Networks & Graphs part 2

CS 579 Online Social Network Analysis

Dr. Cindy Hood
9/2/25

Homework Assignments

- HW #1 Assigned last week
  - Due by midnight Sept 3
- HW #2 assigned on Thursday
  - Due by midnight Monday 9/15

Exams and Final Project Poster Presentation

- Exam 1 – Oct 9 in class
- Exam 2 – Dec 2 in class

- Final Project Poster Session – Dec 4 in class

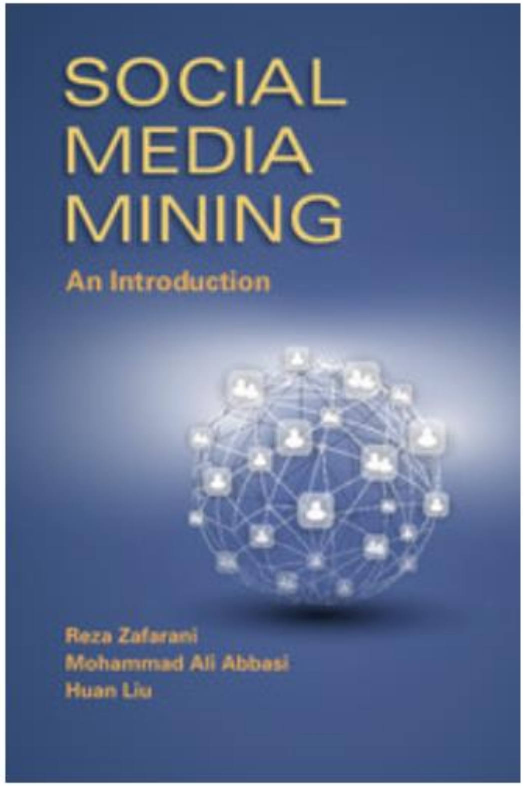
- Online students (sections 2 and 3) will have remote options

Teaching Assistants

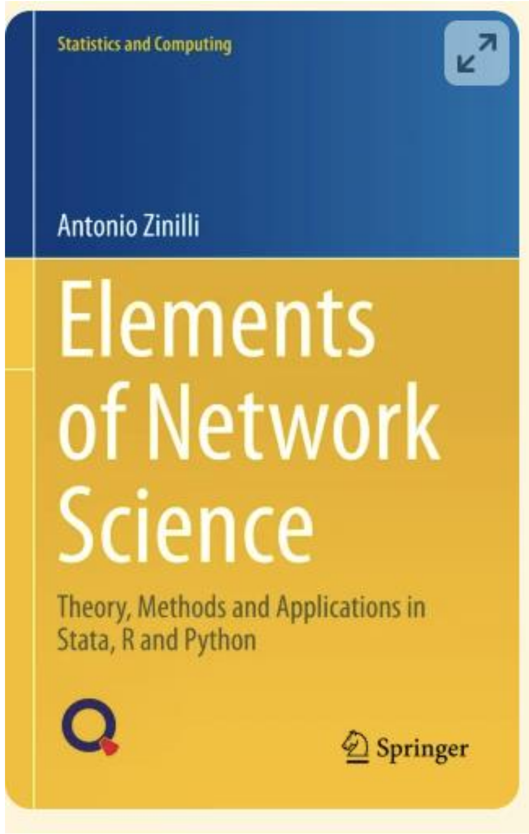
- Siva Krishna Golla
  - sgolla2@hawk.illinoistech.edu
  - Mondays 2-3pm on zoom
- Khush Dhiren Patel
  - kpatel210@hawk.illinoistech.edu
- Aswith Sama
  - asama@hawk.illinoistech.edu
  - Thursdays 3-4pm on zoom

