Social Network Analysis

Course Objectives, Outline, Grading, Readings

Instructors
Sougata Mukherjea Email: smukherj@in.ibm.com
Amit A. Nanavati Email: namit@in.ibm.com

Course Objectives

- Social Network Analysis (SNA) is about analysing networks arising in various contexts, especially those arising in social contexts: as a result of people connecting with each other on online social networks such as twitter and facebook, as well as "who-calls-whom" graph arising out of Telecom networks.

- However, the techniques for analysing social networks can be extended to other non-social networks as well.

Course Objectives

- In this course, we will learn about techniques for analysing networks, both social and others, and also learn about how (algorithms) to do this in a *scalable* manner.
- Also, we will learn how to visualise some of these large networks.
- And lastly, we will also explore the applications of SNA in various domains such as Telecom, Biology, etc

Course Outline

- Graph Theory and Social Networks
- Visualizing Social Networks
- Game Theory
- Information Networks and the World Wide Web
- Network Dynamics
- Applications of SNA in various domains

Grading (Tentative)

- Assignment 1 (5%) – Tooling assignment
- Assignment 2 (5%) – Coding assignment
- Assignment 3 (10%) – Open-ended problems

- Mid Term (15%)

- Project (40%) – 2 or 3 in a group
  - Problem Definition and Literature Survey (15%)
  - First set of Analysis and Results (10%)
  - Final Presentation & Submission of report in LaTeX/Word (15%)

- Final Exam (25%)

Readings

- "Networks, Crowds, and Markets: Reasoning About a Highly Connected World" by David Easley and Jon Kleinberg Cambridge University Press 2010.
- Online: https://www.cs.cornell.edu/home/kleinber/networks-book/
- Research Papers discussed in class.

Social Network Analysis

Introduction

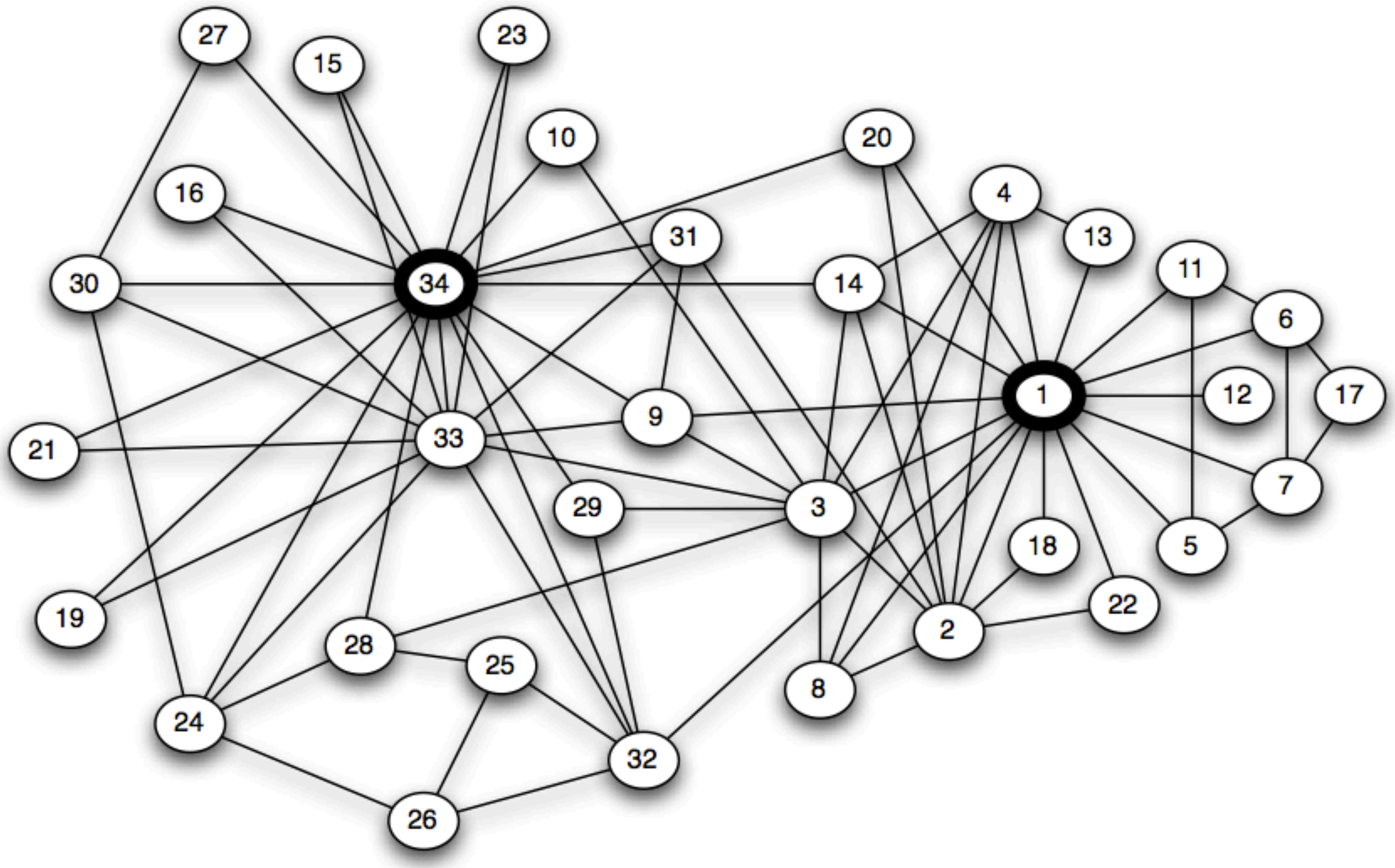
Introduction

…

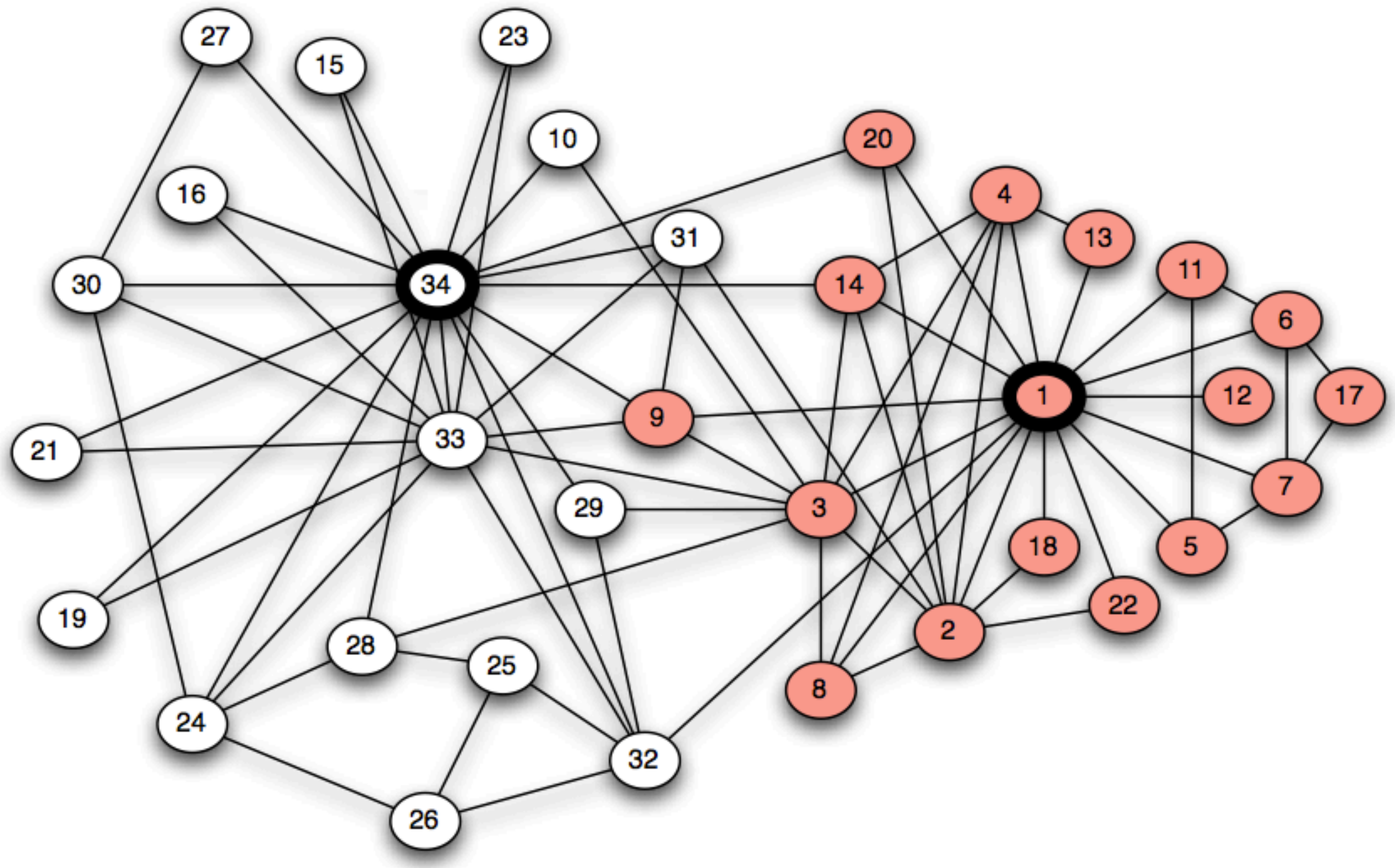
Course Outline

- Graph Theory and Social Networks
- Visualizing Social Networks
- Game Theory
- Information Networks and the World Wide Web
- Network Dynamics
- Applications of SNA in various domains

