Knowledge Discovery in Social and Information Networks

Cornelia Caragea

Computer Science and Engineering
University of North Texas

Acknowledgement: Lada Adamic and Jure Leskovec

June 9, 2016

Overview of Social and Information Network Analysis

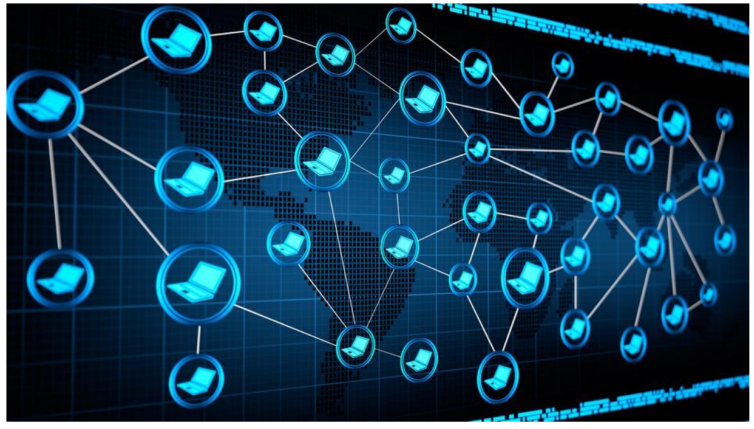
Recently...

- There has been a growing public fascination with the complex “connectedness" of modern society

Internet

Connecting us to the world!

... to millions of users!



... and to information!

Bogotá

From Wikipedia, the free encyclopedia

Coordinates: 4°35′53″N 74°4′33″W

For a suburban borough located in New Jersey, see Bogota, New Jersey.

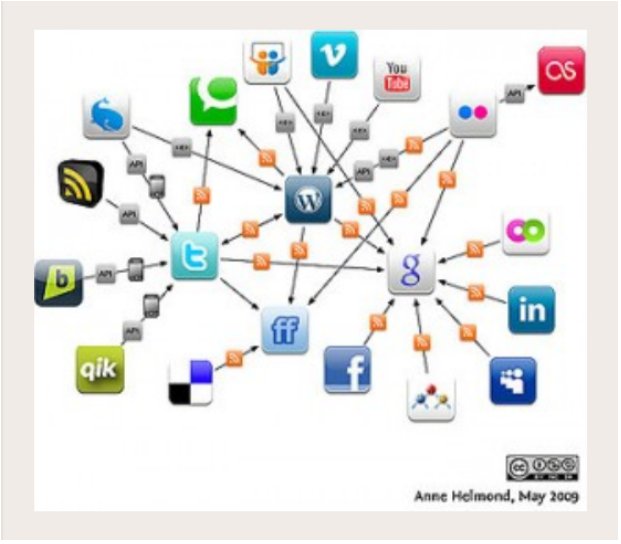
This article **needs additional citations for verification**. Please help improve this article by adding citations to reliable sources. Unsourced material may be challenged and removed. *(December 2011)*

Bogotá (Spanish pronunciation: [boɣoˈta] (listen)), from 1991 to 2000 called **Santafé de Bogotá**, is the capital, and largest city, of Colombia. It is also designated by the national constitution as the capital of the department of Cundinamarca, though the city of Bogotá now comprises an independent capital district and no longer belongs administratively to that department. Bogotá is the most populous city in the country, with 7,363,782 inhabitants as of 2010.[4] Bogotá and its metropolitan area, which includes municipalities such as Chía, Cota, Soacha, Cajicá and La Calera, had a population of around 8 million in 2010.[5]

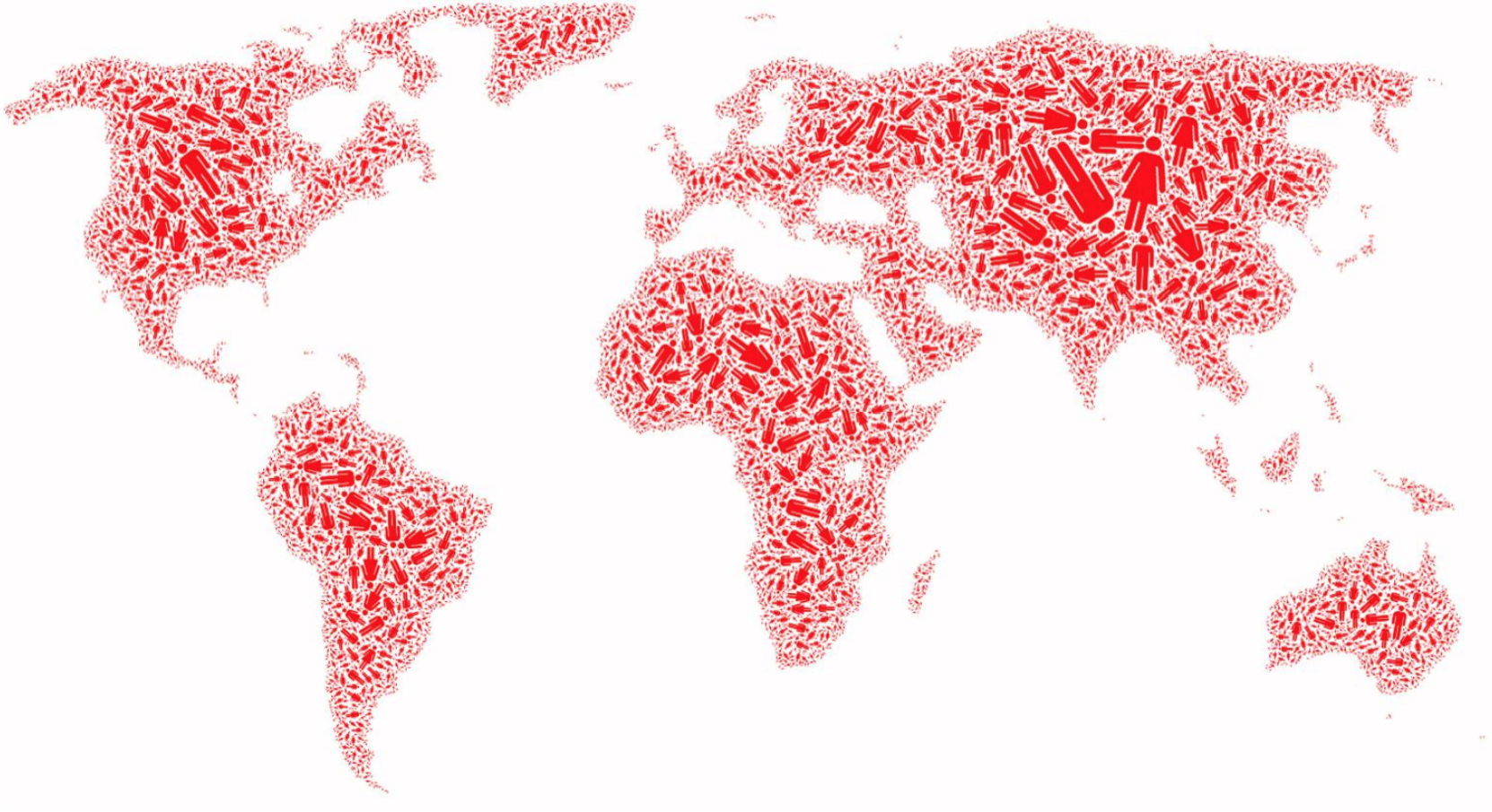
Bogotá

— Capital District —

We are Surrounded by Friends!



