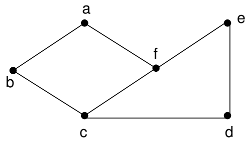
CSI 445/660 – Part 1
(Graph Theory Basics)

Ref: Chapter 2 of [Easley & Kleinberg].

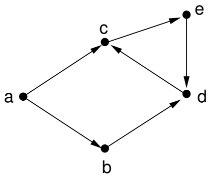
Types of Graphs

- **Undirected** and **Directed**.



Undirected graph:

- **Example:** Friendship relation among people.
- A **symmetric** relationship.

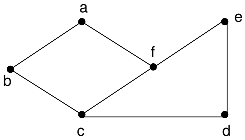


Directed graph:

- **Example:** Follower relationship in Twitter.
- May not be symmetric.

Undirected Graphs: Notation and Definitions

Example:



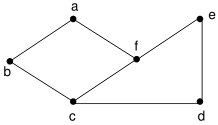
Notation: $G(V, E)$

$$V = \{\text{a, b, c, d, e, f}\} \quad (\textbf{nodes} \text{ or } \textbf{vertices})$$

$$E = \{ \{\text{a,b}\}, \{\text{a,f}\}, \{\text{b,c}\}, \{\text{c,d}\}, \{\text{c,f}\}, \{\text{d,e}\}, \{\text{e,f}\} \} \quad (\textbf{edges})$$

$|V| =$ No. of nodes $= 6$ $\qquad$ $|E| =$ No. of edges $= 7$

Notation and Definitions (continued)



Definition: The **degree** of a node v is the number of edges **incident on** v.

Example: Degree of a = 2, degree of f = 3.

Some observations:

- Sum of the degrees of all the nodes
  $$= \text{Degree(a)} + \text{Degree(b)} + \ldots + \text{Degree(f)}$$
  $$= 2 + 2 + 3 + 2 + 2 + 3 = 14 \ (\textbf{even})$$
  $$= 2 \times \text{No. of edges.}$$

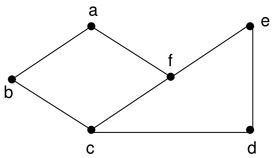
- Nodes with **odd** degree = {c, f}; thus, the number of nodes of odd degree is **even**.

Notation and Definitions (continued)

Theorem: **[First Theorem of Graph Theory]**

In any undirected graph, **the sum of the degrees of all the nodes is equal to twice the number of edges**.

Corollary: In any undirected graph, the **number of nodes of odd degree is even**.

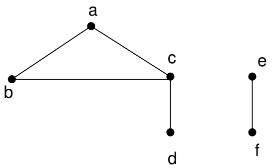


Examples of paths in graph G:

- **a – f – e – d**
- **a – b – c – f – e – d**

- There is a path between every pair of nodes.
- Graph G is **connected**.

Notation and Definitions (continued)

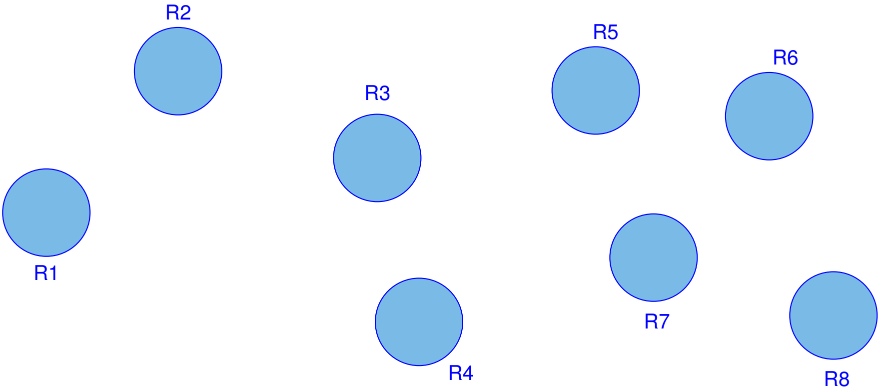


- **Disconnected** graph.
- Has two **connected components**.