- **Not yet officially working, waiting for authorization (US govt)**

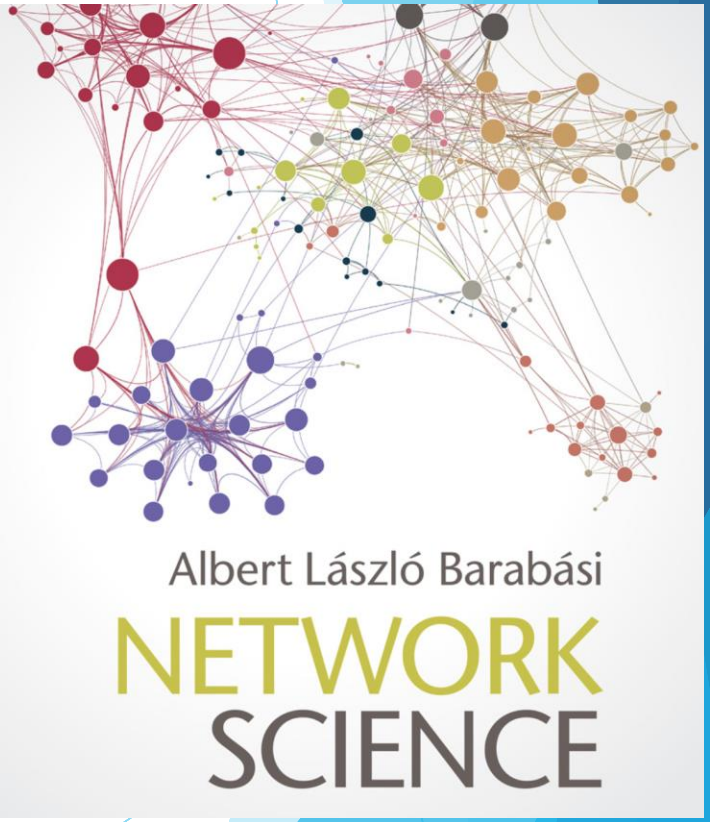
References



http://www.socialmediamining.info



https://link.springer.com/book/10.1007/978-3-031-84712-7



http://networksciencebook.com

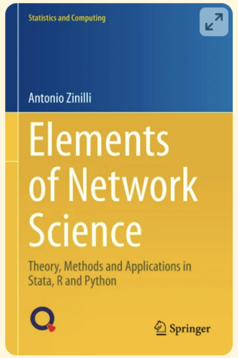
Networks are the heart of complex systems

A **complex system** is a system composed of many components that may interact with one another.[1] Examples of complex systems are Earth's global climate, organisms, the human brain, infrastructure such as power grid, transportation or communication systems, complex software and electronic systems, social and economic organizations (like cities), an ecosystem, a living cell, and, ultimately, for some authors, the entire universe.[2][3][4]

- ”In many cases, it is useful to represent such a system as a network where the nodes represent their components and the links represent their interactions”

https://en.wikipedia.org/wiki/Complex_system

Network Science



Elements of Network Science

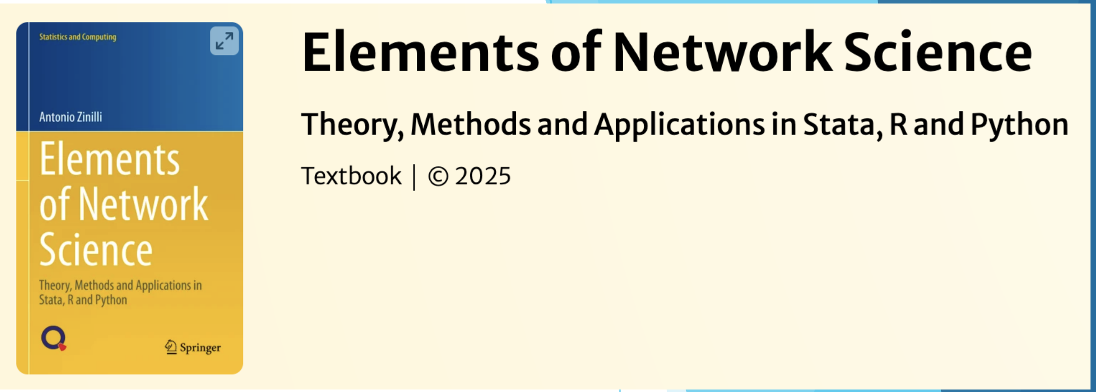
Theory, Methods and Applications in Stata, R and Python