We are Surrounded by the Society as a Whole!



Networks!

At the heart of this fascination is the idea of a network

- A pattern of interconnections among a set of objects, where the interconnections can be of various types
- The diversity of contexts in which networks are invoked is vast:
  - Social systems
  - Information systems
  - Technological systems
  - Economic systems
  - $\cdots$ and many more!

Examples of Networks

- Global manufacturing operations now have networks of suppliers
- Web sites have networks of users
- Media companies have networks of advertisers

Examples of Networks

- Global manufacturing operations now have networks of suppliers
- Web sites have networks of users
- Media companies have networks of advertisers

We will never understand these systems unless we understand the networks behind it.

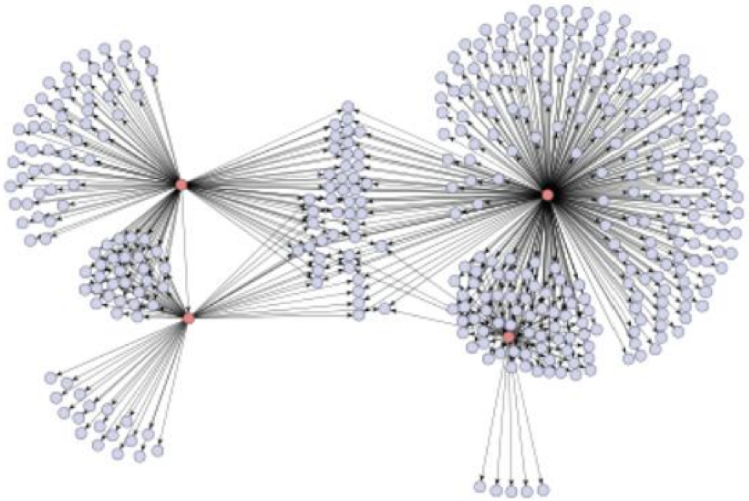
Reasoning about Networks

- What can we get out of modeling the world around us as networks?
  - Understand:
    - How information diffuses in social networks.
    - How resilient different infrastructure networks are (e.g., an electrical power network).
    - Why networks are organized the way they are (i.e., predict behavior of networked systems).
    - Random or intentional failures / Conflicts.
  - Identify:
    - Patterns and statistical properties of network data.

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?
- More precisely, we learn to characterize a network structure:
  - Are nodes connected through the network?
  - How far apart are they?
  - Are some nodes more important than others due to their position in the network?
  - Is the network composed of communities?



What Do We Study in Networks?

Model Network Formation:

- Understand how the network structure is formed:
  - Randomly generated networks - adding edges at random
  - Preferential attachment - “the rich gets richer"
  - Small-world networks - friend of a friend relationship
    - Any two people in the world are connected through a small number of nodes
    - A Facebook study showed that any two people in the Facebook graph are connected within 4.7 hops, on average
  - Optimization, strategic network formation
    - Airline networks optimized to be efficient (it does not cost the airline much money)
    - Individuals may choose to connect to some nodes, but not with others

What Do We Study in Networks?

Processes and dynamics:

- Networks provide “skeleton" for spreading of information, behavior, diseases
- How do information and diseases spread?
- Understand how network structure affects processes:
  - Information diffusion
    - If any two people in the network are connected through a small number of nodes, the information will diffuse very rapidly.
    - Virus diffusion - how the SN influences how quickly a virus can be spread and what immunization processes to use?
  - Opinion formation
    - Your opinion may be influenced by what your friends think.
  - Coordination/cooperation
    - Performing a certain task depends on the inputs from the nodes you are tied to, how quickly can you establish the task?
  - Resilience to attacks
    - If a subset of the nodes are removed from the SN, can the network still function?

Why Networks?

Why is the role of networks expanding?

- Data availability
  - Rise of Mobile, Web 2.0 and Social media
- Universality
  - 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
- Impact!
  - Social networking, Social media, Drug design

Networks: Size Matters

- Network data: Orders of magnitude
  - 436-node network of email exchange at a corporate research lab [Adamic-Adar, SocNets '03]
  - 43,553-node network of email exchange at an university [Kossinets-Watts, Science '06]
  - 4.4-million-node network of declared friendships on a blogging community [Liben-Nowell et al., PNAS '05]
  - 240-million-node network of communication on Microsoft Messenger [Leskovec-Horvitz, WWW '08]
  - 800-million-node Facebook network [Backstrom et al. '11]
- Aspects of Networks!

