Lecture 2

Network Science

GRAPHS

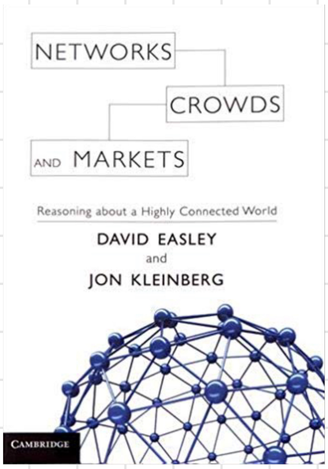
Todays Topics

Basic Definitions

Paths and Connectivity

Distance and BFSearch

Network Data Sets



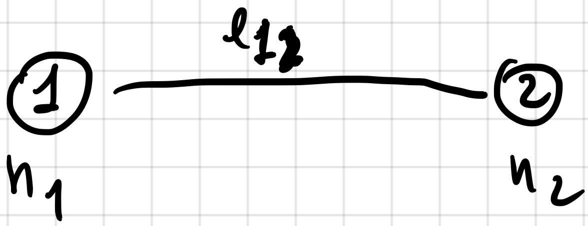
Chapter 2
"Graphs"

Basic Definitions

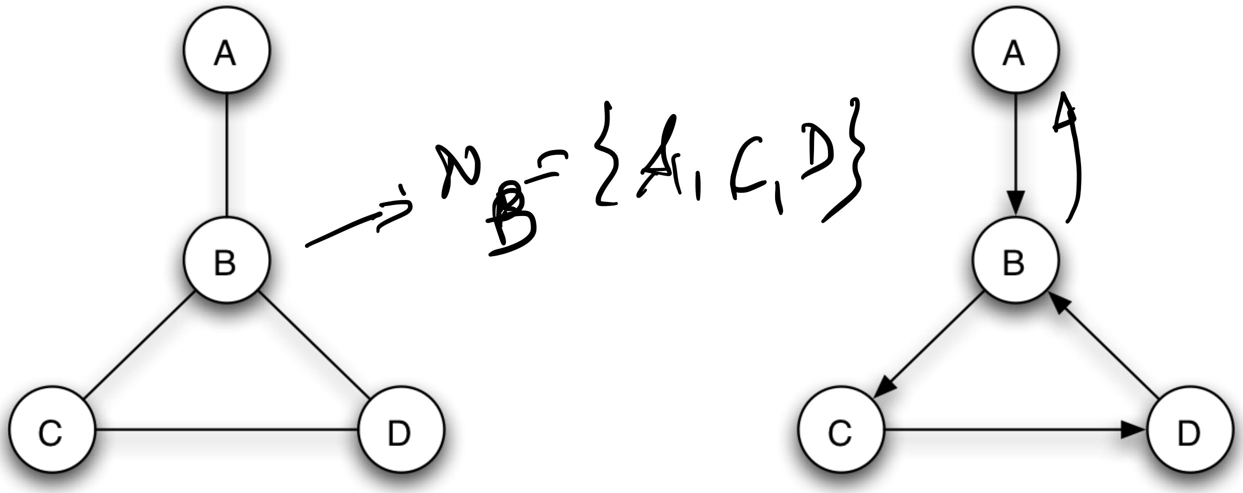
$$G = (N, E)$$

N : nodes $= \{n_1, n_2, \ldots, n_t\}$

E : edges $= \{(n_i, n_j) \ldots\}$



N_i : set of Neighbors of node i



(a) *A graph on 4 nodes.*

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

Figure 2.1: Two graphs: (a) an undirected graph, and (b) a directed graph.