Textbook | © 2025

- "Network Science is the study of complex systems composed of basic components known as nodes and their interactions or relationships, referred to as edges (Newman, 2003). By studying an interacting system as a whole through this mathematical abstraction, we can uncover important insights that would otherwise be overlooked if we considered the system merely as a collection of isolated units. Network Science provides a unified formalism that can be applied to describe systems from different scientific fields while still being represented as a network of interconnected elements. Indeed, using this all-inclusive approach, a variety of complex systems emerging from various disciplines have been studied."

Newman, M. E. J. (2003). The structure and function of complex networks. *SIAM Review, 45*, 167–256. https://doi.org/10.1137/S003614450342480

Network Science vs Social Network Analysis (SNA)

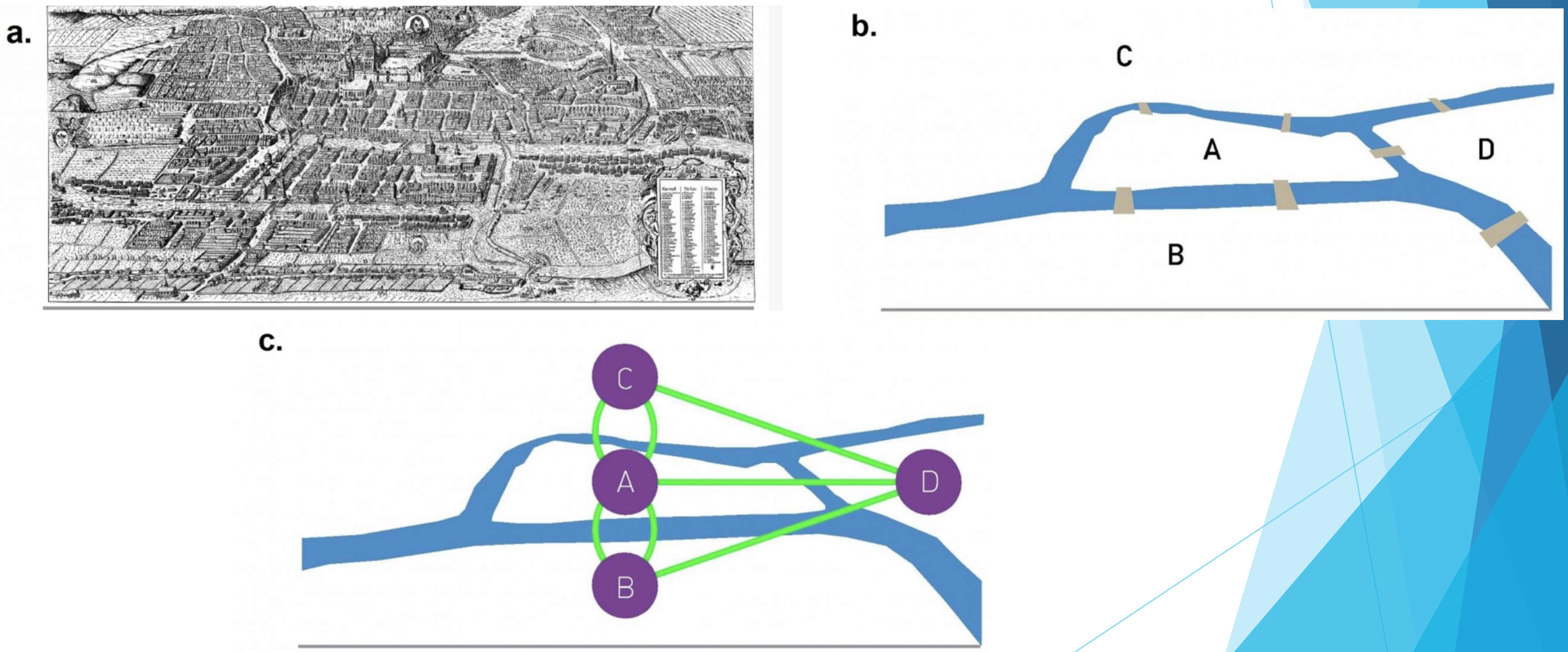


- Network Science is a highly interdisciplinary resesarch area focused on developing theoretical and practical approaches and techniques (Börner, et al,. 2007)
- Network Science, characterized by its emphasis on complex networks, encompasses a broader and more structured scope than SNA.
- It aims to identify universal structural features in diverse complex networks, understand the mechanisms underlying their formation, and explore the relationship between features and network processes.