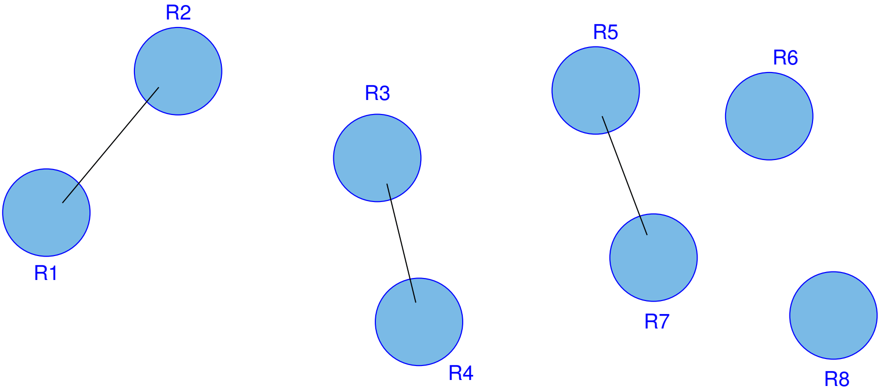
Evolution of a large social network: Imagine the following global friendship graph.

- One node per person in the world
(No. of nodes $\approx$ 7.3 billion).
- An edge between each pair of friends.

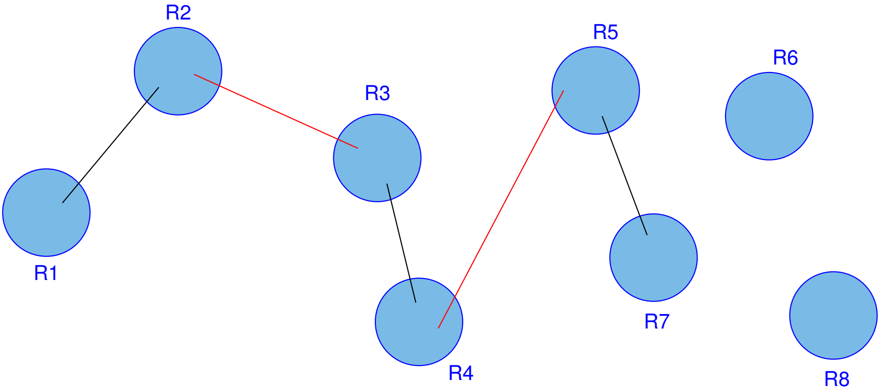
Friendship Network Evolution



Friendship Network Evolution (continued)



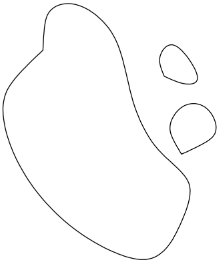
Friendship Network Evolution (continued)



Friendship Network Evolution (continued)

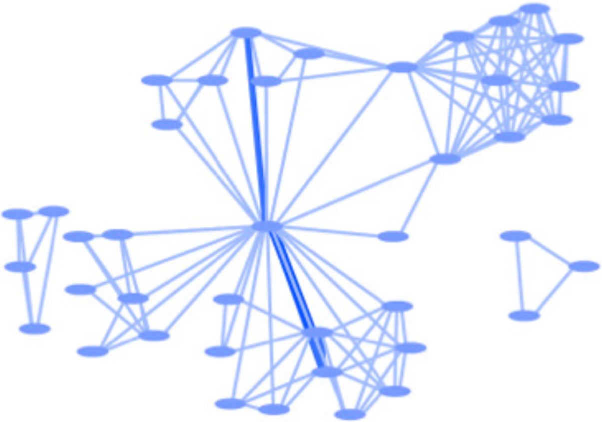
- Components get **merged** over time.
- The graph is likely to contain paths between people in remote parts of the world.
- A large subset of the nodes are in one component, called the **giant component**. (This is typical of many social networks arising in practice.)