Graphs as Mathematical Models

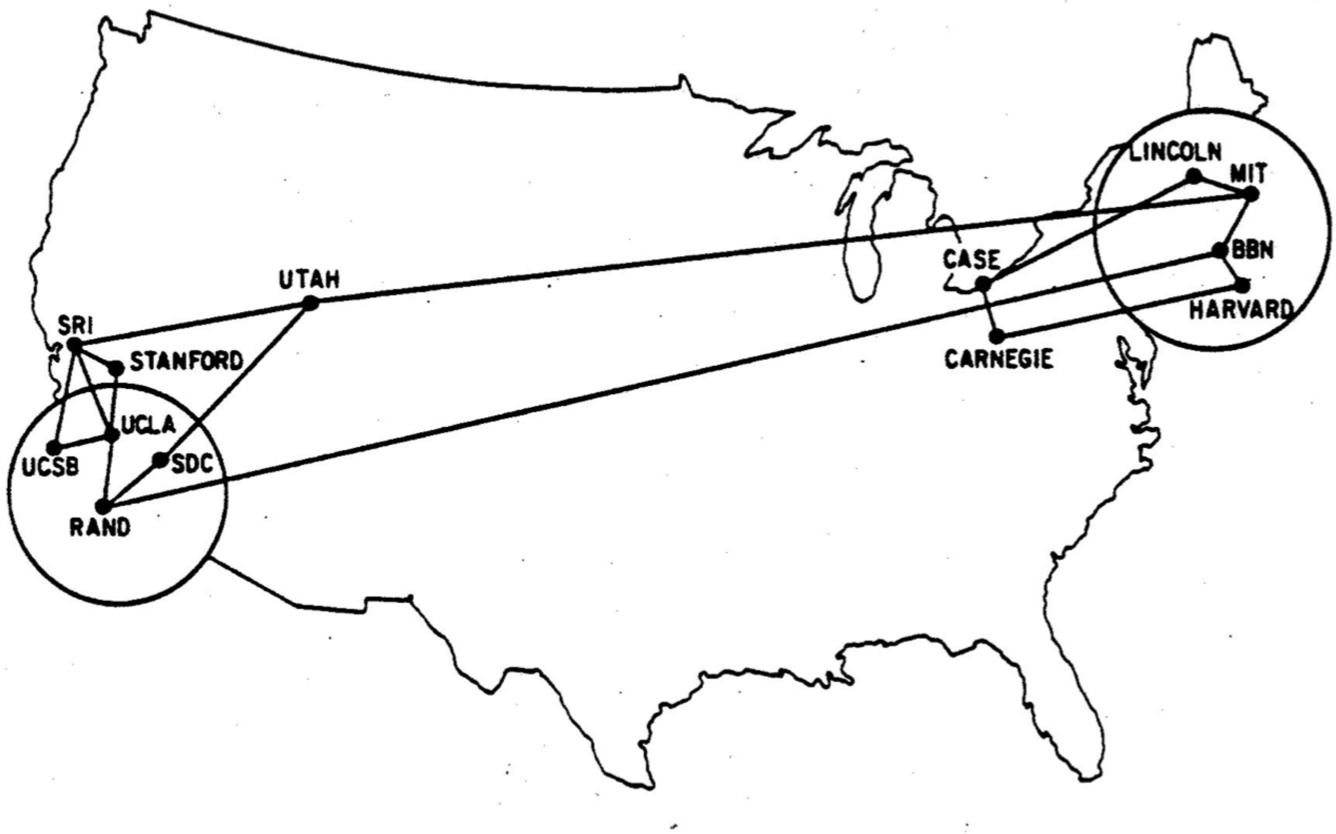


Figure 2.2: A network depicting the sites on the Internet, then known as the Arpanet, in December 1970. (Image from F. Heart, A. McKenzie, J. McQuillian, and D. Walden [214]; on-line at http://som.csudh.edu/cis/lpress/history/arpamaps/.)

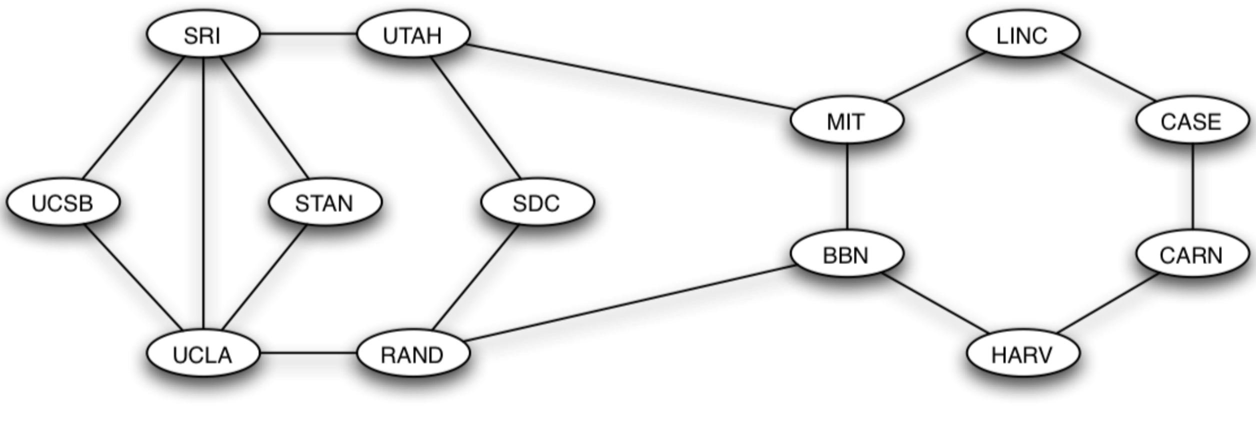
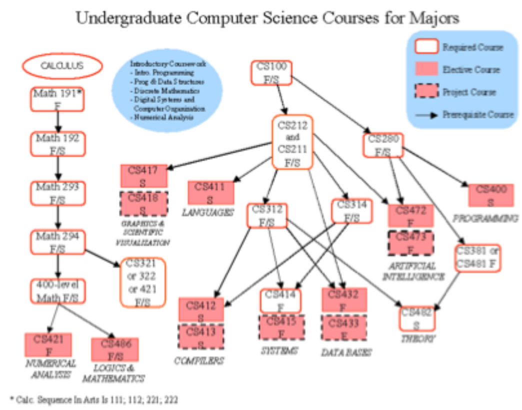


Figure 2.3: An alternate drawing of the 13-node Internet graph from December 1970.

(a) *Airline routes*

(b) *Subway map*



(c) *Flowchart of college courses*

(d) *Tank Street Bridge in Brisbane*

Figure 2.4: Images of graphs arising in different domains. The depictions of airline and subway systems in (a) and (b) are examples of *transportation networks*, in which nodes are destinations and edges represent direct connections. Much of the terminology surrounding graphs derives from metaphors based on transportation through a network of roads, rail lines, or airline flights. The prerequisites among college courses in (c) is an example of a *dependency network*, in which nodes are tasks and directed edges indicate that one task must be performed before another. The design of complex software systems and industrial processes often requires the analysis of enormous dependency networks, with important consequences for efficient scheduling in these settings. The Tank Street Bridge from Brisbane, Australia shown in (d) is an example of a *structural network*, with joints as nodes and physical linkages as edges. The internal frameworks of mechanical structures such as buildings, vehicles, or human bodies are based on such networks, and the area of *rigidity theory*, at the intersection of geometry and mechanical engineering, studies the stability of such structures from a graph-based perspective [388]. (Images: (a) www.airlineroutemaps.com/USA/Northwest_Airlines_asia_pacific.shtml, (b) www.wmata.com/metrorail/systemmap.cfm, (c) www.cs.cornell.edu/ugrad/flowchart.htm.)

Paths and Connectivity

N : nodes (vertices, actors, sites)
L: links (edges, arcs)

<u>Path</u>

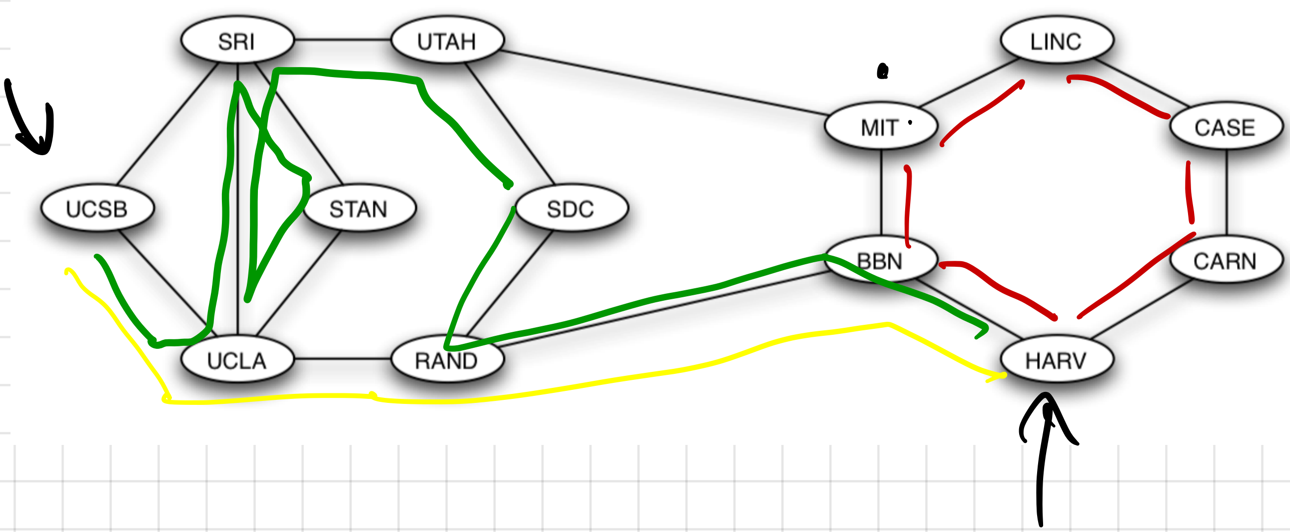
$p = \{n_1, n_2, \ldots, n_e\}$
length $= e$
$\forall i \quad (n_i, n_{i+1}) \in L$

repeating nodes are allowed!
(cycles ...)