- The field seeks models that are theoretically grounded and capable of replicating the network’s global or systemic characteristics by revealing insights into the data generation process.
- Network Science is broader than Social Network Analysis

Börner, K., Sanyal, S., & Vespignani, A. (2007). Network science. *Annual Review of Information Science and Technology, 41*(1), 537–607.

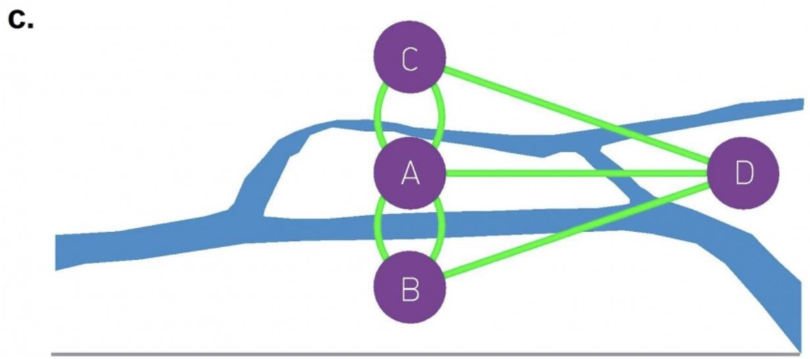
How to represent networks?

- Leonard Euler used a graph to represent Bridges of Konigsberg, Eastern Prussia 1735



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

Graph representation



- Euler offered a rigorous mathematical proof that such path does not exist.
  - He made a simple observation: if there is a path crossing all bridges, but never the same bridge twice, then nodes with odd number of links must be either the starting or the end point of this path. Indeed, if you arrive to a node with an odd number of links, you may find yourself having no unused link for you to leave it.
  - A walking path that goes through all bridges can have only one starting and one end point. Thus such a path cannot exist on a graph that has more than two nodes with an odd number of links. The Königsberg graph had four nodes with an odd number of links, A, B, C, and D, so no path could satisfy the problem.