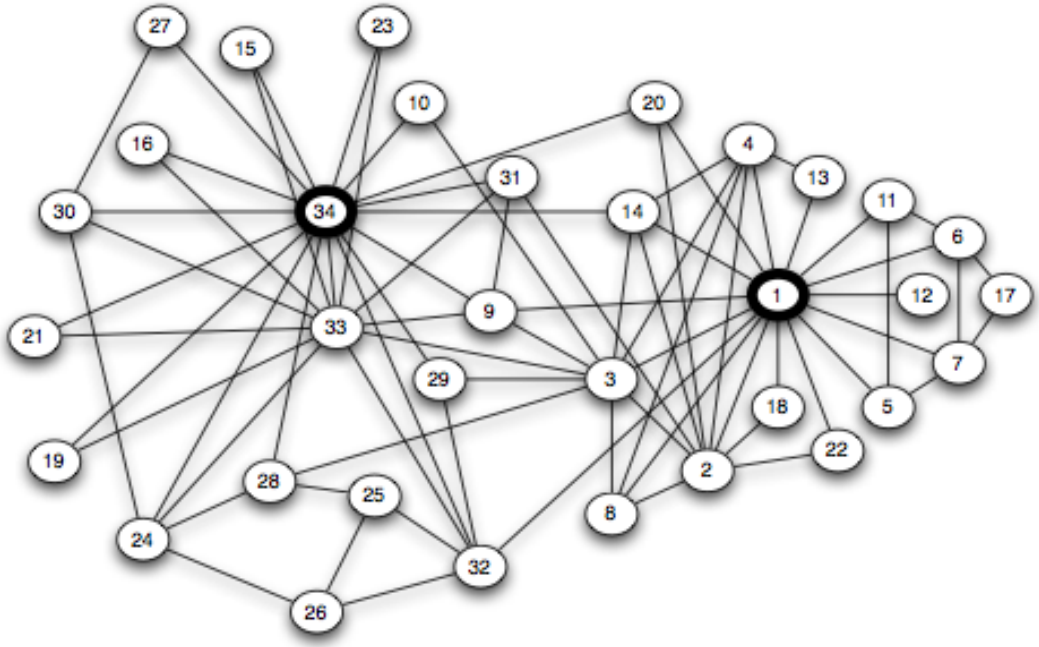
Aspects of Networks - Social

The Facebook Social Graph

- 4-degrees of separation [Backstrom-Boldi-Rosa-Ugander-Vigna, 2011].

Aspects of Networks - Friendship

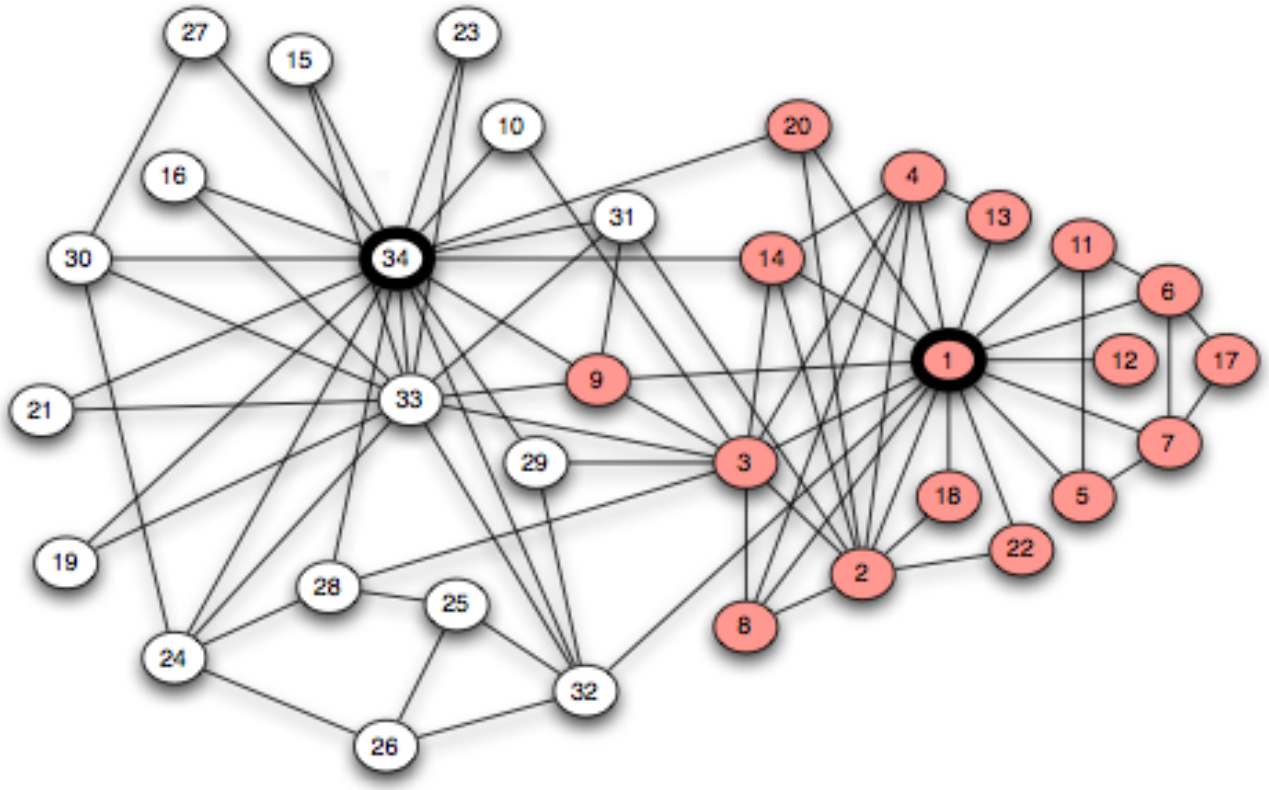
The Karate Club Network



- A social network among 34 people in a university karate club studied by the anthropologist Wayne Zachary in the 1970s.
- The people are represented by small circles, with lines joining pairs of people who are friends outside the context of the club.