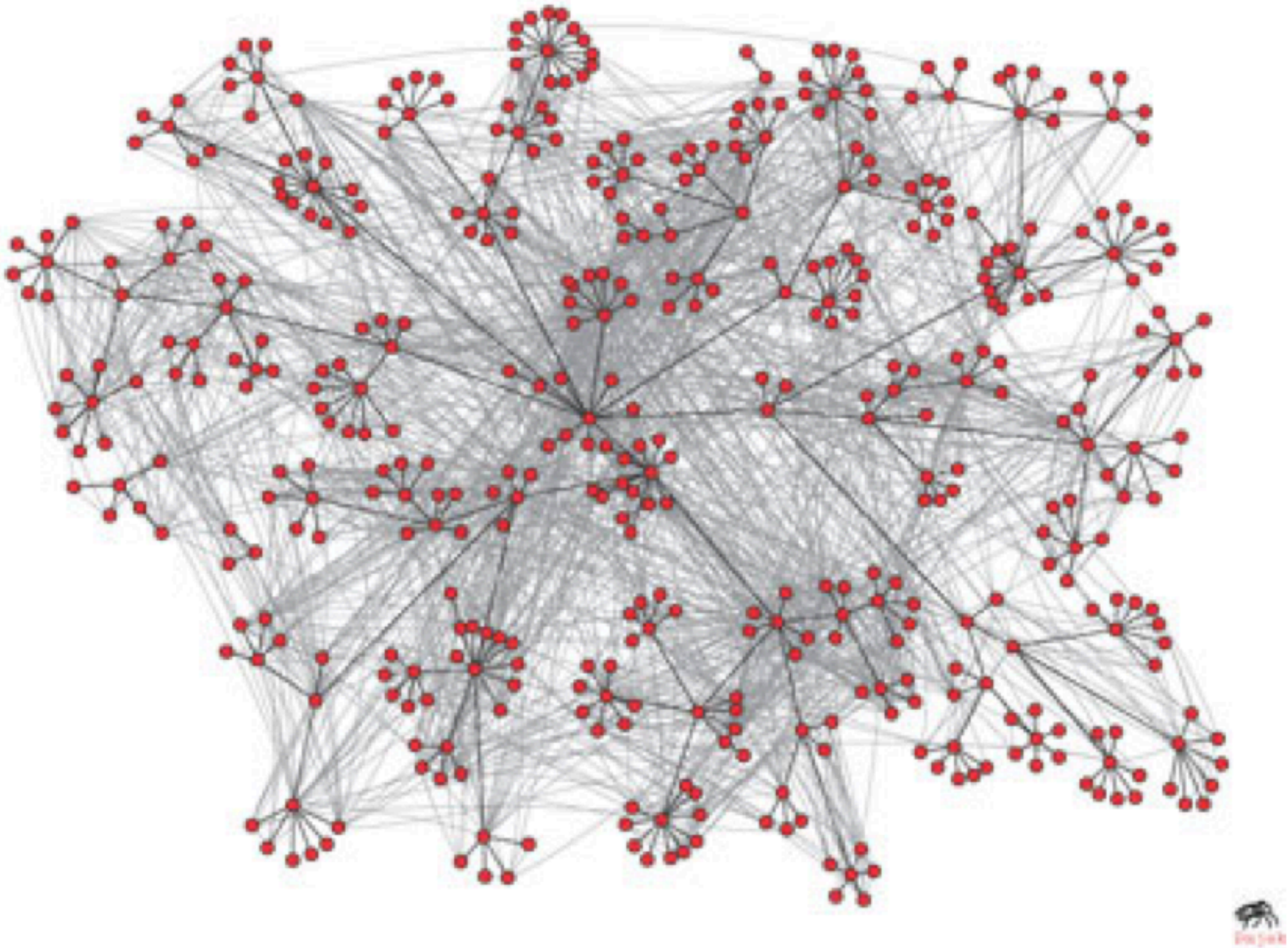
Introduction



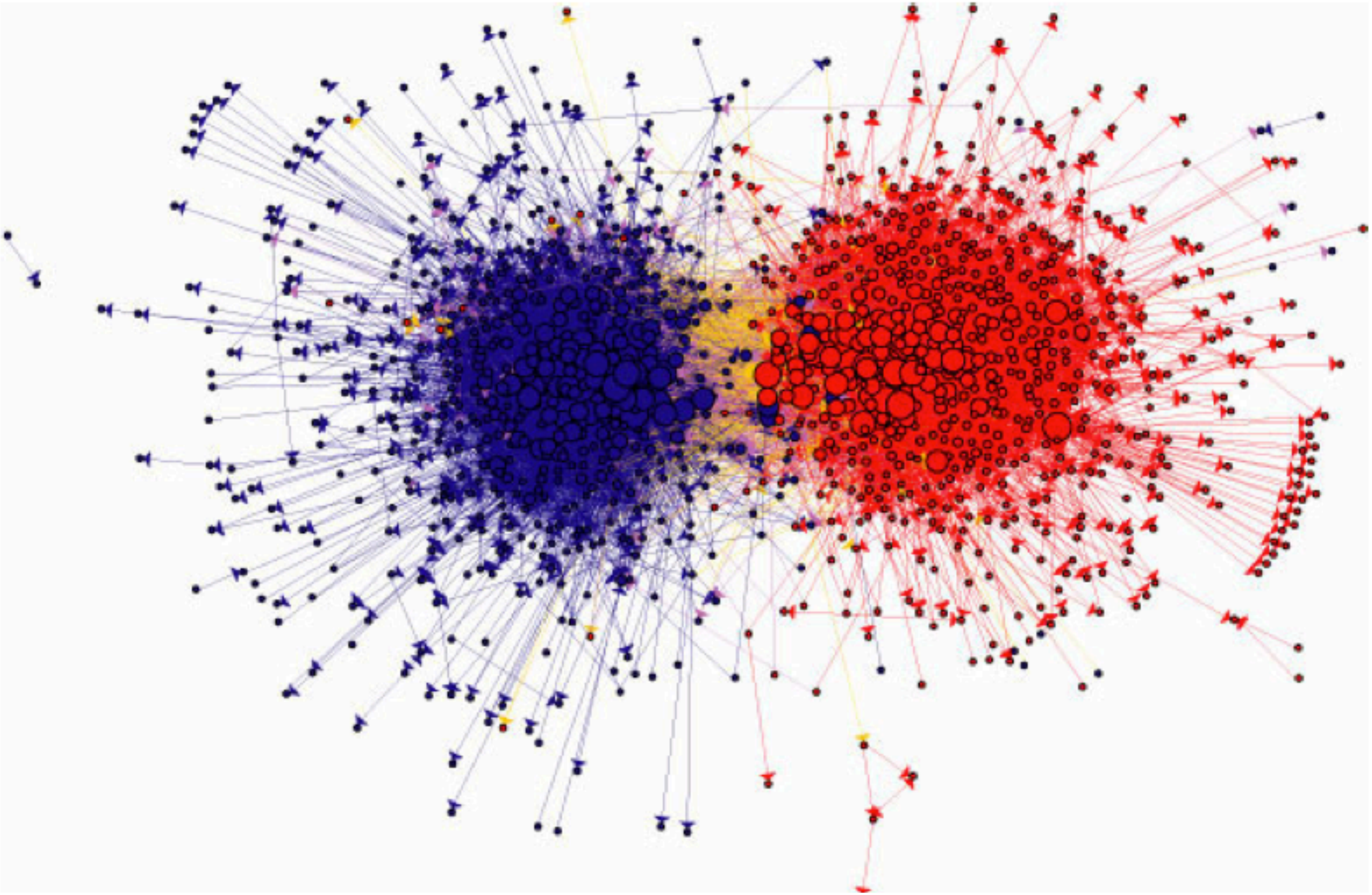
The social network of friendships within a 34-person karate club [2]