Birth of Graph Theory

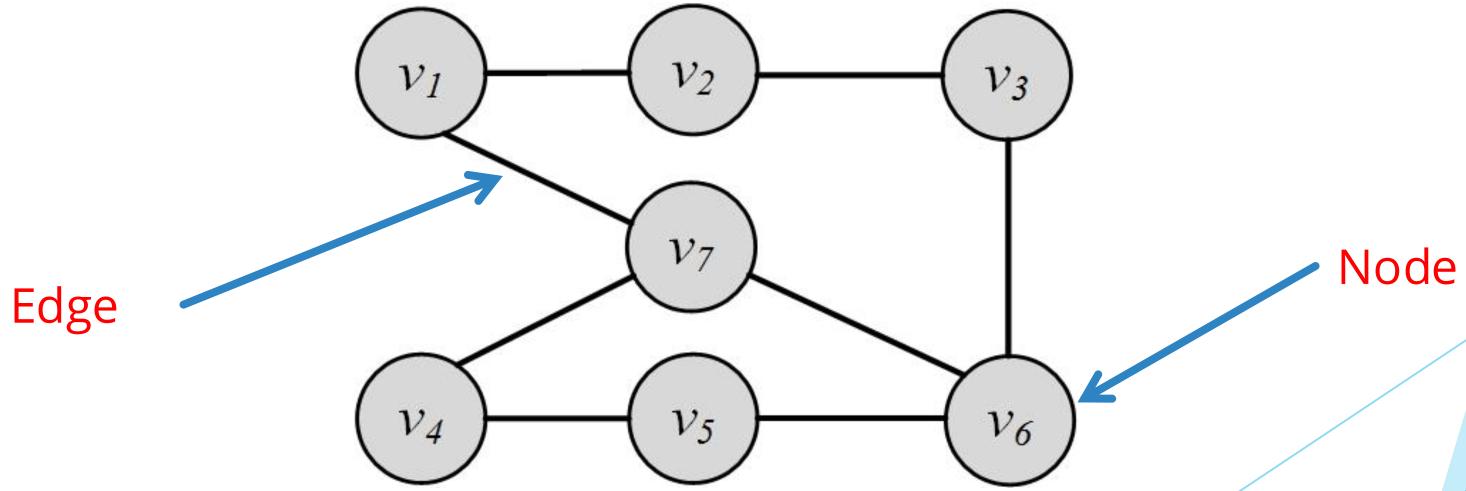
- Euler’s proof was the first time someone solved a mathematical problem using a graph.
- Two key takeaways
  - Some problems become simpler and more tractable if they are represented as a graph.
  - The existence of the path does not depend on our ingenuity to find it. Rather, it is a property of the graph.
    - Networks have properties encoded in their structure that limit or enhance their behavior.

Graph Basics

Nodes and Edges

A network is a graph, or a collection of points connected by lines

- Points are referred to as **nodes, actors,** or **vertices** (plural of **vertex**)
- Connections are referred to as **edges**, **ties or links**



Nodes or Actors

- In a friendship social graph, nodes are people and any pair of people connected denotes the friendship between them
- Depending on the context, these nodes are called nodes, or actors
  - In a web graph, "*nodes*" represent sites and the connection between nodes indicates web-links between them
  - In a social setting, these nodes are called actors
  - The size of the graph is $V = \{v_1, v_2, \ldots, v_n\}$

$$|V| = \mathbf{n}$$

Edges

- Edges connect nodes and are also known as **ties, links** or **relationships**

- In a social setting, where nodes represent social entities such as people, edges indicate internode relationships and are therefore known as relationships or (social) ties