Aspects of Networks - Friendship

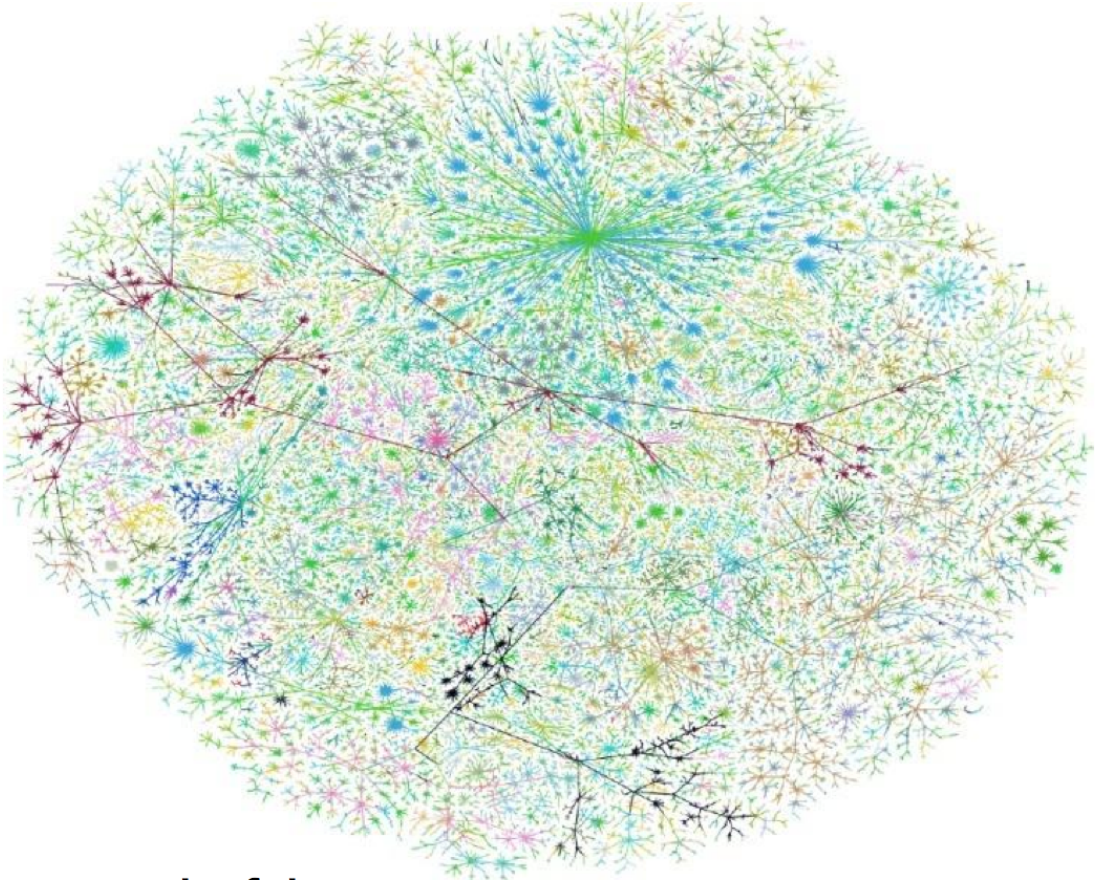
The Karate Club Network



- Networks can capture the sources of conflict within a group.

Aspects of Networks - Communication

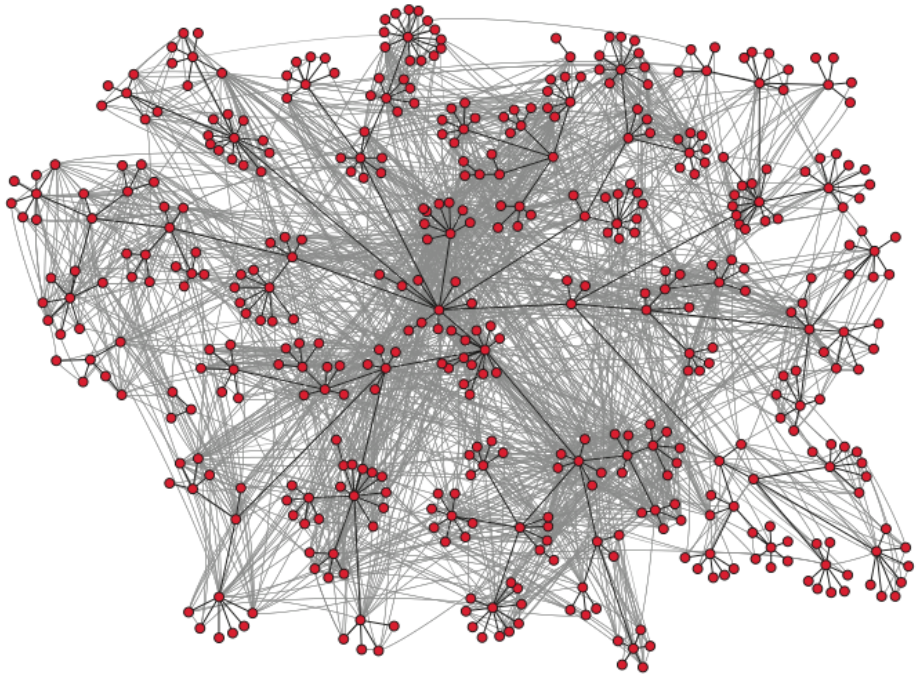
The Graph of the Internet



- Robustness [Doyle-Willinger, 2005].

Aspects of Networks - Communication

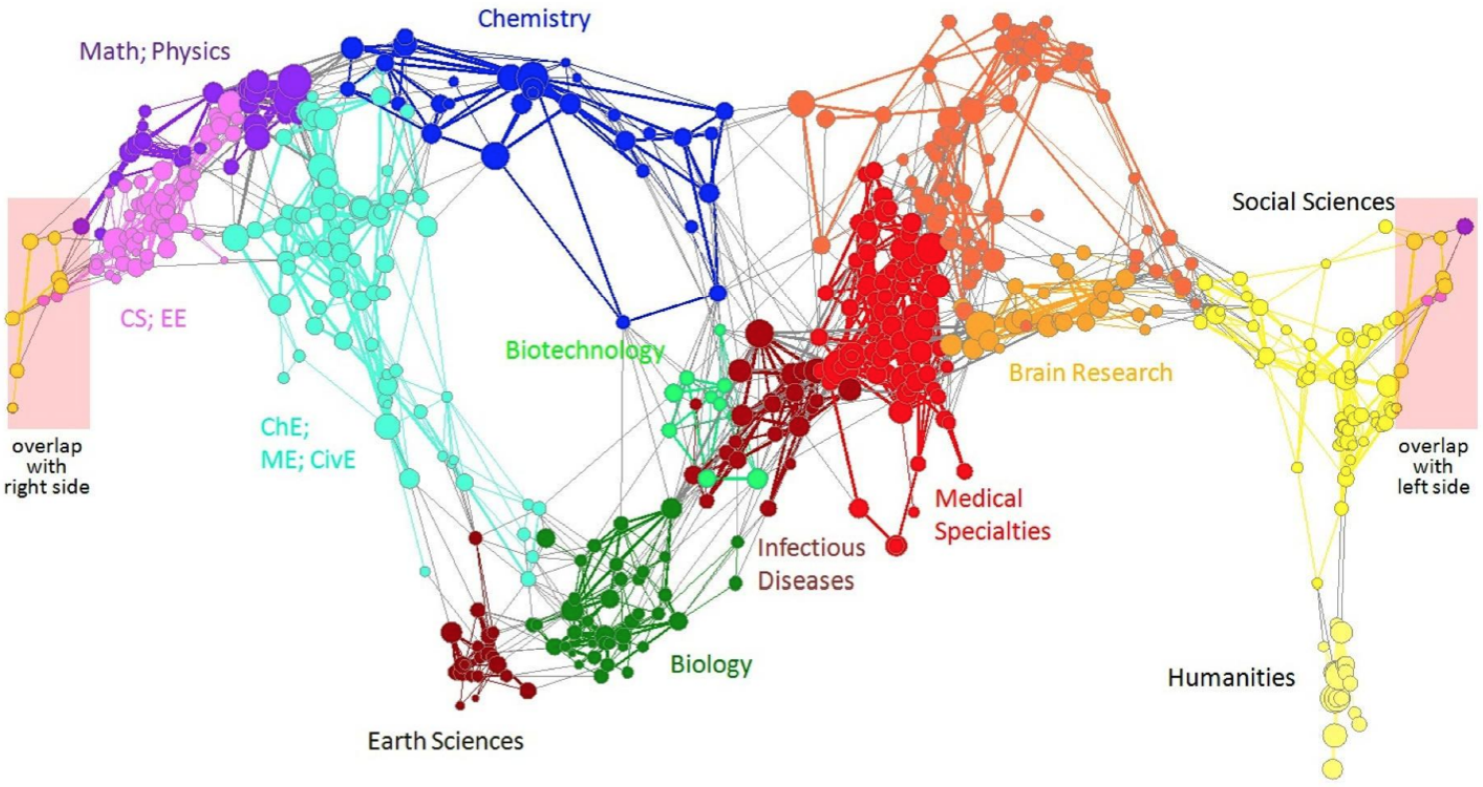
HP Research Labs Email Exchange [Adamic-Adar, 2003].