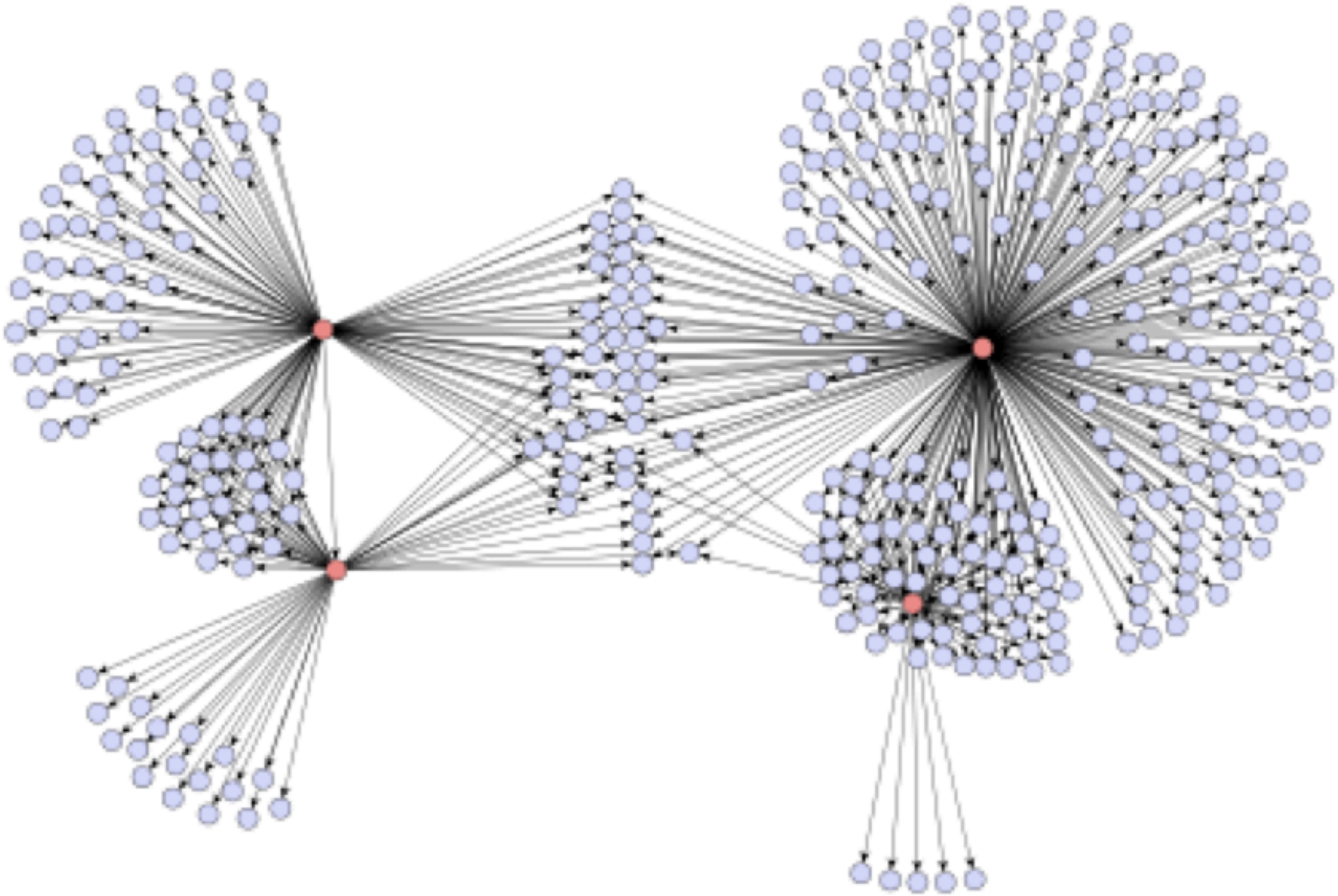
The social network of friendships within a 34-person karate club [2]



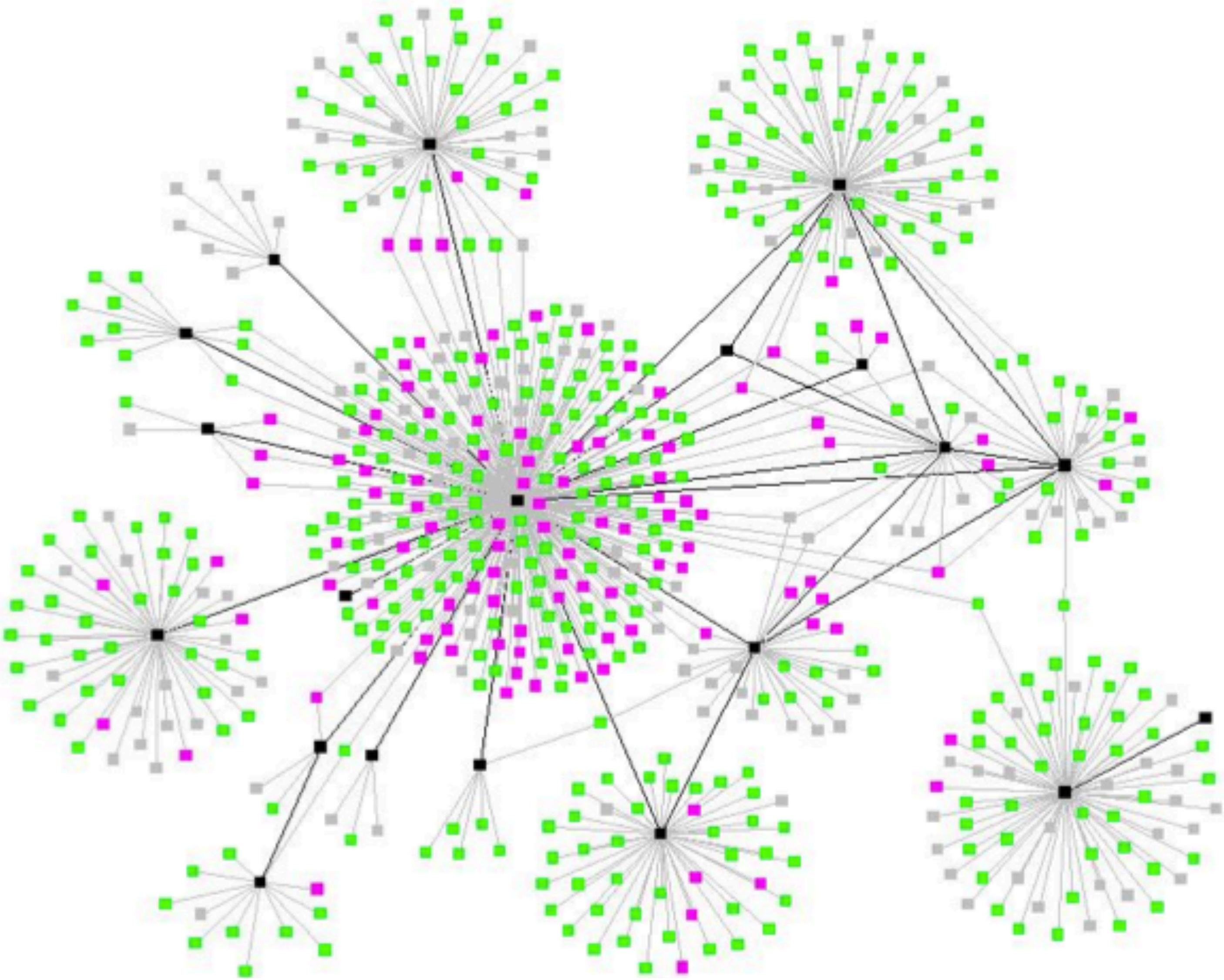
The pattern of e- mail communication among 436 employees of Hewlett Packard Research Lab is superimposed on the official organizational hierarchy [3]. (Image from http://www- personal.umich.edu/ ladamic/img/ hplabsemailhierarchy.jpg)



