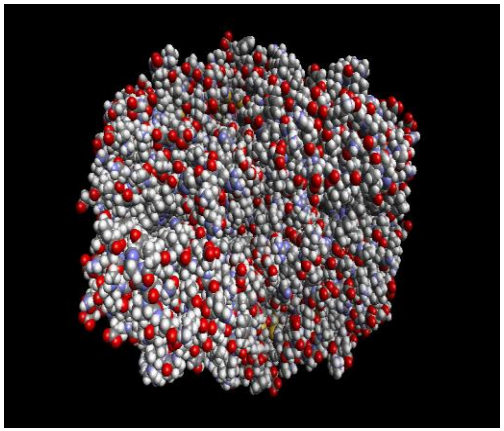
Applying Your Superpowers

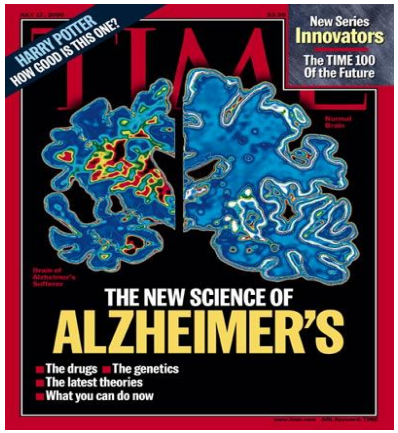
- **Predicting epidemics: Ebola**



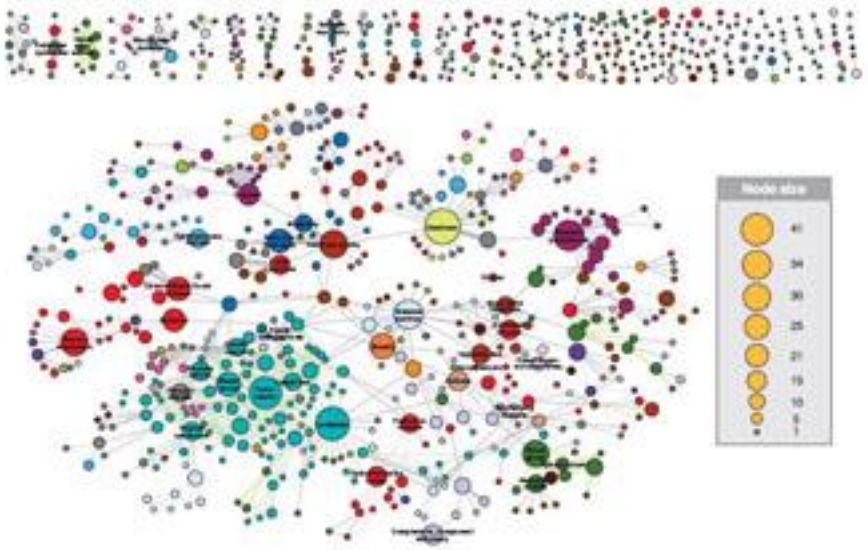
Applying Your Superpowers

- **Interactions of human diseases**





- **Drug design**



What's Next?

What's Next?

- **Final exam**
  - **Thu, Apr 20th, 10am-12pm**
    - Short answers
    - Room **SC303** (Stong College, same building, different room)
- **Project presentation**
  - **Tue, Apr 4th, in-class**
    - 25 minutes + 5 min QA
    - See course website for more info
- **Project final report**
  - **Sun, Apr 30th Midnight (11:59PM) Pacific Time**
    - Email PDF report, 7-8 pages
    - see course website for more info

What Next? Seminars

- **EECS6xxx: Data Analytics and Visualization**
  - Fall 2017, Project course
  - Data mining, graph mining, data visualization
- **Conferences / Journals:**
  - **Conferences**
    - **KDD**: Conf. on Knowledge Discovery & Data Mining
    - **WWW**: ACM World Wide Web Conference
    - **WSDM**: ACM Web search and Data Mining
    - **ICDM:** IEEE International Conference on Data Mining
    - **ICWSM**: AAAI Int. Conf. on Web-blogs & Social Media
  - **Journals**
    - **Complex Networks:** Journal of Complex Networks
    - **TKDD:** ACM Transactions on Knowledge Discovery from Data
    - **TKDE:** IEEE Transactions on Knowledge and Data Engineering

In Closing…

You have worked a lot…

…and (hopefully) learned a lot!

thank you & happy holidays