- Gray edge - at least two email exchange over a period of a few months, black edges - who reports to whom.
- Individual who are closer to each other are more likely to exchange emails, however, any two individuals are connected through a small number of nodes.

Aspects of Networks - Information

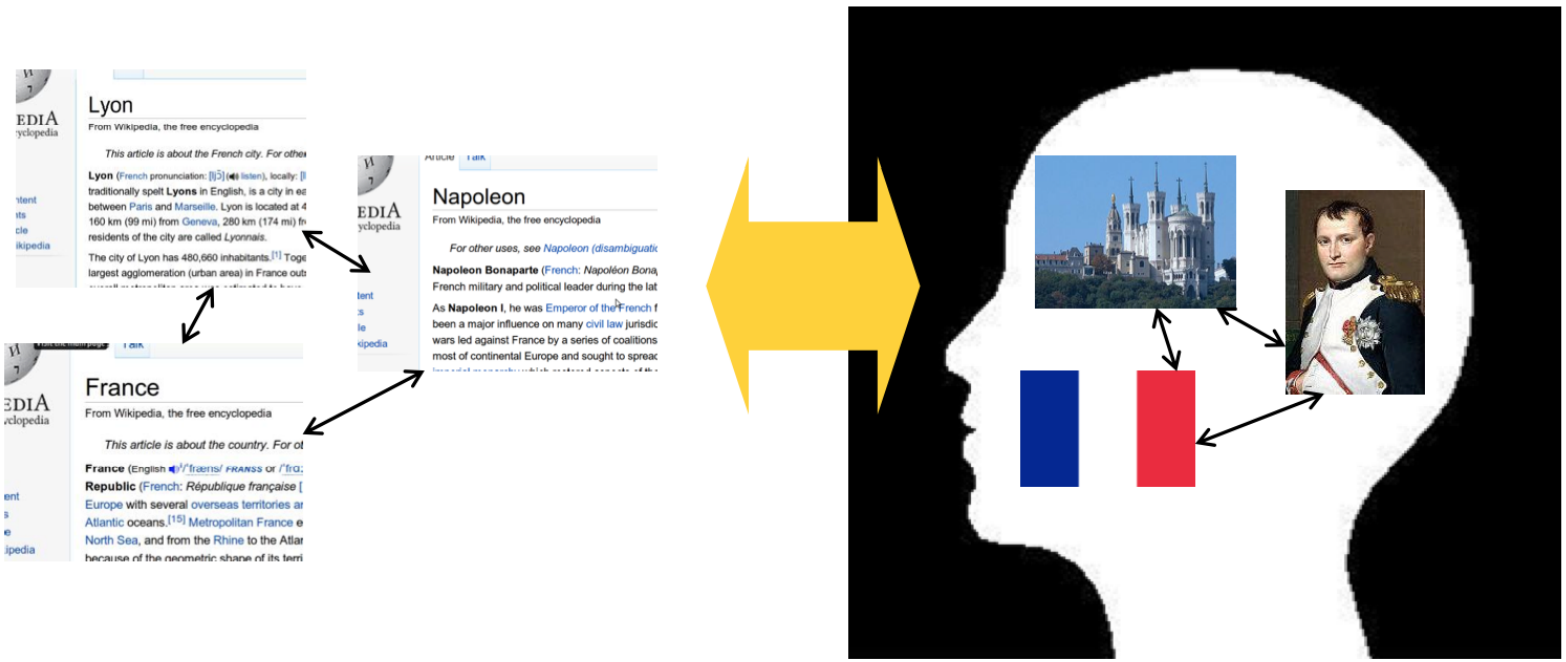
Citation Networks



- Maps of Science [Börner et al., 2012].

Aspects of Networks - Knowledge

The Graph of Wikipedia Pages



- Understand how people navigate Wikipedia → Understand how people connect concepts [West-Leskovec, 2012].