no repeating nodes
$\Rightarrow$ simple path

paths

↳ simple path



loop (cycle)

$n_1 = n_e$

Cycle in a social network

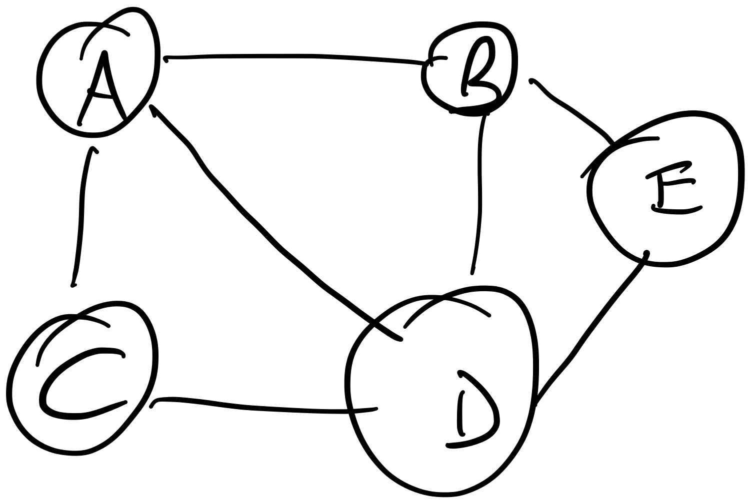
Alice

Bob

Carmen

Diego

Edward



Components

A graph is connected if for every pair of nodes there is at least one path connecting them

connected components:
a subset of the nodes s.t.
i) every nodes in the subset has a path to every other noded in the subset
ii) the subset is not part of some larger set of nodes that have property (i)

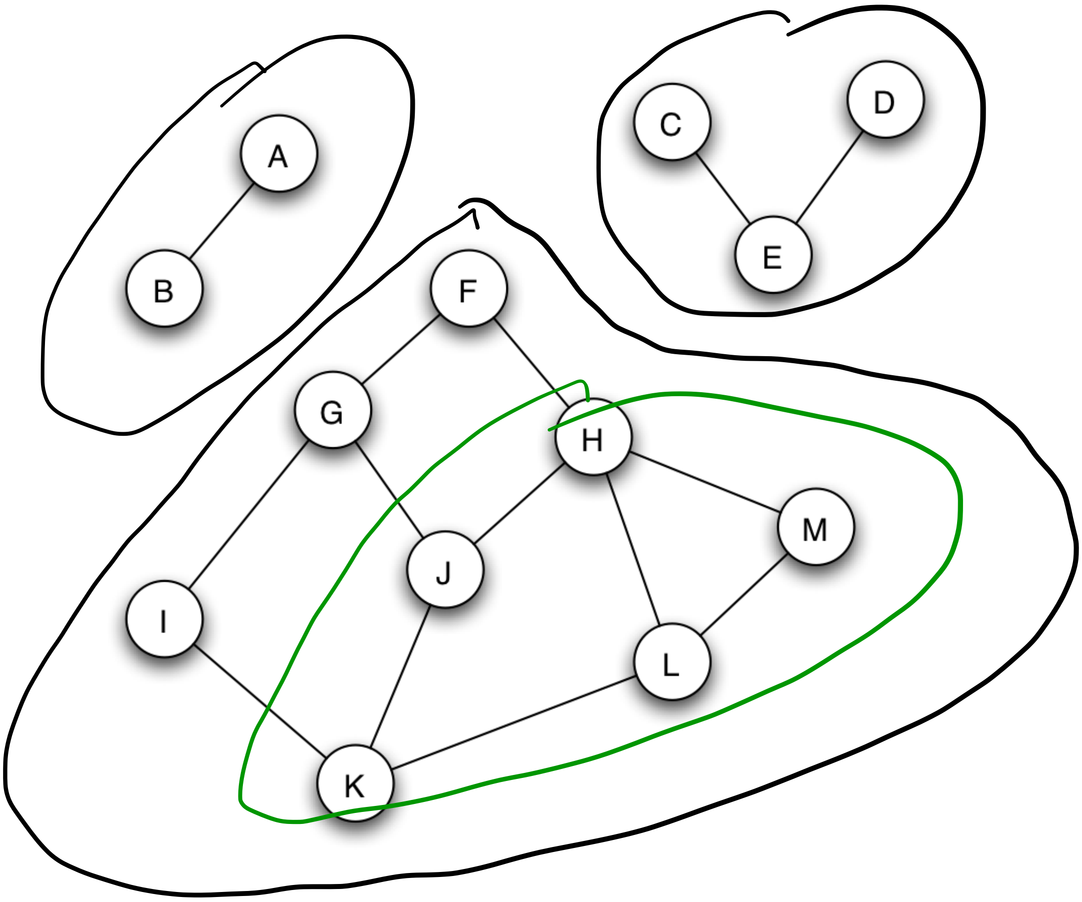


Figure 2.5: A graph with three connected components.