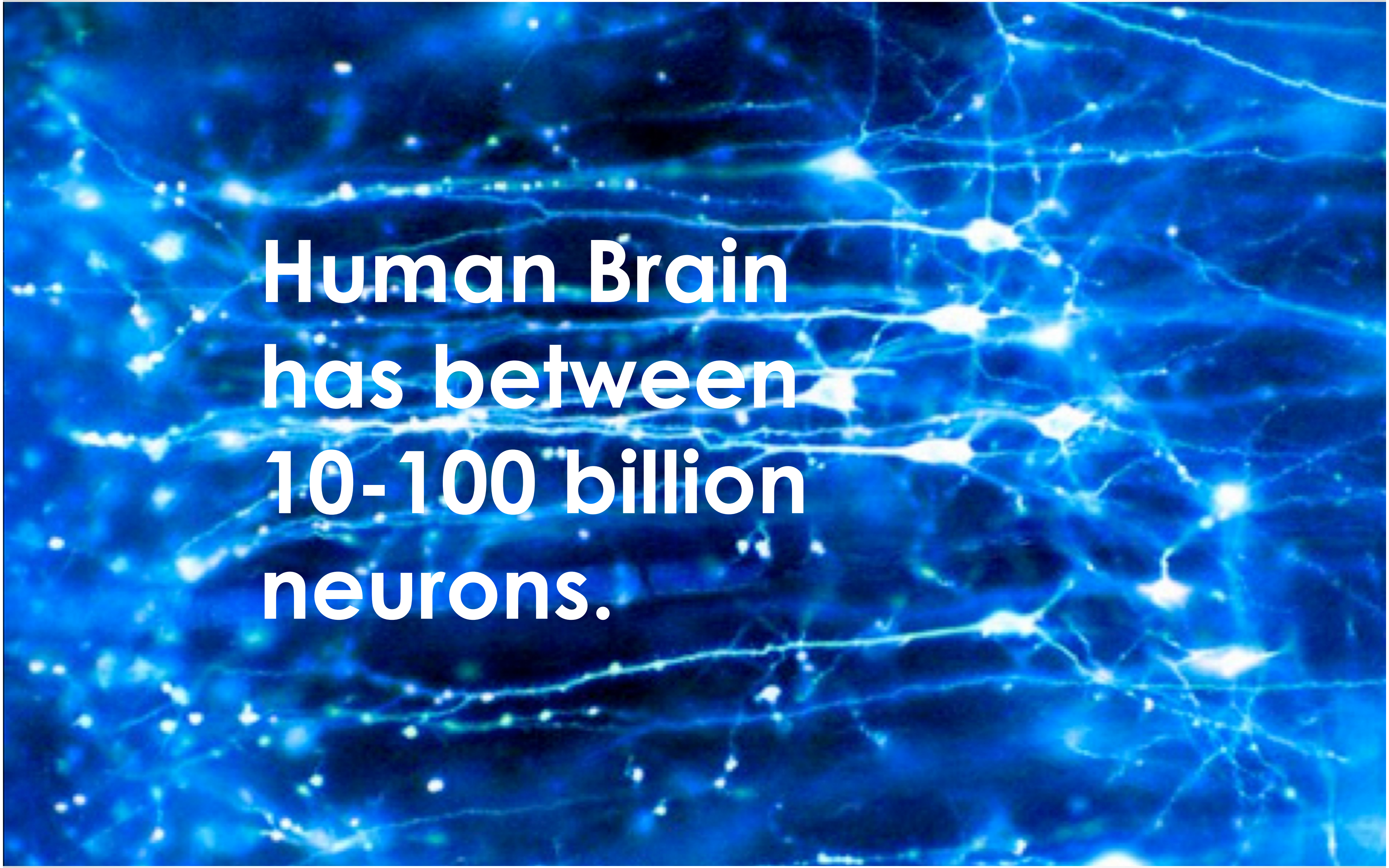
The network structure of political blogs prior to the 2004 U.S. Presidential election reveals two natural and well-separated clusters [4]. (Image from http://www- personal.umich.edu/ ladamic/img/politicalblogs.jpg)



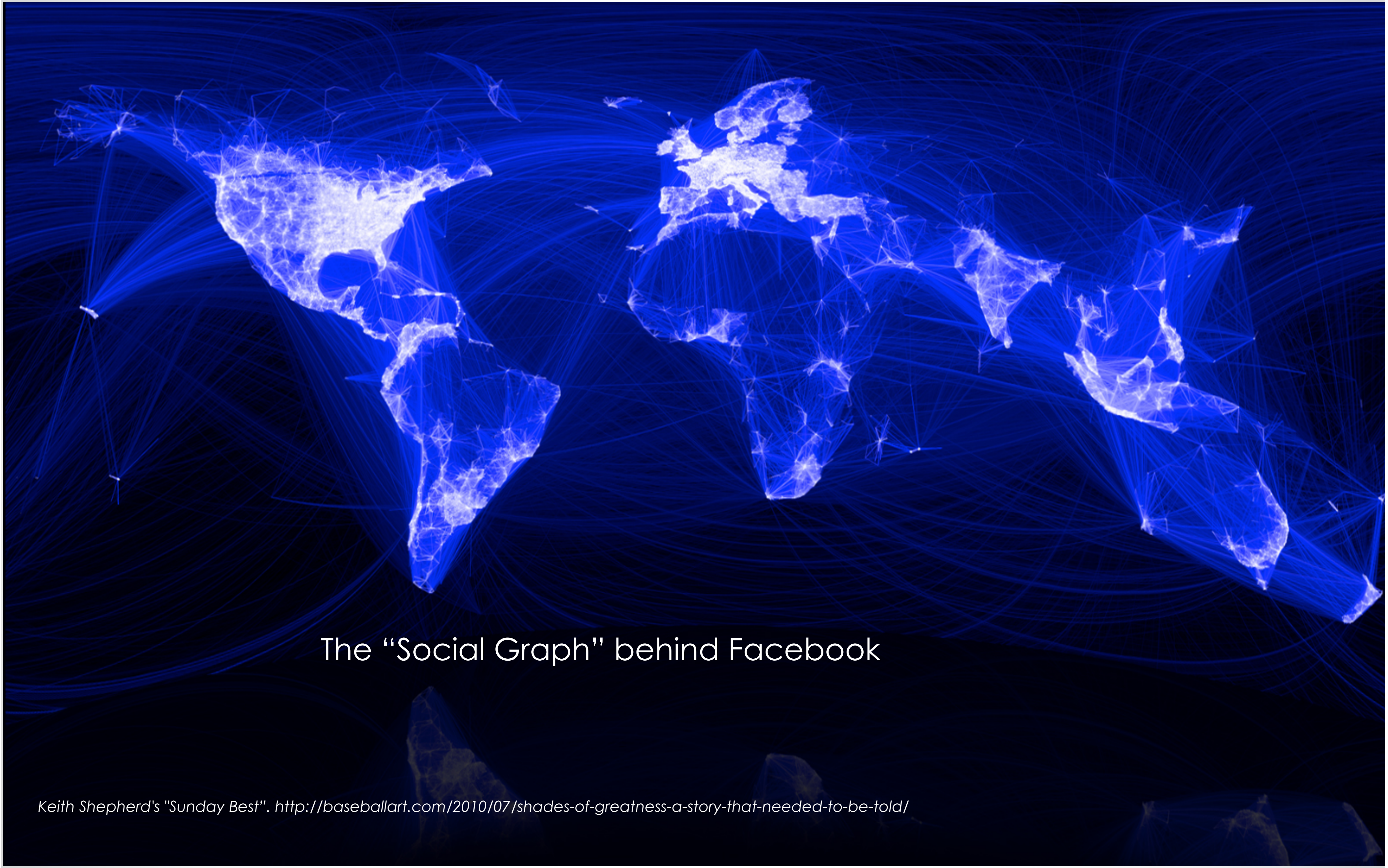
When people are influenced by the behaviors their neighbors in the network, the adoption of a new product or innovation can cascade through the network structure. Here, e-mail recommendations for a Japanese graphic novel spread in a kind of informational or social contagion. (Image from Leskovec et al. [5].)



The spread of an epidemic disease (such as the tuberculosis outbreak shown here) is another form of cascading behaviour in a network. The similarities and contrasts between biological and social contagion lead to interesting research questions. (Image from Andre et al. [6].)