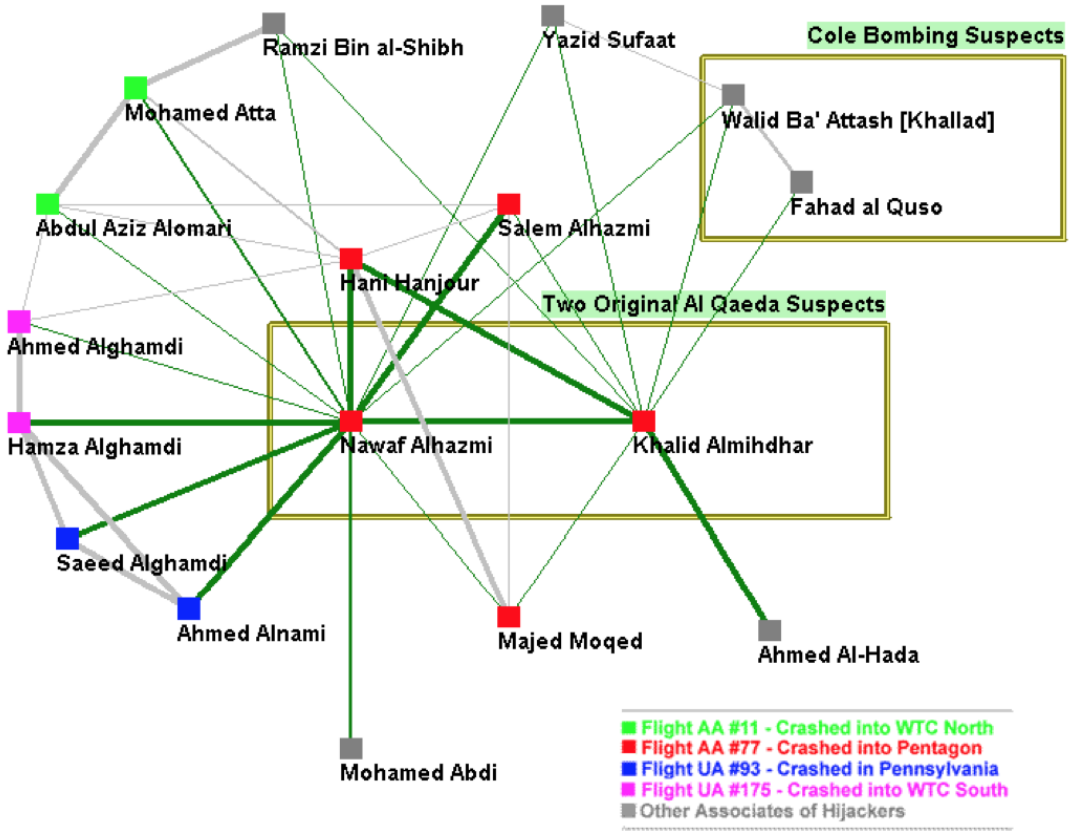
Aspects of Networks - Brain

The Human Neurons' Network

- Human brain has between 10-100 billion neurons [Sporns, 2011].

Aspects of Networks - Organizations

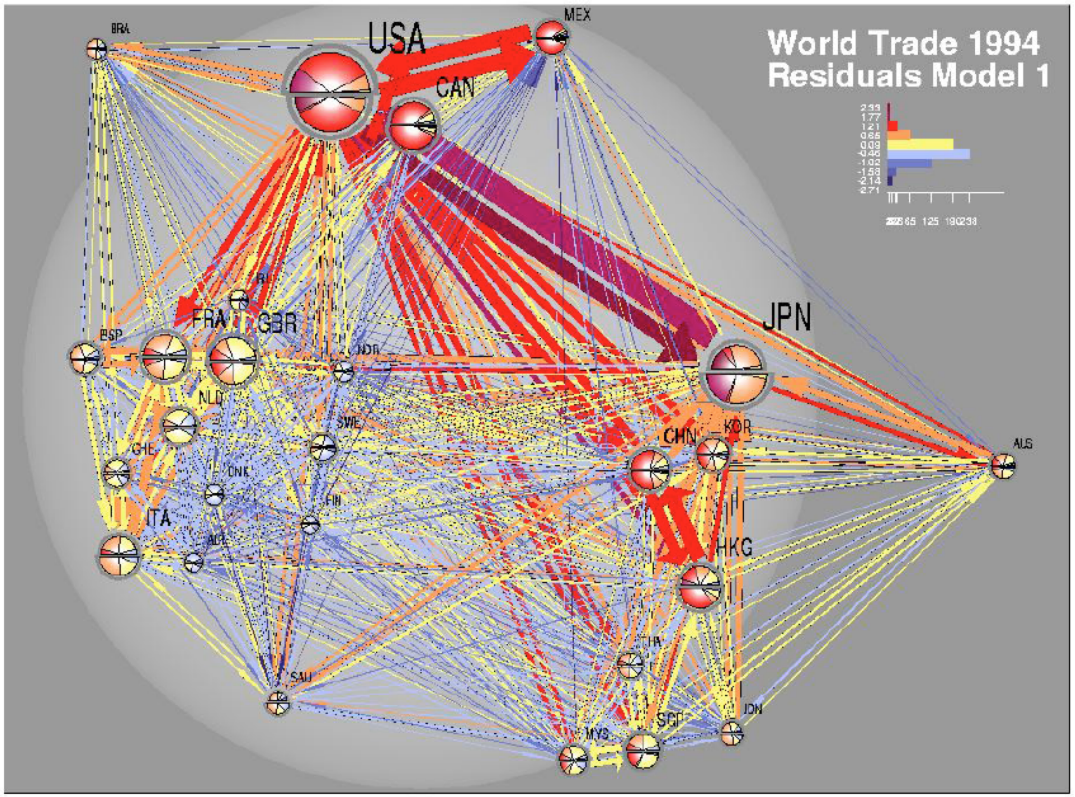
The 9/11 Terrorist Network [Krebs, 2002]



- Understand who are the leaders of a terrorist attack

Aspects of Networks - Economy

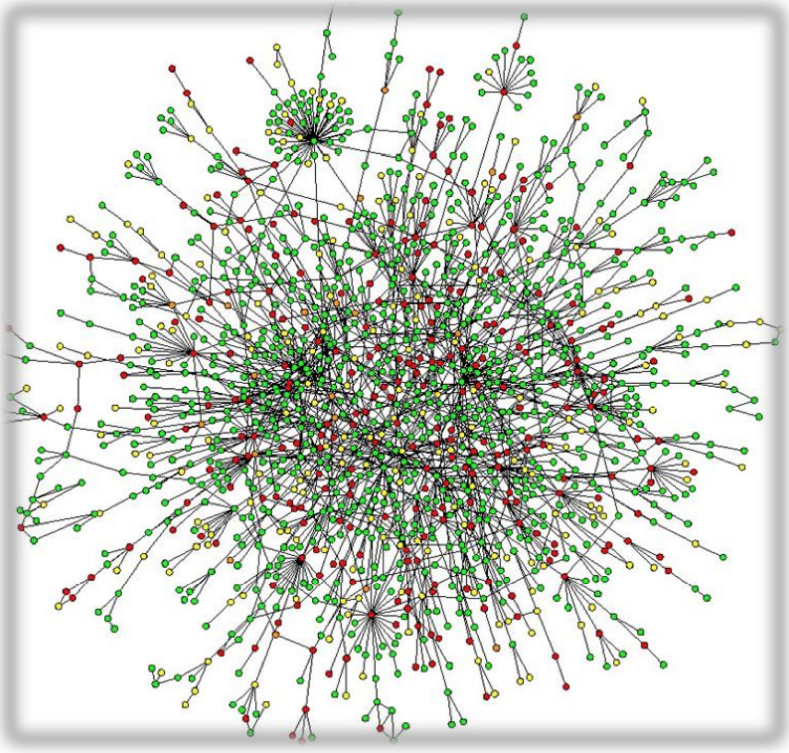
The International Trade Network [Krempel and Plümper, 2002]