— : not a component

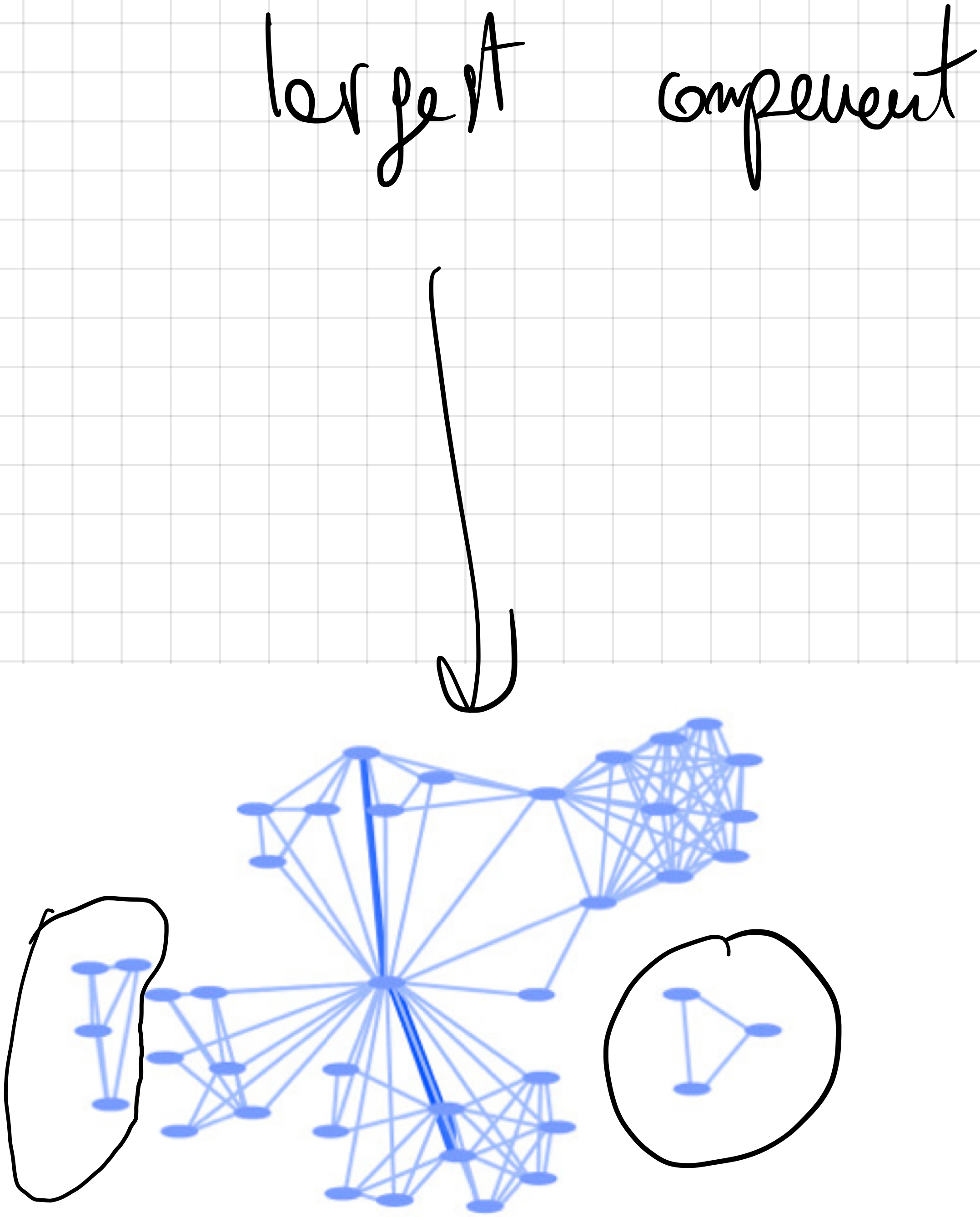
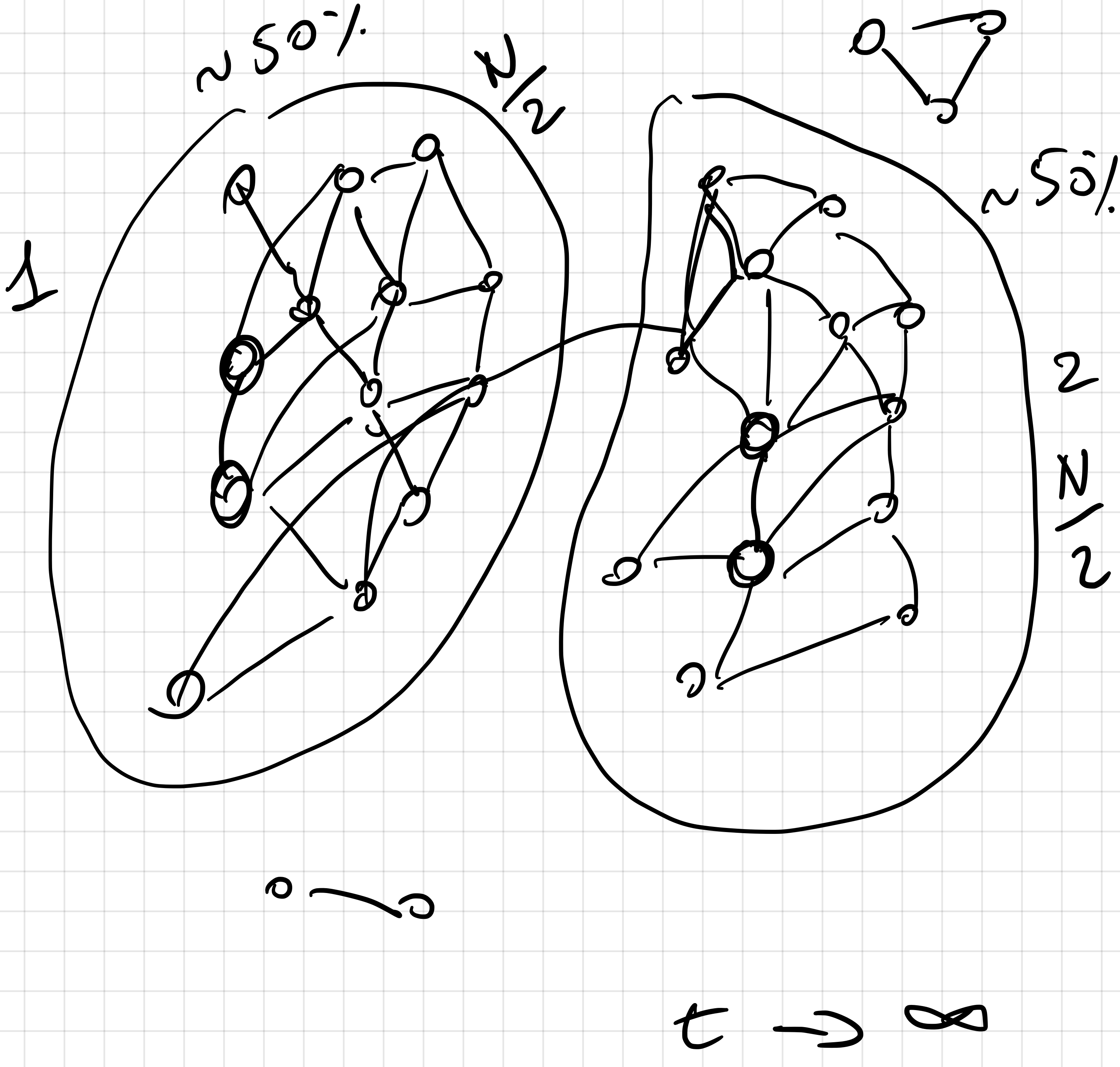