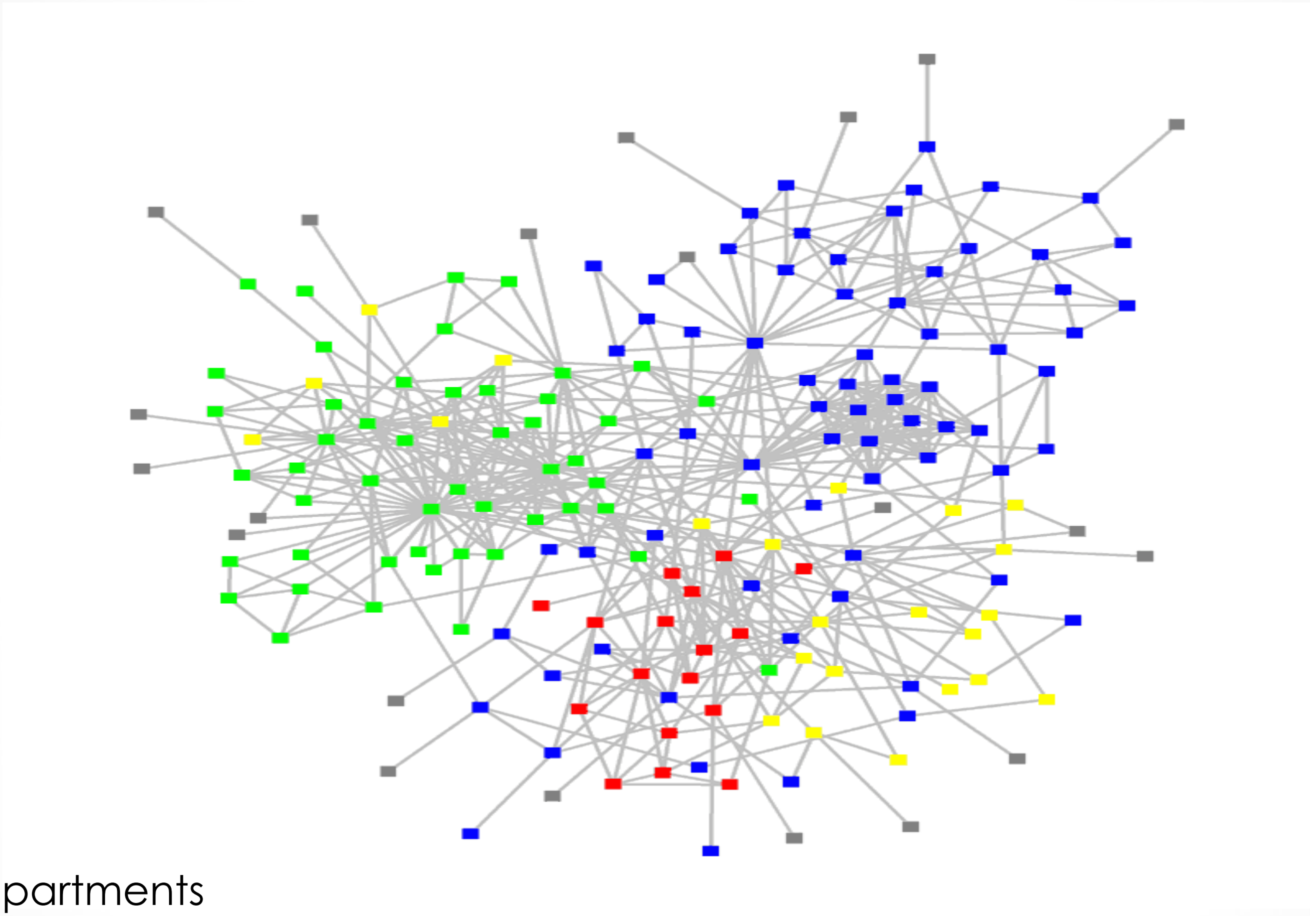
Human Brain has between 10-100 billion neurons.



The "Social Graph" behind Facebook

Keith Shepherd's "Sunday Best". http://baseballart.com/2010/07/shades-of-greatness-a-story-that-needed-to-be-told/

Structure of an Organisation

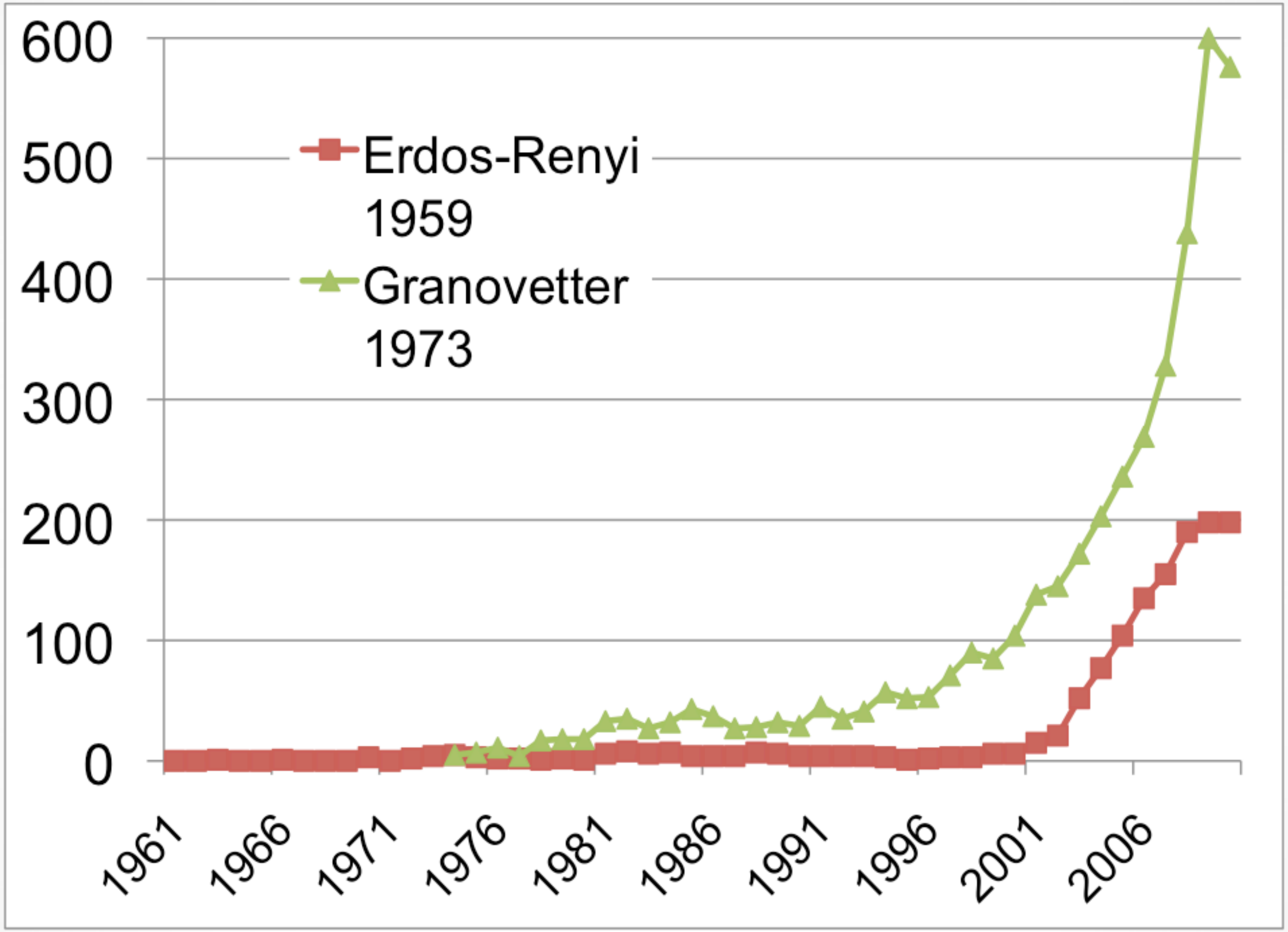


: departments

: consultants

: external experts

THE HISTORY OF NETWORK ANALYSIS



Characteristics of Network Science

Interdisciplinary

Empirical, data driven

Quantitative and Mathematical

Computational

Use of Network Science

How to understand the WWW structure?
It's searchability?

How to understand the spread of diseases?

How to understand the formation of links and communities?

How to visualise networks for specific insights?

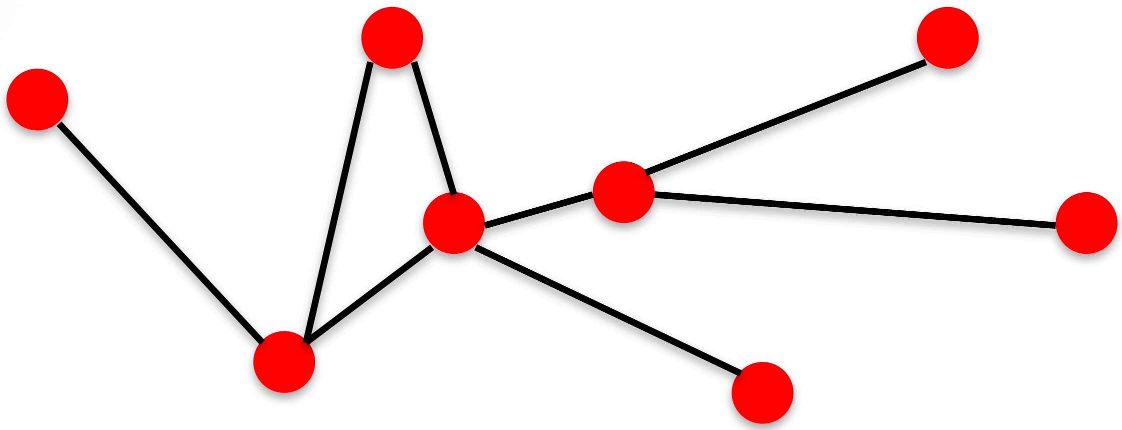
Course Outline

- Graph Theory and Social Networks
- Visualizing Social Networks
- Network Dynamics
- Information Networks and the World Wide Web
- Game Theory
- Applications of SNA in various domains