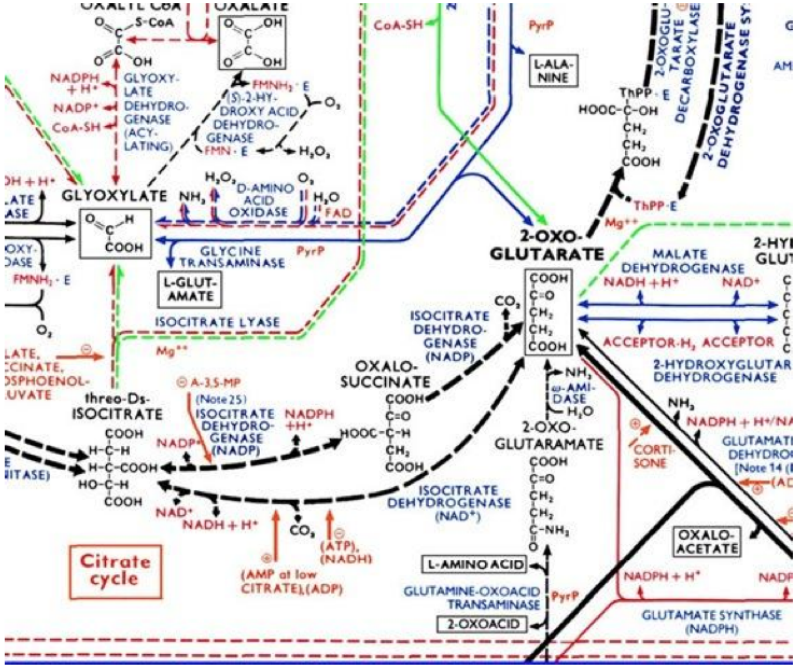


- Understand what countries occupy powerful positions and derive economic benefits from these positions

Aspects of Networks - Biology

The Protein and Metabolic Networks

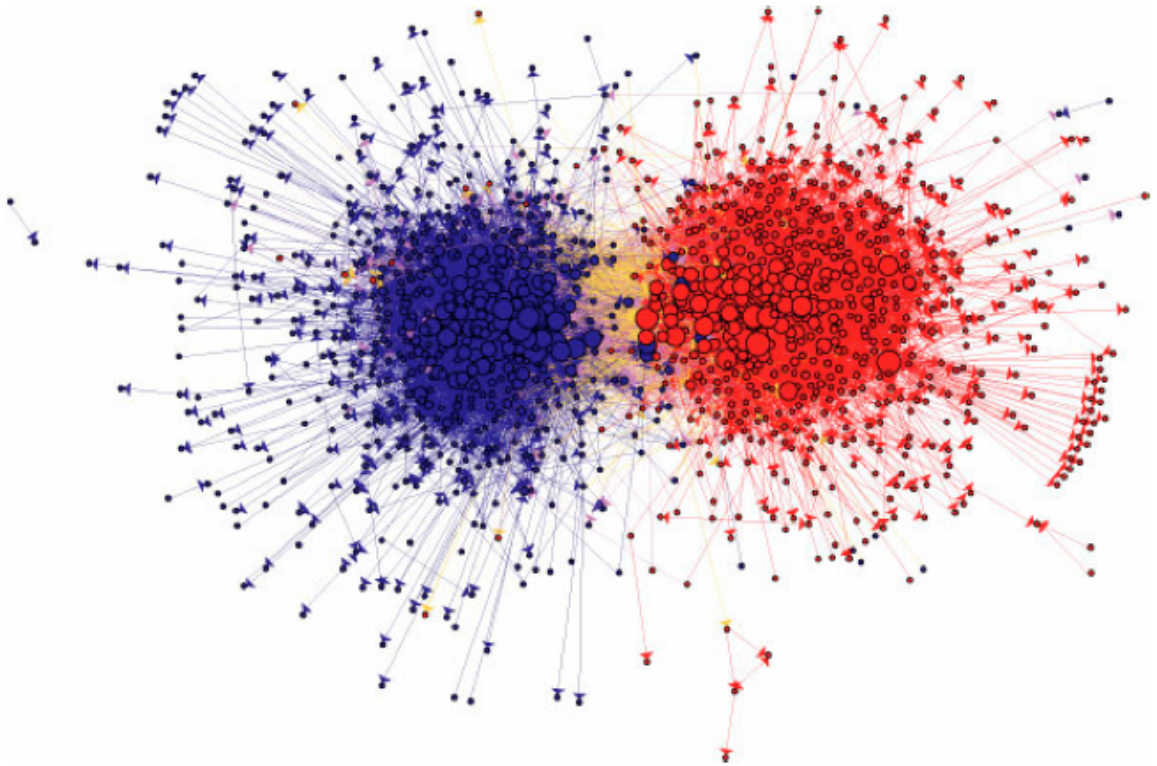




- Proteins and their “physical" interactions
- Metabolites and their chemical reactions

Aspects of Networks - Media

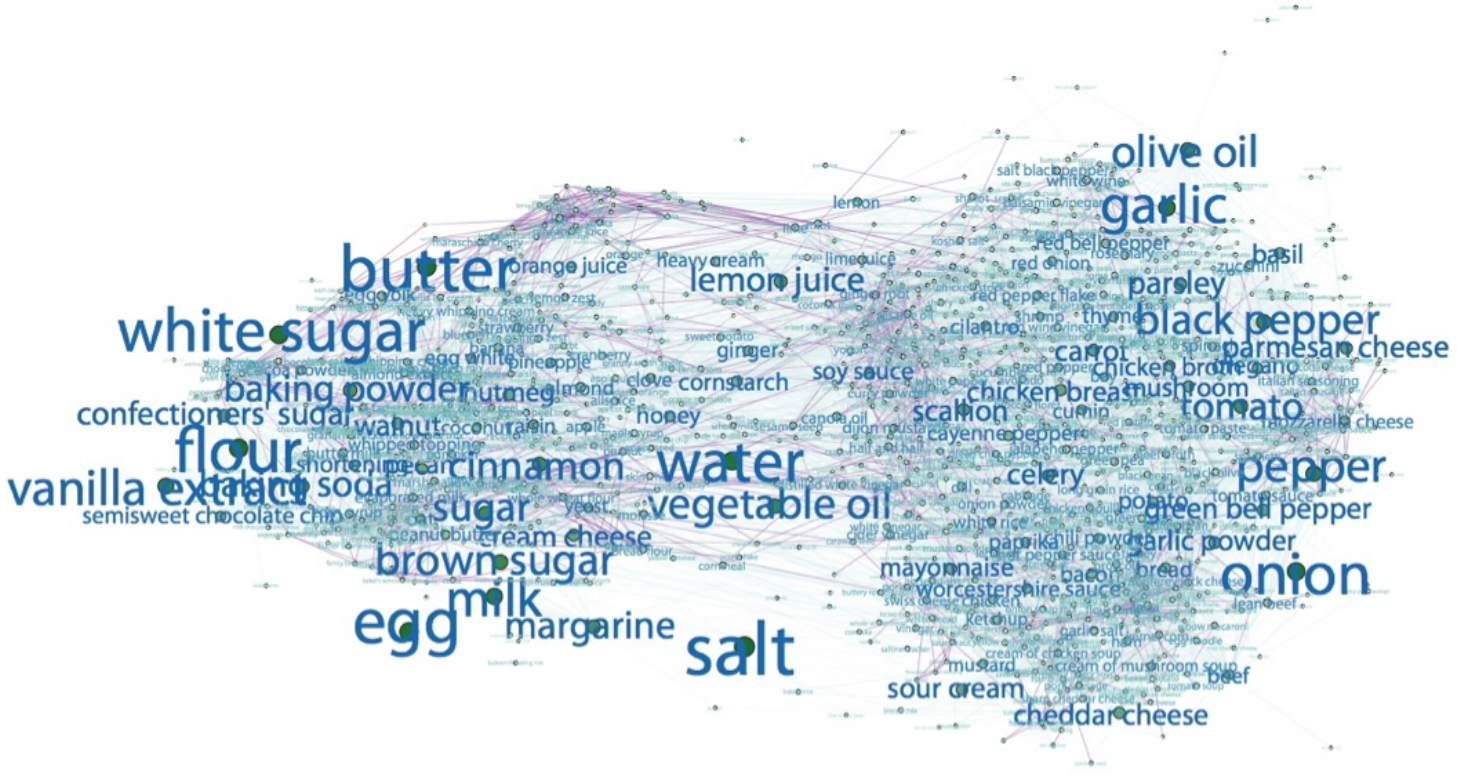
Connections between Political Blogs [Adamic-Glance, 2005].