Figure 2.6: The collaboration graph of the biological research center *Structural Genomics of Pathogenic Protozoa (SGPP)* [134], which consists of three distinct connected components. This graph was part of a comparative study of the collaboration patterns graphs of nine research centers supported by NIH's Protein Structure Initiative; SGPP was an intermediate case between centers whose collaboration graph was connected and those for which it was fragmented into many small components.

Giant Component



⇒ the probability that two largest component co-exist in a large Network is very low!

two facts about Giant Component in Random Graphs:

1. Given N nodes and creating a link between two randomly selected nodes at each step, after a while only one Giant Component will emerge

2. The Giant Component emerges surprisingly soon: when the average degree is > 1 !

See agent based simulation with NetLogo !

Expectations come true with many dataset!

one Giant Component (and average degree very small)

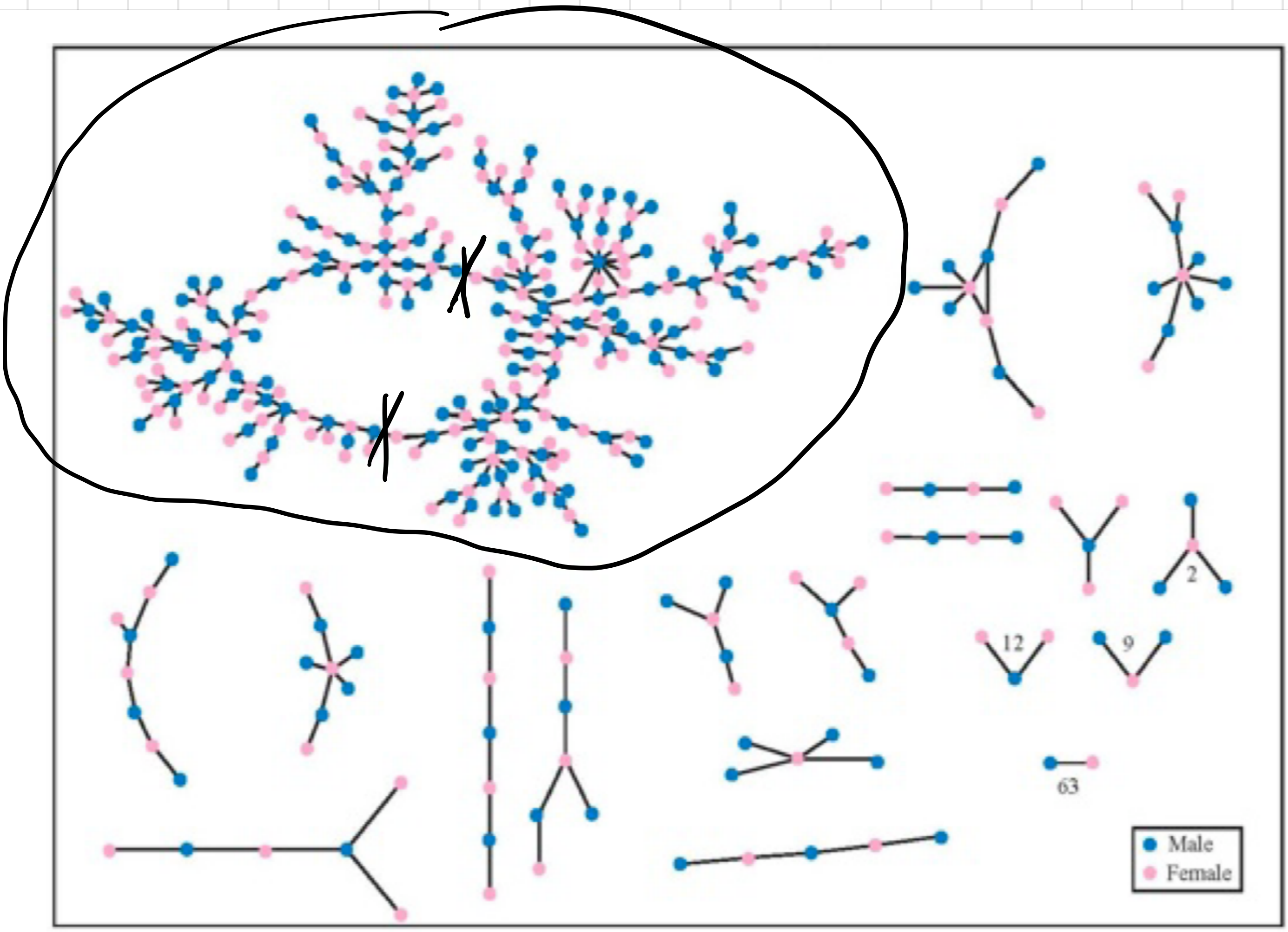
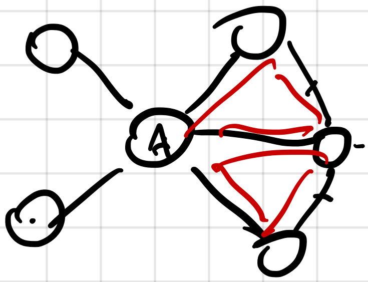


Figure 2.7: A network in which the nodes are students in a large American high school, and an edge joins two who had a romantic relationship at some point during the 18-month period in which the study was conducted [49].