An Illustration by Prof. Alistair Sinclair (UC Berkeley):

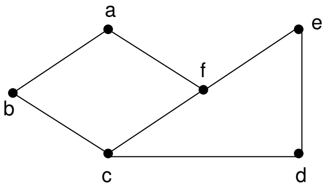


Giant Component: Another Example

Collaboration graph at a research center (from [EK]):



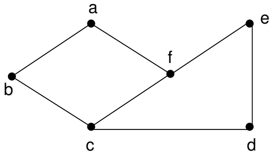
Shortest Paths

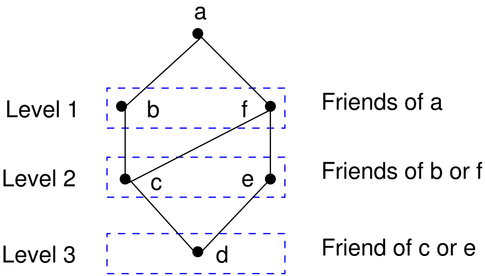


Paths between a **and** e:

- **a – f – e** : Length = 2 (No. of edges)
- **a – b – c – f – e** : Length = 4
- There is no path between a and e with length < 2.
- So, **a – f – e** is a **shortest path** between a and e.
- Shortest paths can be found using a procedure called **breadth-first-search** (BFS).

Breadth-First-Search: Example I



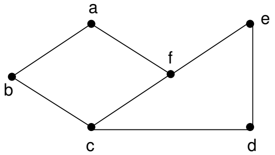


Observation: Each node is within a distance of 3 from node a.

Breadth-First-Search: Example II

Observation: Each node is within a distance of 2 from node f.

Definition of Diameter



Shortest Path Lengths: (Partial list)

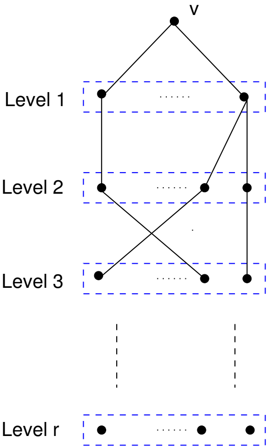
Node pair	**Shortest Distance**
a, b	1
a, c	2
a, d	**3**
⋮	⋮
b, e	**3**
⋮	⋮
e, f	1

- **Diameter**: Maximum among the shortest path lengths.
- Diameter of the above graph $=$ 3.

Some Notes About Diameter

- Diameter is meaningful only for **connected** graphs. (Some references use ∞ as the diameter of a disconnected graph.)
- If a graph is disconnected, one needs to consider the diameter each connected component.
- For a connected graph with n nodes, the diameter is at most $n - 1$.
- In communication networks, diameter gives an indication of the worst-case delay for message delivery.
- Typically, giant components of social networks have small diameters (**small world phenomenon**).

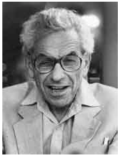
BFS and Diameter



Observation: For any **connected** graph, if a BFS produces r levels, then the diameter of the graph is at most $2r$.

Small World Phenomenon

Example: Erdős Collaboration Network



- Paul Erdős (1913 – 1996)
- Hungarian Mathematician

- Each node is a researcher and edge $\{x, y\}$ means that researchers x and y co-authored at least one paper.
- **Level 0**: Node corresponding to Erdős.
- **Level 1**: Nodes corresponding to researchers who co-authored a paper with Erdős.

Erdős Collaboration Network (continued)

- **Level 2**: Nodes corresponding to researchers who co-authored a paper with some researcher in Level 1.