Preliminaries

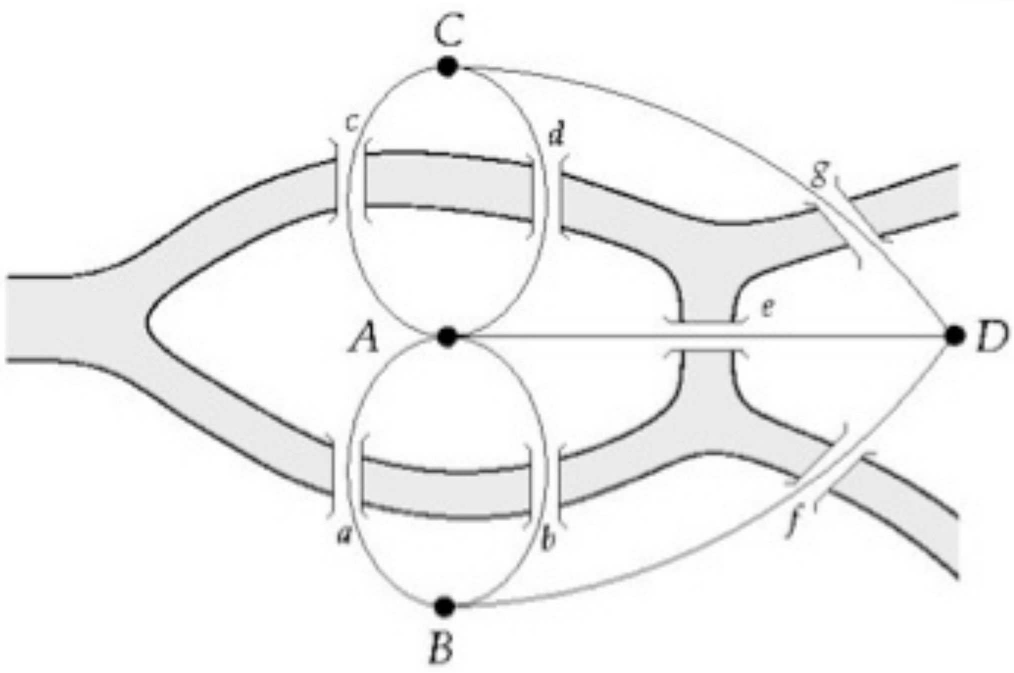
• • •

A Graph (Network)



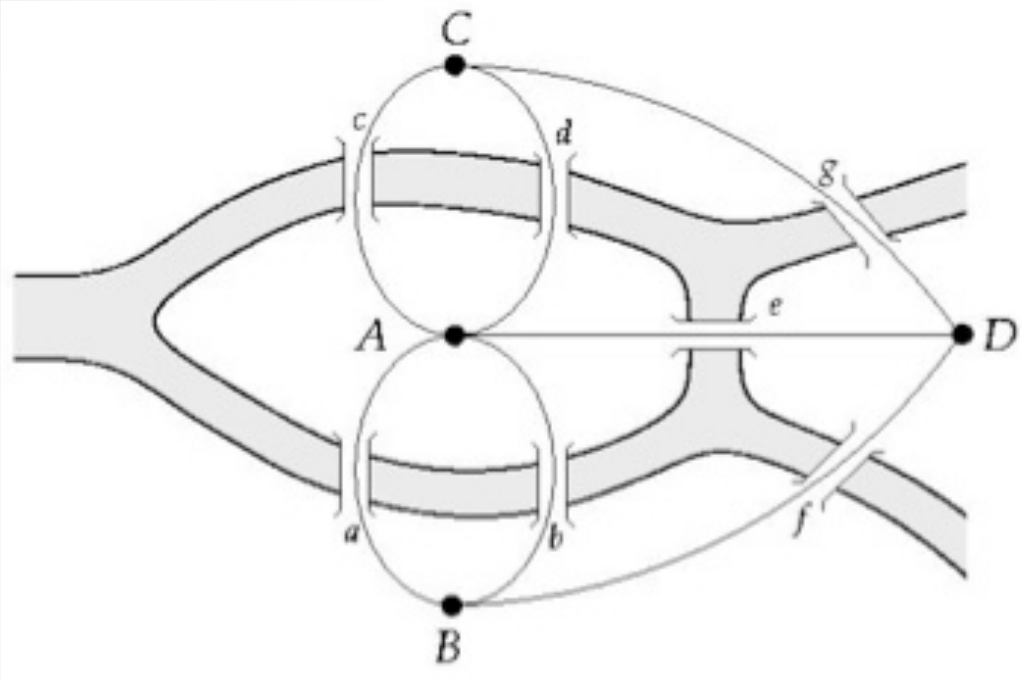
- **components**: nodes, vertices N
- **interactions**: links, edges L
- **system**: network, graph (N,L)

The Bridges of Königsberg



Can one walk across the seven bridges and never cross the same bridge twice?

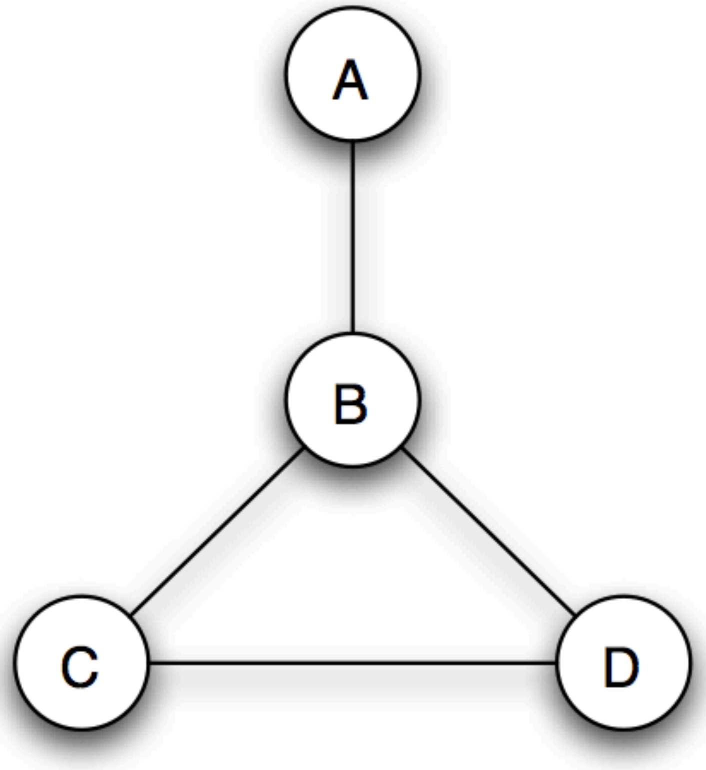
The Bridges of Königsberg



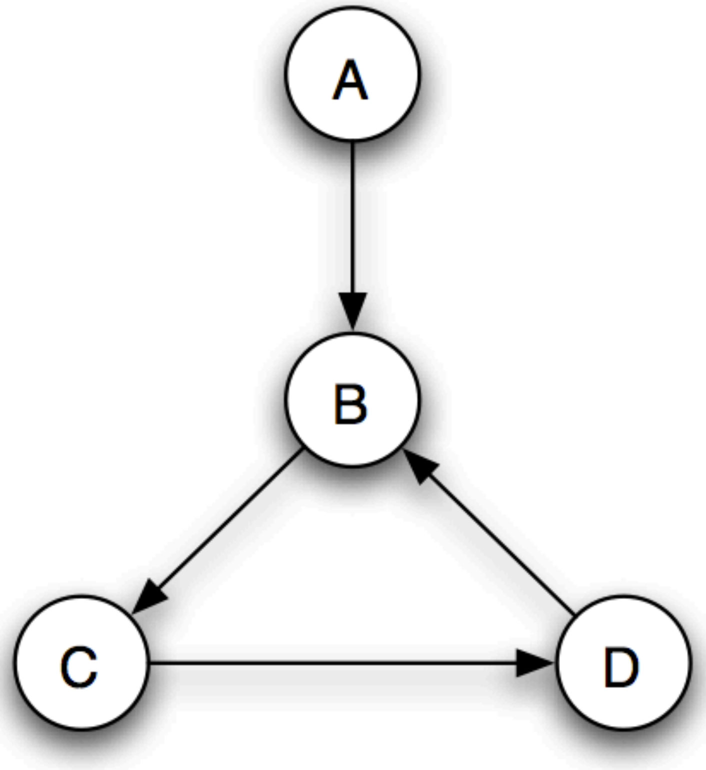
Can one walk across the seven bridges and never cross the same bridge twice?

1735: **Euler's theorem:**

(a) If a graph has more than two nodes of odd degree, there is no path.
(b) If a graph is connected and has no odd degree nodes, it has at least one path.