Other Measures

- clustering coefficient

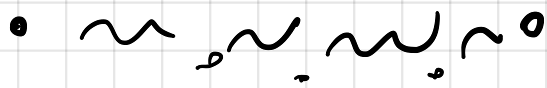


$$cc(A) = \frac{2\cdot 2}{5\cdot 4} = \frac{1}{5}$$

- degree, average degree
  degree distribution
- centralities
- assortativity
- homophily

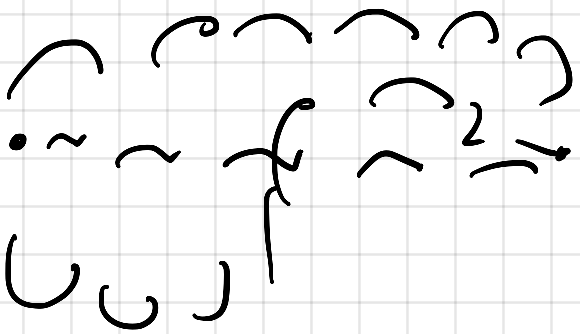
...later in this course

Distance and Bredth - First Search



a path between two nodes

=> length



the shortest path

=> length

=> distance

BFS

Breedth First Search

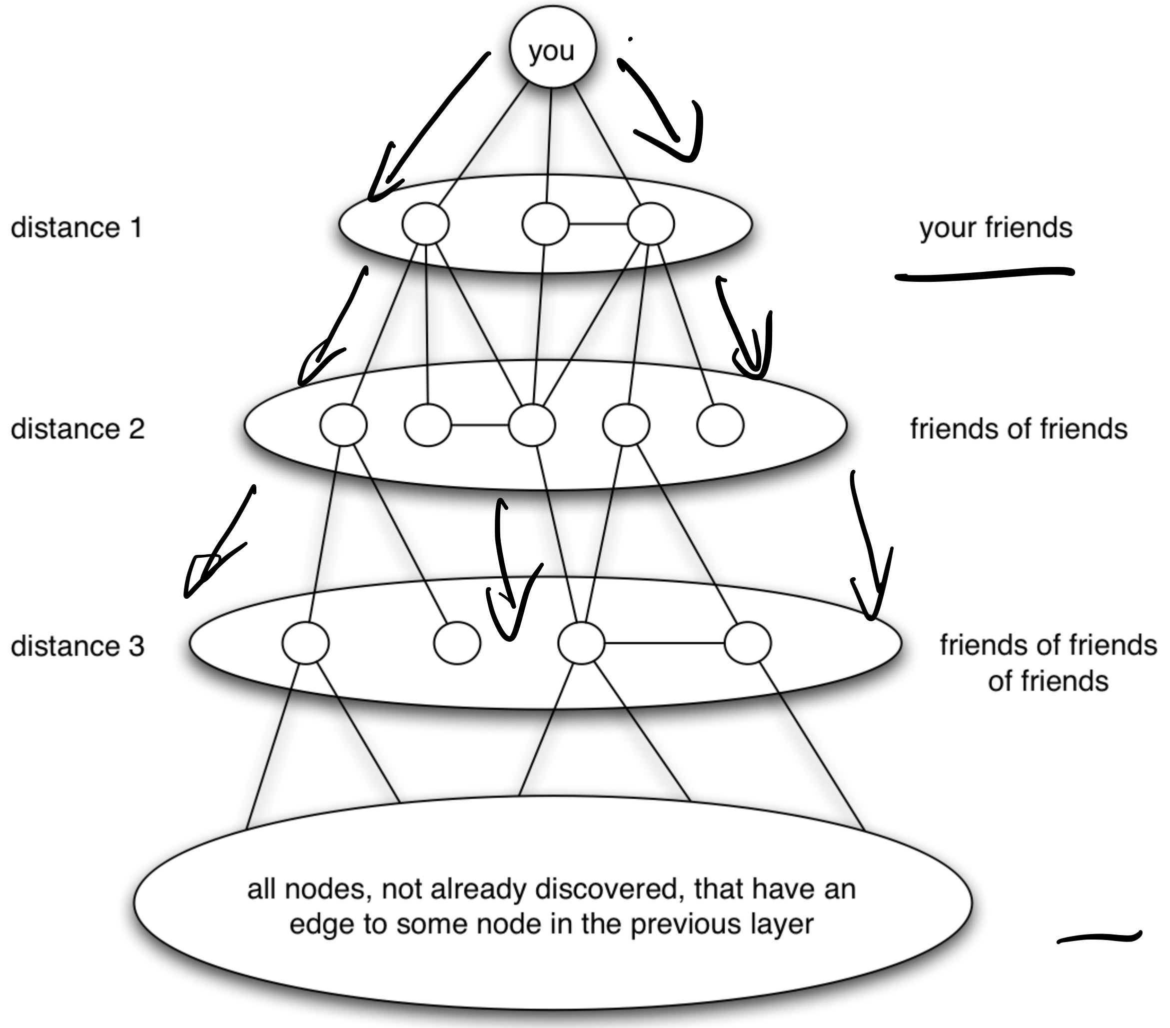


Figure 2.8: Breadth-first search discovers distances to nodes one "layer" at a time; each layer is built of nodes that have an edge to at least one node in the previous layer.

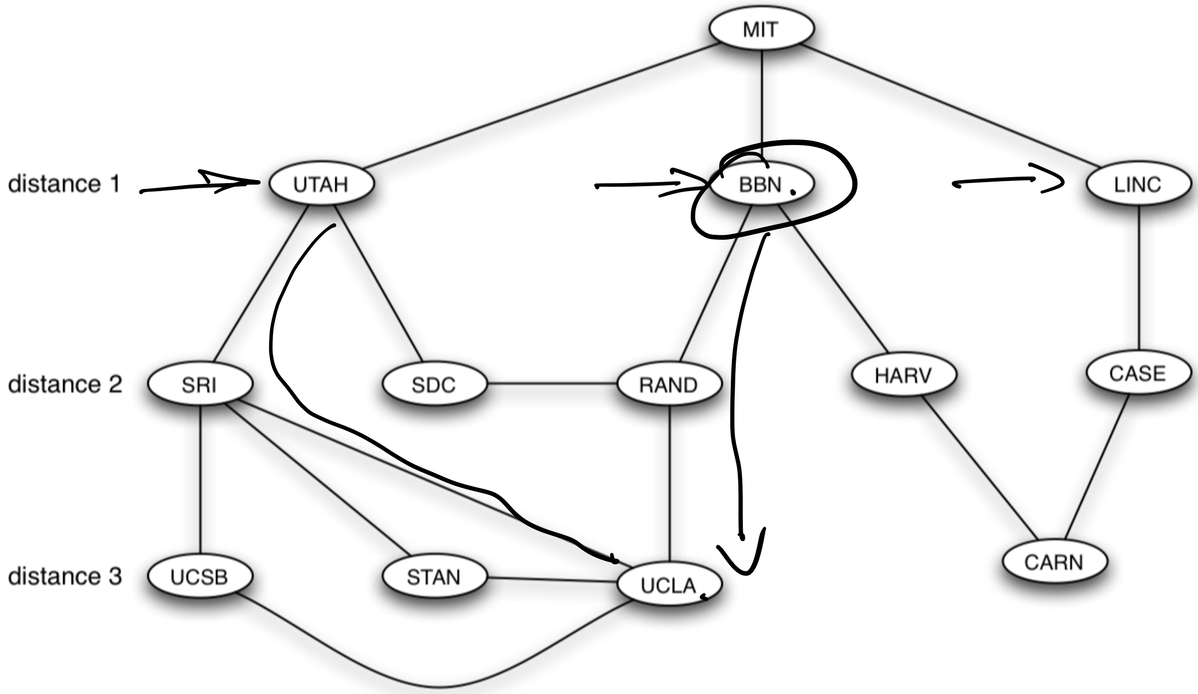


Figure 2.9: The layers arising from a breadth-first of the December 1970 Arpanet, starting at the node MIT.