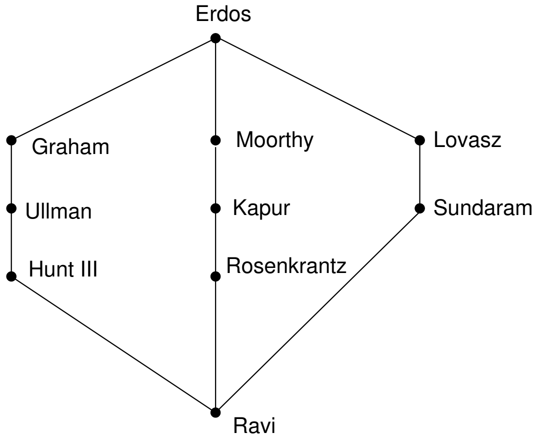
$$\vdots \qquad \vdots$$

- **Level j**: Nodes corresponding to researchers who co-authored a paper with some researcher in Level $j-1$.

- **Erdős Number** of a researcher: The level number in the graph for the node corresponding to the researcher.

Largest known Erdős Number $= 8$.

An Example for Erdős Number



- Ravi's Erdős Number ≤ 3.
- Erdős Numbers of Teri Harrison, Catherine Dumas and Dan Lamanna ≤ 4.

Small World Phenomenon

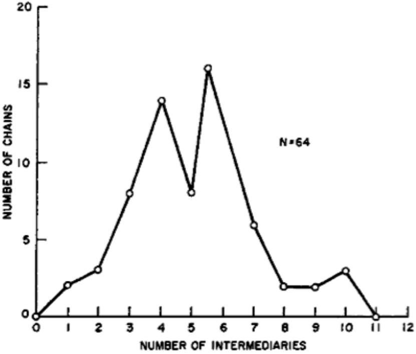
- Stanley Milgram (1933 – 1984)
- American Sociologist/Psychologist (Yale University)

Milgram's Experiment:

- Done during the 1960's. (**Budget: $680**)
- Chose 296 random starters (in Nebraska and Kansas).
- Asked each starter to forward a letter to a target person in Boston.
- **Rule:** Each person should forwarded the letter to another person whom they knew on a first name basis (to eventually reach the target as quickly as possible).

Milgram's Experiment (continued)

- 64 letters eventually reached the destination.
- Each letter that reached the destination forms a chain of people.
- Median length of the chain $= 6$ (**"six degrees of separation"**).

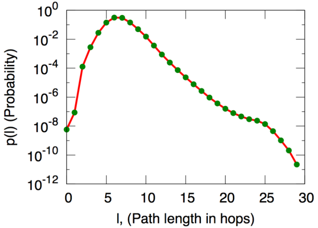


Milgram's Experiment (continued)

- The experiment suggested that social networks exhibit the **small world phenomenon**: they contain short paths between nodes (i.e., they have small diameters).
- **Kevin Bacon Game** popularized the idea.
- Milgram's work was influenced by the work of Ithiel de Sola Pool and Manfred Kochen.
- The "small world" idea also appeared in a short story by the Hungarian author Frigyes Karinthy in 1929.

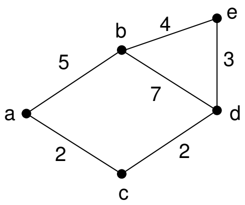
A Recent Large Scale Study

- By Eric Horovitz and Jure Leskovec [2008].
- Large social network with $\approx$ 240 million users of Internet Messenger.
- An edge in the graph indicates that the two users engaged in a two-way conversation during the observation period.
- The giant component includes almost all the nodes.
- Median path length $= 7$.



- **So far:** Distance = No. of edges.
- **More general situation:** Each edge has a non-negative "weight" (which may represent distance, time, etc.).

Example:



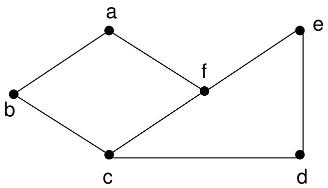
- Length of path **a – b – e** = $5 + 4 = 9$.
- Length of path **a – c – d – e** = $2 + 2 + 3 = 7$.
- So path **a – c – d – e** is shorter (even though it uses more edges).

Generalization – Edge Weights (continued)

- When all edge weights are 1, we get the previous case (i.e., unweighted graphs).
- Software for obtaining travel directions uses weighted graphs (constructed from road maps).
- With edge weights, BFS cannot be used to find shortest paths; a more sophisticated algorithm is used.
- Diameter can be defined as before (except that shortest paths are based on edge weights).