(a) *A graph on 4 nodes.*



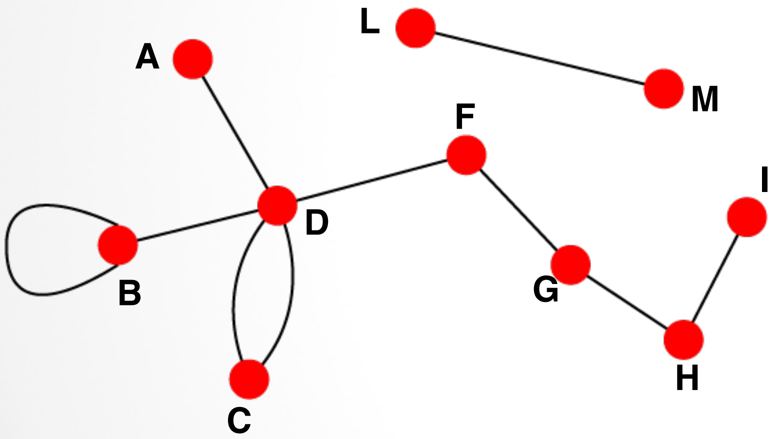
(b) *A directed graph on 4 nodes.*

Directed & Undirected Graphs

Undirected

Links: undirected (*symmetrical*)

Graph:

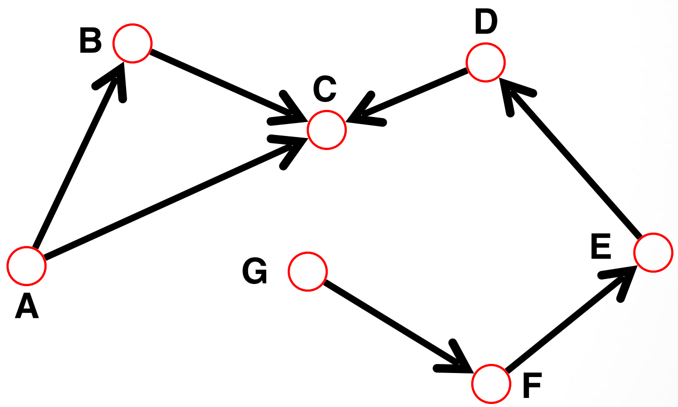


Undirected links :
coauthorship links
Actor network
protein interactions

Directed

Links: directed (*arcs*).

Digraph = directed graph:



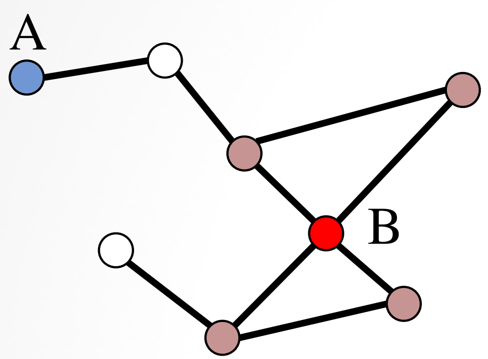
An undirected link is the superposition of two opposite directed links.

Example Networks

NETWORK	NODES	LINKS	DIRECTED UNDIRECTED	N	L
Internet	Routers	Internet connections	Undirected	192,244	609,066
WWW	Webpages	Links	Directed	325,729	1,497,134
Power Grid	Power plants, transformers	Cables	Undirected	4,941	6,594
Mobile Phone Calls	Subscribers	Calls	Directed	36,595	91,826
Email	Email addresses	Emails	Directed	57,194	103,731
Science Collaboration	Scientists	Co-authorship	Undirected	23,133	93,439
Actor Network	Actors	Co-acting	Undirected	702,388	29,397,908
Citation Network	Paper	Citations	Directed	449,673	4,689,479
E. Coli Metabolism	Metabolites	Chemical reactions	Directed	1,039	5,802
Protein Interactions	Proteins	Binding interactions	Undirected	2,018	2,930

Degree

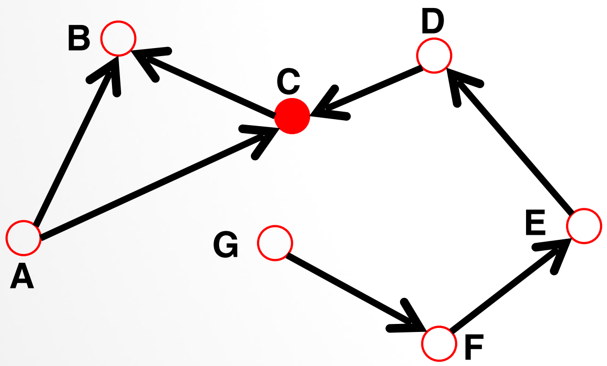
Undirected



Node degree: the number of links connected to the node.

$$k_A = 1 \qquad k_B = 4$$

Directed



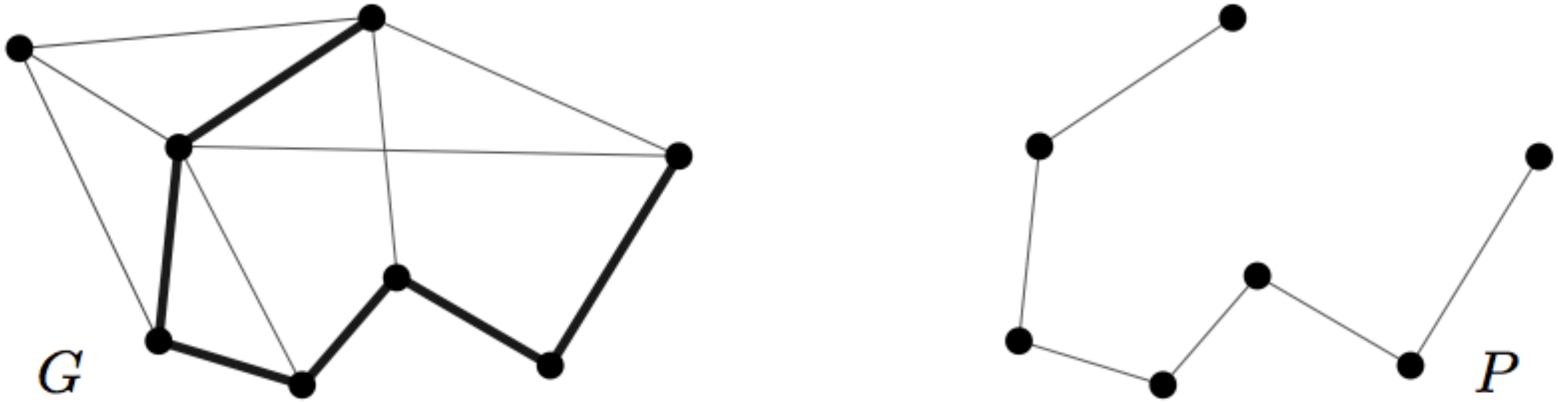
In *directed networks* we can define an in-degree and out-degree.

The (total) degree is the sum of in- and out-degree.

$$k_C^{in} = 2 \qquad k_C^{out} = 1 \qquad k_C = 3$$

Source: a node with $k^{in} = 0$; Sink: a node with $k^{out} = 0$.

Paths

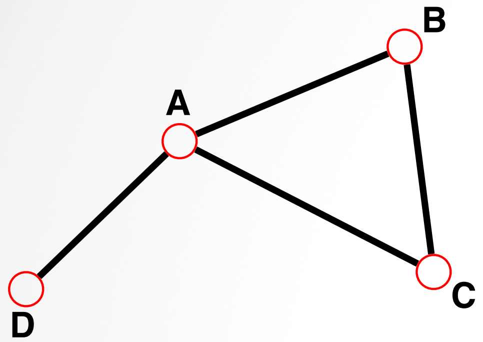


A *path* is a non-empty graph $P = (V, E)$ of the form

$$V = \{ x_0, x_1, \dots, x_k \} \qquad E = \{ x_0x_1, x_1x_2, \dots, x_{k-1}x_k \},$$

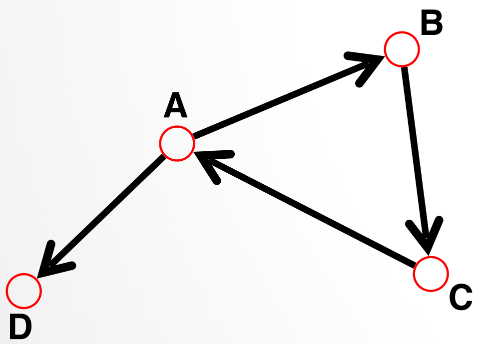
where the x_i are all distinct.

Distance



The *distance (shortest path, geodesic path)* between two nodes is defined as the number of edges along the shortest path connecting them.