Small World phenomenon

Stanley Milgram (1960)

Omaha : 296 random addresses.

"send the letter to THIS stock brocker in Boston"

a relevant percentage of these letters arrived to the target through some hops.

median = 6

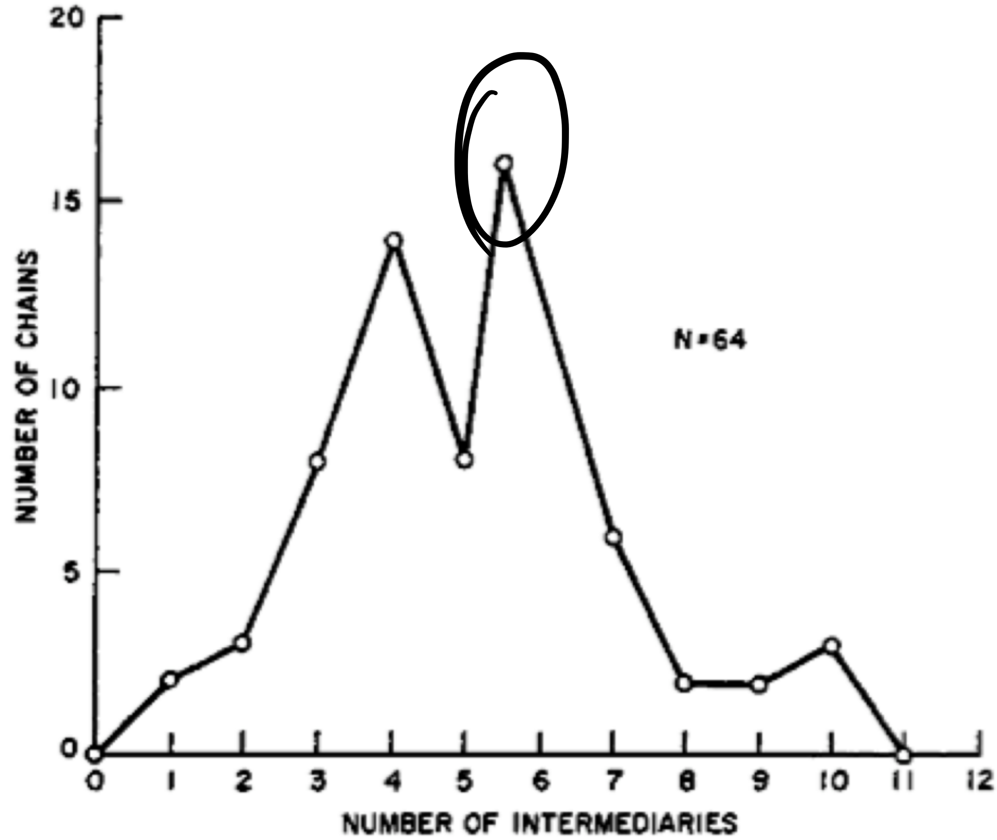


Figure 2.10: A histogram from Travers and Milgram's paper on their small-world experiment [391]. For each possible length (labeled "number of intermediaries" on the x-axis), the plot shows the number of successfully completed chains of that length. In total, 64 chains reached the target person, with a median length of six.

"six degrees of separation"

myth

Instant messaging

Leskovec & Horvitz

240 millions of nodes

median: 7

average: 6.6

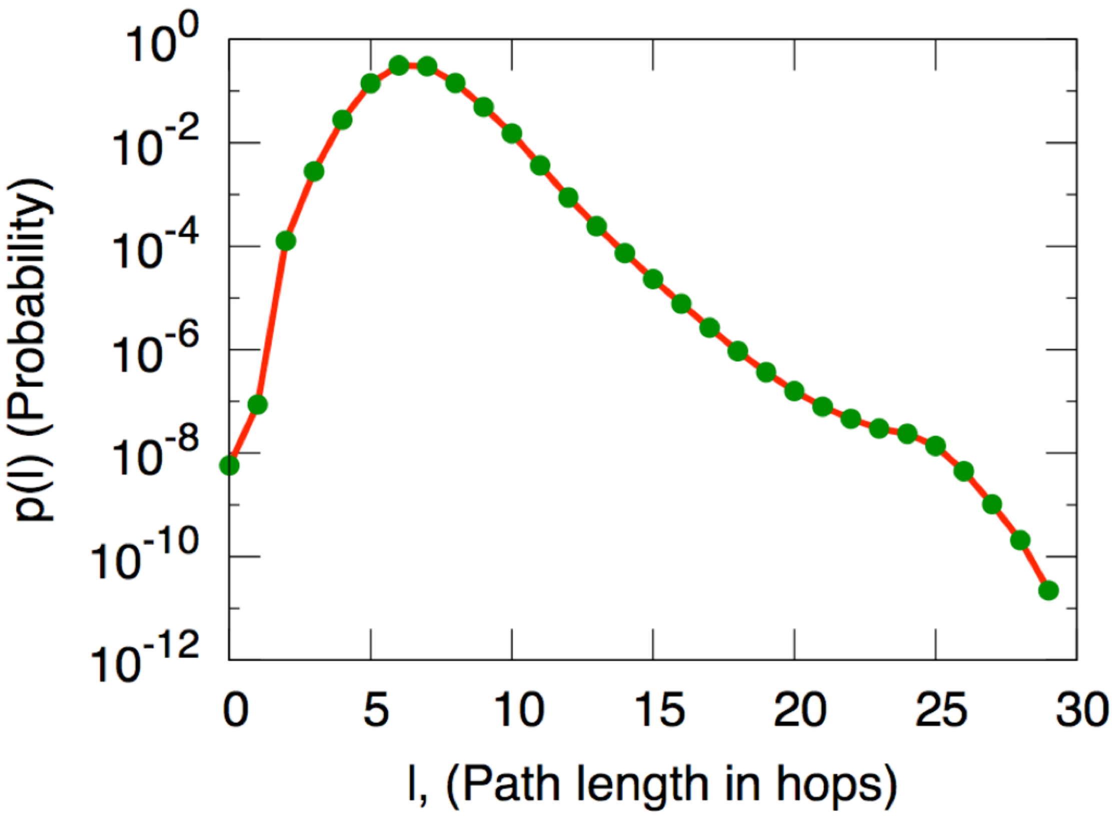


Figure 2.11: The distribution of distances in the graph of all active Microsoft Instant Messenger user accounts, with an edge joining two users if they communicated at least once during a month-long observation period [273].