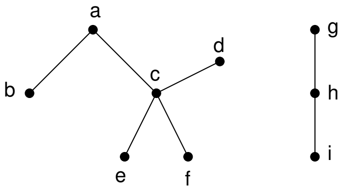
Cycles in graphs

Cycle: A path that starts and ends at the same node.



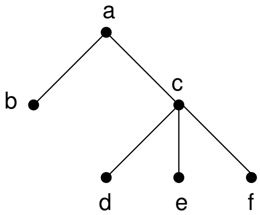
- Cycle 1: **a – b – c – f – a**.
- Cycle 2: **c – f – e – d – c**.

Acyclic graph: A graph with **no** cycles.



- Each connected component is a **tree**.
- The graph is a **forest**.

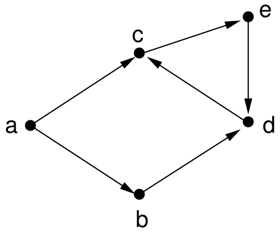
Standard Way of Displaying Trees



- Node a: **Root** of the tree.
- Nodes b, c: **Children** of the root (**siblings**).
- Nodes d, e, f: Children of node c.
- Nodes b, d, e, f: **Leaves**. (They don't have any children.)
- Note the BFS structure.

Directed Graphs: Notation and Definitions

Example:



- Edges can be traversed only in the indicated direction.

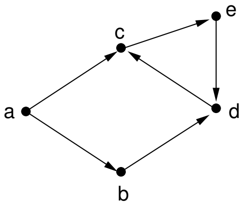
$$V \;=\; \{\text{a, b, c, d, e}\} \quad (\textbf{nodes} \text{ or } \textbf{vertices})$$

$$E \;=\; \{\text{ (a,b), (a,c), (b,d), (c,e), (d,c), (e,d) }\}$$
$$(\textbf{directed edges})$$

$|V|$ = No. of nodes = 5 $\qquad$ $|E|$ = No. of directed edges = 6

Note: Directed edges are indicated as **ordered pairs**.

Directed Graphs (continued)

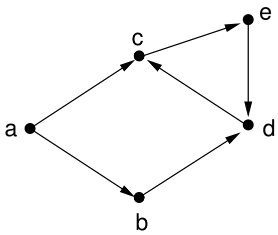


- **Outdegree** of a node v: No. of edges leaving v.
- **Indegree** of a node v: No. of edges entering v.
- **Total Degree** of a node v $=$ Outdegree(v) $+$ Indegree(v).

Example: Indegree of a $= 0$, Outdegree of a $= 2$.

Observation: Sum of the outdegrees of all the nodes $=$ Sum of the indegrees of all the nodes $=$ No. of directed edges.

Paths and Cycles in Directed Graphs



Directed paths:

- **a** $\to$ **c** $\to$ **e**: Length $= 2$.
- **a** $\to$ **b** $\to$ **d** $\to$ **c** $\to$ **e**: Length $= 4$.
- There is **no directed path** from e to a.

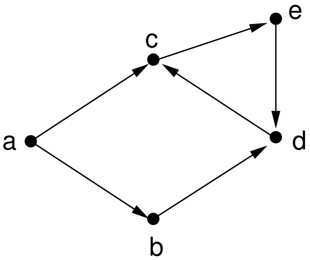
Directed cycle: **d** $\to$ **c** $\to$ **e** $\to$ **d**: Length $= 3$.

Connectivity in Directed Graphs

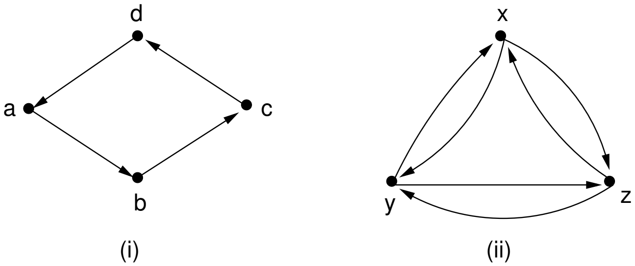
Weakly connected: Undirected graph obtained by **erasing** all edge directions is connected.