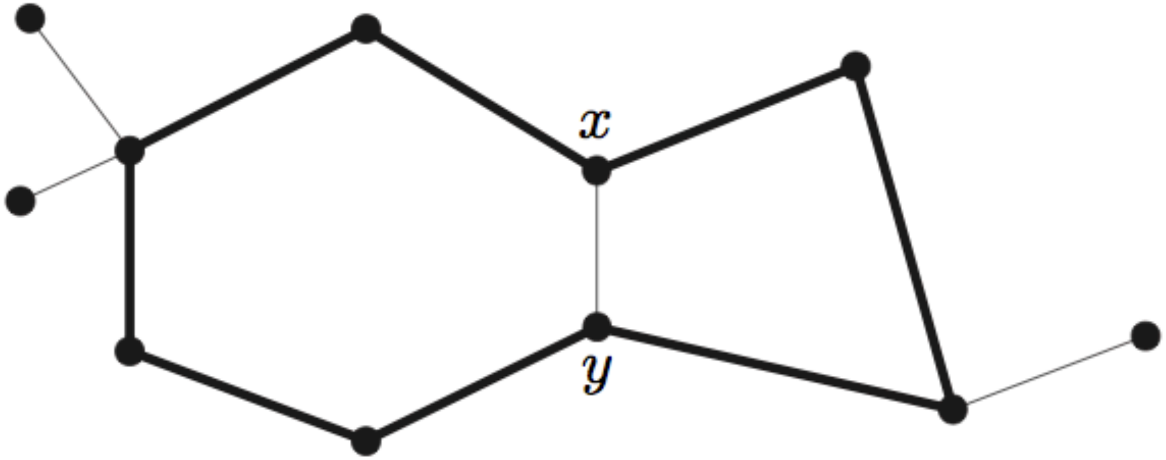
*If the two nodes are disconnected, the distance is infinity.



In directed graphs each path needs to follow the direction of the arrows.

Thus in a digraph the distance from node A to B (on an AB path) is generally different from the distance from node B to A (on a BCA path).

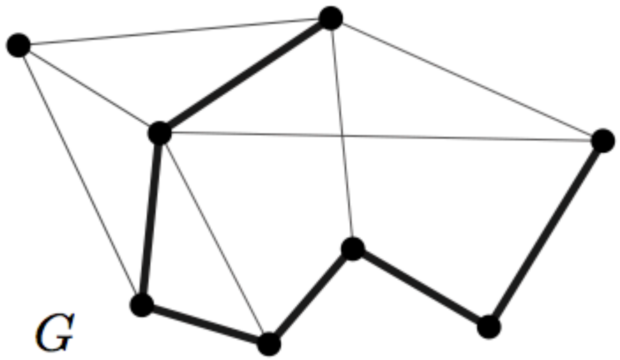
Cycles

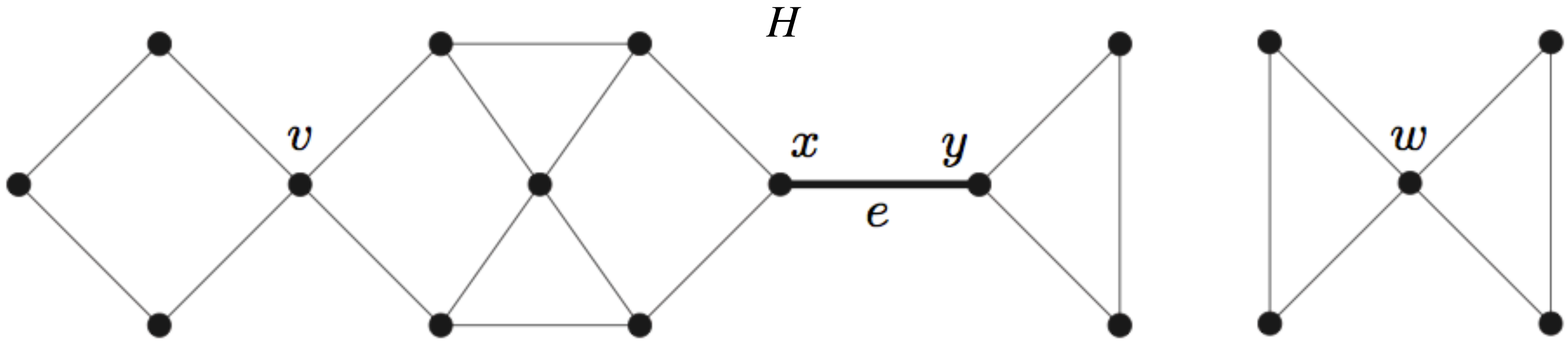


If $P = x_0 \ldots x_{k-1}$ is a path and $k \geqslant 3$,

then the graph $C := P + x_{k-1}x_0$ is called a *cycle*.

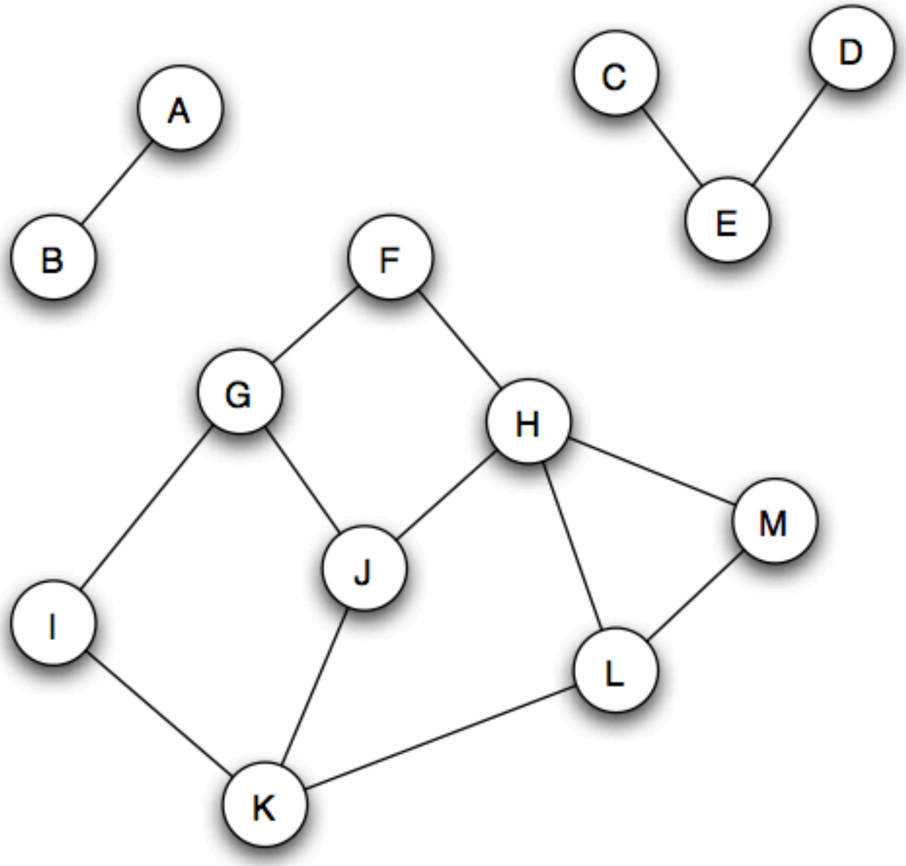
Connectivity





A non-empty graph G is called *connected* if any two of its vertices are linked by a path in G. If $U \subseteq V(G)$ and $G[U]$ is connected, we also call U itself connected (in G).

Connected Components



References

1. http://math.iisc.ernet.in/~nmi/NBSSEWSLectures.php
   Nosh's lecture 2, 3 and 4
2. Wayne Zachary. An information flow model for conflict and fission in small groups. Journal of Anthropological Research, 33(4):452–473, 1977.
3. Lada A. Adamic and Eytan Adar. How to search a social network. Social Networks, 27(3):187–203, 2005.
4. Lada Adamic and Natalie Glance. The political blogosphere and the 2004 U.S. election: Divided they blog. In Proceedings of the 3rd International Workshop on Link Discovery, pages 36–43, 2005.
5. Jure Leskovec, Lada Adamic, and Bernardo Huberman. The dynamics of viral marketing. ACM Transactions on the Web, 1(1), May 2007.
6. McKenzie Andre, Kashef Ijaz, Jon D. Tillinghast, Valdis E. Krebs, Lois A. Diem, Bev- erly Metchock, Theresa Crisp, and Peter D. McElroy. Transmission network analysis to complement routine tuberculosis contact investigations. American Journal of Public Health, 97(3):470–477, 2007.
7. F. Heart, A. McKenzie, J. McQuillian, and D. Walden. ARPANET Completion Report. Bolt, Beranek and Newman, 1978.
8. Images on slides 11, 12, 13 from http://barabasi.com/networksciencebook/
9. P. Erdős and A. Rényi. On Random Graphs, 1959 (http://www.renyi.hu/~p_erdos/1959-11.pdf)
10. Granovetter, Mark S. "The strength of weak ties." *American journal of sociology* (1973): 1360-1380.

Prime Reference for the Course and Thanks to:

https://www.cs.cornell.edu/home/kleinber/networks-book/