- The network structure of political blogs prior to the 2004 U.S. Presidential election - who follows whom. Liberals' blogs are colored in blue, the conservatives' are colored in red.

Aspects of Networks - Cooking!

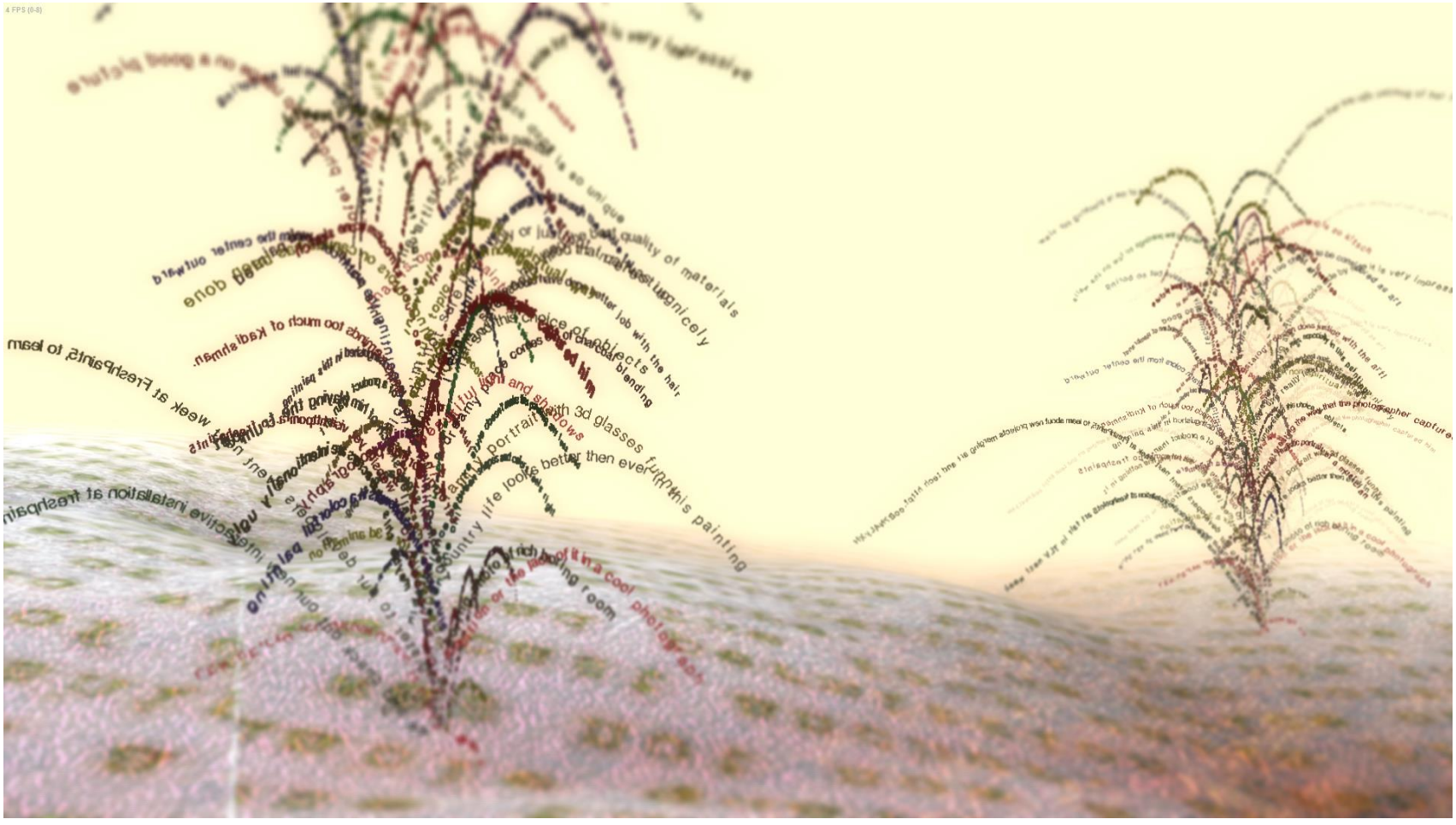
The Ingredients Network [Teng-Lin-Adamic, 2011].



- The network of recipe ingredients. Analyzed 1000s of recipes to find out what ingredients work well with what others.

Aspects of Networks - Natural Language!

The Word Networks



Networks Really Matter

Help us understand:

- The spread of diseases (hopeless without social networks)
- The WWW structure and information (hopeless without the web topology)
- The dissemination of news or the evolution of science (hopeless without information networks).
- And many, many others!

Applications

- Expert Finding
- Maximizing the Spread of Influence through a Social Network
- Minimizing the Spread of Misinformation in Social Networks
- Document or Citation Recommendation
- Predicting Collaborators in an Academic Network
- Predicting the “Follower" Relationship in an Academic Network
- Reviewers to Papers Assignment
- Committee Formation
- Topic Identification
- Recommender Systems

Use automated, machine learning and information retrieval techniques to solve them!

What is Machine Learning?
Next!