Strongly connected: There is a **directed path** from any node to any other node.

Examples:



- Weakly connected but not strongly connected. (There is **no** directed path from e to a.)



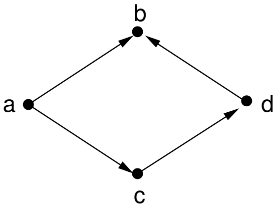
- Directed graphs (i) and (ii) are both strongly connected.

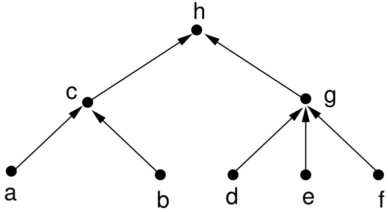
Simple Facts:

- Every strongly connected graph is also weakly connected; however, a weakly connected graph need not be strongly connected.
- Every strongly connected graph contains a directed cycle.

Directed Acyclic Graph (dag): A directed graph **without any directed cycle**.

Examples:

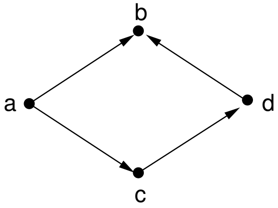




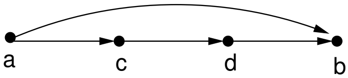
Note: The dag on the right is a model of the **hierarchy** in an organization.

Fact: The nodes of any dag can be arranged along a line so that **each directed edge goes from left to right**.

Example:



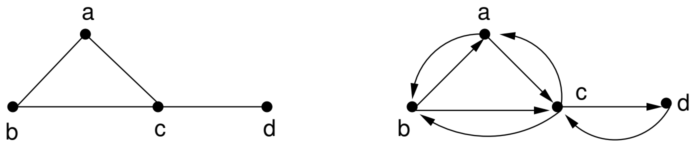
(i)



(ii)

- Such an arrangement of the nodes of a dag is called a **topological sort**.
- A topological sort of a dag can be constructed efficiently.

Representing an Undirected graph as a Directed Graph

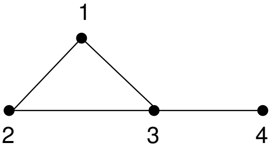


- An undirected graph can be thought of as a directed graph by replacing each undirected edge by a pair of edges in opposite directions.
- Software tools that work only with directed graphs can handle undirected graphs using this transformation.

Representing Graphs in a Computer

- Visual representation is not useful in developing algorithms.
- Two common forms: **Adjacency Matrix** and **Adjacency List**.

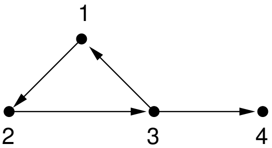
Adjacency Matrix for an Undirected Graph:



	1	2	3	4
1	0	1	1	0
2	1	0	1	0
3	1	1	0	1
4	0	0	1	0

- For an undirected graph with n nodes, the adjacency matrix has n rows and n columns.
- The entry in row i and column j is 1 if $\{i, j\}$ is an edge; the entry is 0 otherwise.
- The matrix is **symmetric**.

Adjacency Matrix for a Directed Graph:



	1	2	3	4
1	0	1	0	0
2	0	0	1	0
3	1	0	0	1
4	0	0	0	0

- For a directed graph with n nodes, the adjacency matrix has n rows and n columns.
- The entry in row i and column j is 1 if (i, j) is an edge; the entry is 0 otherwise.
- The matrix is not necessarily symmetric.

Representing Graphs ... (continued)

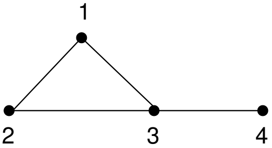
Remarks on Adjacency Matrix Representation:

- For a graph with n nodes, the memory space needed for the adjacency matrix is n^2. (This is not practical for large graphs.)
- For weighted graphs, we can store the weight of each edge in the adjacency matrix.