Facebook

average < 4.4

iPad 10:12 85%

research.fb.com

facebook research

Research Areas | Publications | People | Academic Programs | Downloads & Projects | Careers | Blog

June 22, 2012

Four Degrees of Separation

ACM Web Science Conference (WebSci)

By: Lars Backstrom, Paolo Boldi, Marco Rosa, Johan Ugander, Sebastiano Vigna

Abstract

Frigyes Karinthy, in his 1929 short story “Lancszemek” (in English, “Chains”) suggested that any two persons are distanced by at most six friendship links. Stanley Milgram in his famous experiments challenged people to route postcards to a fixed recipient by passing them only through direct acquaintances. Milgram found that the average number of intermediaries on the path of the postcards lay between 4.4 and 5.7, depending on the sample of people chosen.

We report the results of the first world-scale social-network graph-distance computations, using the entire Facebook network of active users (721 million users, 69 billion friendship links). The average distance we

Erdös Number

Paul Erdös

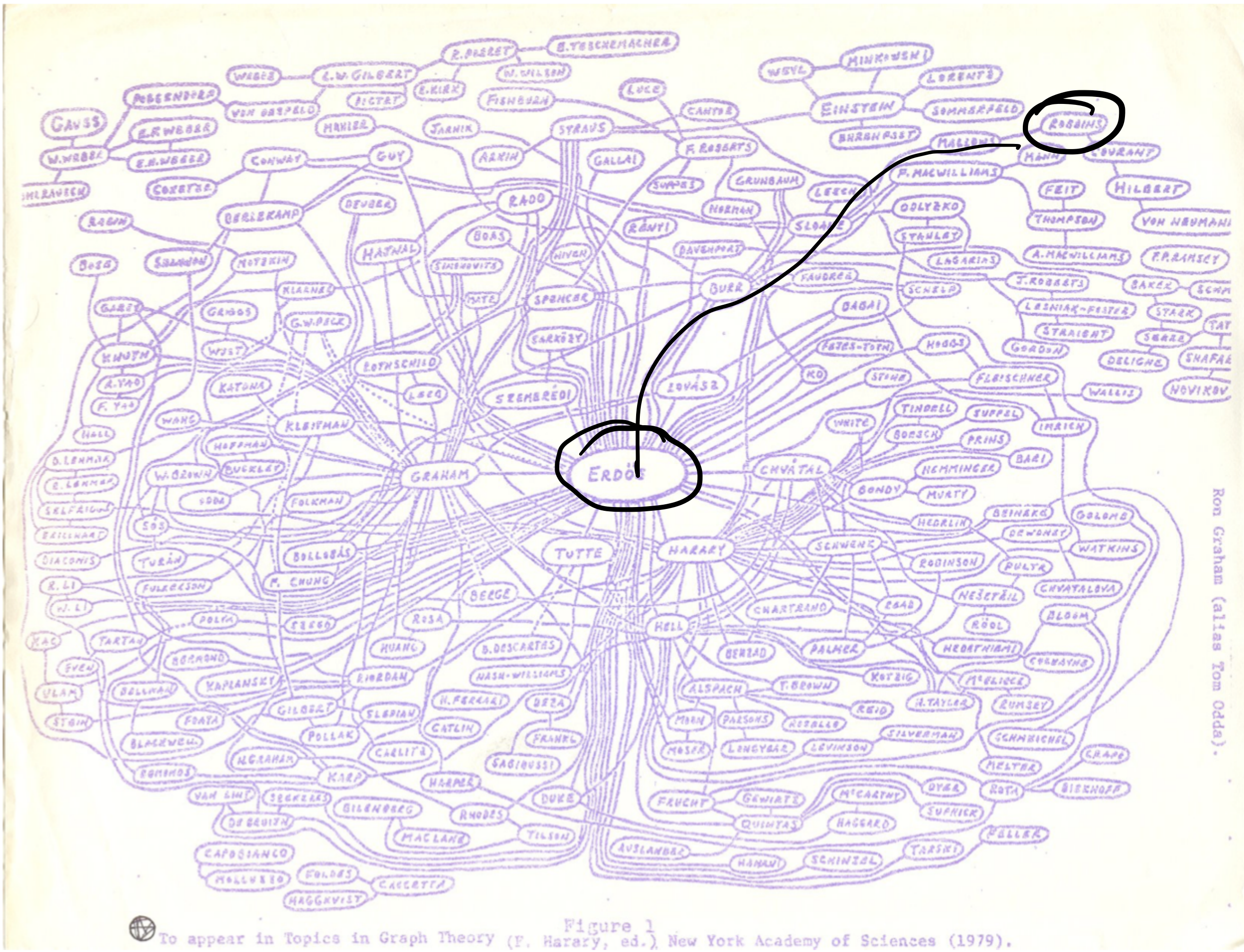


Figure 2.12: Ron Graham's hand-drawn picture of a part of the mathematics collaboration graph, centered on Paul Erdös [189]. (Image from http://www.oakland.edu/enp/cgraph.jpg)

Oracle of Bacon

Not only Kevin Bacon!

Is there a "center" in the universe of Social Networks?

Apperently every node in a social network is both at the center and at the boundaries...

However "centralities" can still say something...

Network Data Sets: an Overview

read

in the

book

many aplications

and domains ...

Take Home Message

REASONS TO STUDY A NETWORK

1. to find <u>structures</u> in your particular domain to detect hidden patterns in your data: you can understand relationships between elements at a wider spectrum.

2. to find <u>universal laws</u> with possible explanations that are not directly tied to the specifics of the domain itself.

□