Adjacency List Representation:

- For each node i, list the nodes to which i has an edge (in some order).
- The size of this representation is **linear** in the number of edges of the graph.
- **Preferred** representation for large graphs.

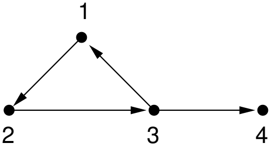
Adjacency List Representation – Undirected Graph:



Node 1: 2 3
Node 2: 1 3
Node 3: 1 2 4
Node 4: 3

Adjacency List Representation – Directed Graph:

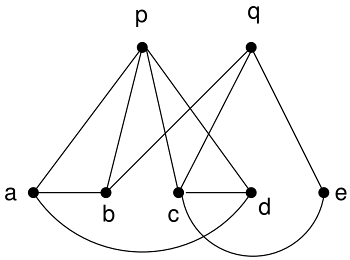
Note: List stores the **outgoing** edges for each node.



Node 1: 2
Node 2: 3
Node 3: 1 4
Node 4:

Egocentric Networks

- Also called **ego networks**.

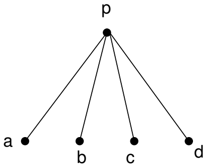


- Each node is called **ego**.
- Neighbors of a node are its **alters**.

Example: With node p as ego, its alters are a, b, c and d.

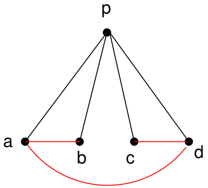
Egocentric Networks (continued)

The 1-Degree Egocentric Network of node p:



Note: This network consists of p, the alters of p and edges between p and its alters.

The 1.5-Degree Egocentric Network of node p:



Note: This network is obtained by adding the edges between the alters of p in the original graph to the 1-degree egocentric network of p.

Settings that Provide Large Network Data Sets

- Manually constructed social networks involving human interactions are small.
- Other settings provide larger data sets representing interactions (which are not necessarily through direct contact).

A. Collaboration Networks: ("Who Works With Whom")

- Co-authorship networks
  - Rich form of interaction over a long period of time (suitable for longitudinal studies).
  - Nodes with high degrees likely to represent influential scientists.
- Co-appearance in movies
- Co-membership in Board of Directors of large companies: used to explain business decisions made by companies.

B. Networks from Communication Among People:
("Who Talks to Whom")

- Internet Messenger example [Horovitz & Leskovec, 2008] discussed earlier.
- Email logs within a company: The most famous example is the Enron data set.
- Call graphs constructed from phone numbers: Privacy of individuals must be protected.

Settings that Provide ... (continued)

C. Information Linkage Graphs:

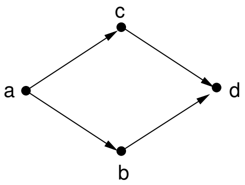
- Web data:
  - Directed graph with nearly 5 billion nodes.
  - Extremely large for effective processing using commodity hardware.
  - Researchers work with reasonable subsets (e.g. linkage among bloggers, linkage among articles of Wikipedia).
- Citation networks:
  - Useful in tracking the development of disciplines (e.g. identifying "central papers" of a discipline).
  - Also useful for longitudinal studies.

D. Technological Networks:

- Computer networks
- Power grid

E. Networks in the Natural World:

- (a) Food Web ("Who Eats Whom" relationship):



- Directed edge $x \rightarrow y$ indicates that species x eats species y.
- Important in studying **cascading extinction** of species.

Settings that Provide ... (continued)

- (b) Neural connections in the brain:
  - Nodes are neurons.
  - The interconnections among the neurons determines cognitive behavior.
- (c) Biological networks:
  - Nodes are chemical compounds that play a role the metabolic process.
  - Edges represent chemical interactions.
  - Study of such networks has applications in medicine (e.g. blocking certain interactions may help in curing certain diseases).