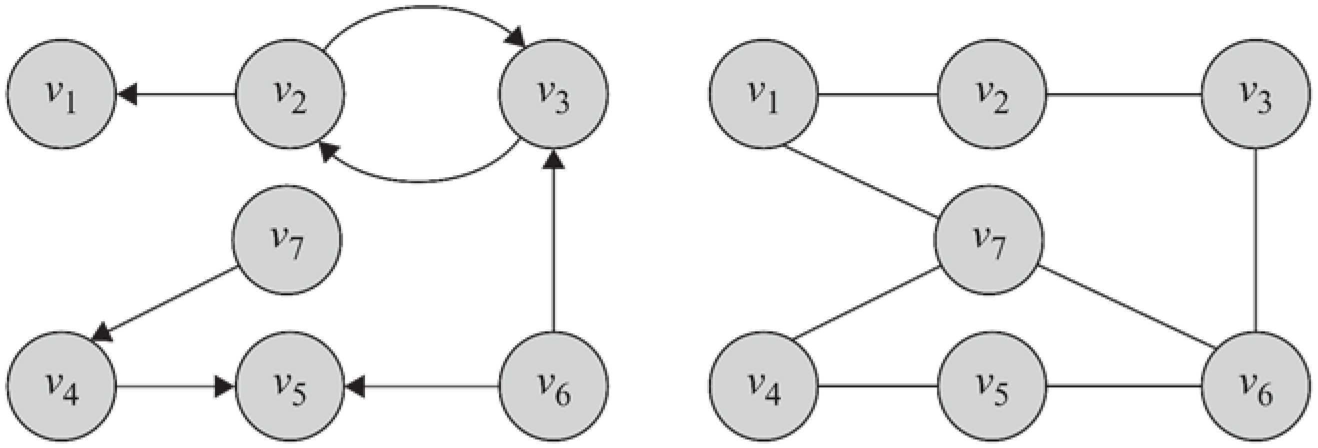
$$E = \{e_1, e_2, \ldots, e_m\}$$

- Number is edges (size of the edge-set) is denoted as

$$|E| = \mathbf{m}$$

Directed Edges and Directed Graphs

- Edges can have directions. A directed edge is sometimes called an arc



(a) Directed Graph (b) Undirected Graph

- Edges are represented using their end-points $e(v_2, v_1)$.
- In undirected graphs both representations are the same

Neighborhood and Degree (In-degree, out-degree)

For any node v, in an undirected graph, the set of nodes it is connected to via an edge is called its neighborhood and is represented as $N(v)$

- *In directed graphs we have incoming neighbors $N_{in}(v)$ (nodes that connect to v) and outgoing neighbors $N_{out}(v)$.*

The number of edges connected to one node is the degree of that node (the size of its neighborhood)

- Degree of a node i is usually presented using notation d_i

In Directed graphs:

d_i^{in}
- In-degrees is the number of edges pointing towards a node

d_i^{out}
- Out-degree is the number of edges pointing away from a node

Degree and Degree Distribution

- **Theorem 1.** The summation of degrees in an undirected graph is twice the number of edges

$$\sum_i d_i = 2|E|$$

- **Lemma 1.** The number of nodes with odd degree is even
- **Lemma 2.** In any directed graph, the summation of in-degrees is equal to the summation of out-degrees,

$$\sum_i d_i^{out} = \sum_j d_j^{in}$$

Degree Distribution

When dealing with very large graphs, how nodes' degrees are distributed is an important concept to analyze and is called ***Degree Distribution***

$$\pi(d) = \{d_1, d_2, \ldots, d_n\}$$

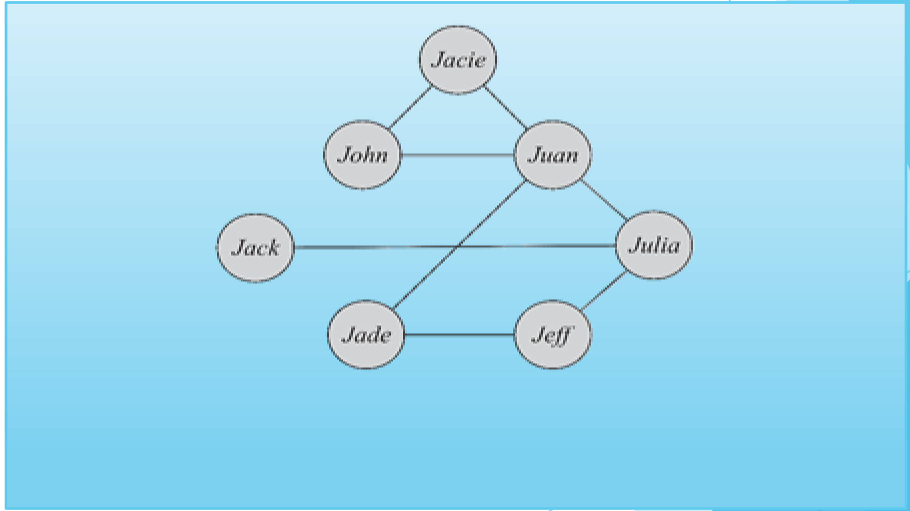
(Degree sequence)

$$p_d = \frac{n_d}{n}$$

Probability of a node with degree d

n_d is the number of nodes with degree d

$$\sum_{d=0}^{\infty} p_d = 1$$



$$p_1 = \frac{1}{7}, p_2 = \frac{4}{7}, p_3 = \frac{1}{7}, p_4 = \frac{1}{7}$$

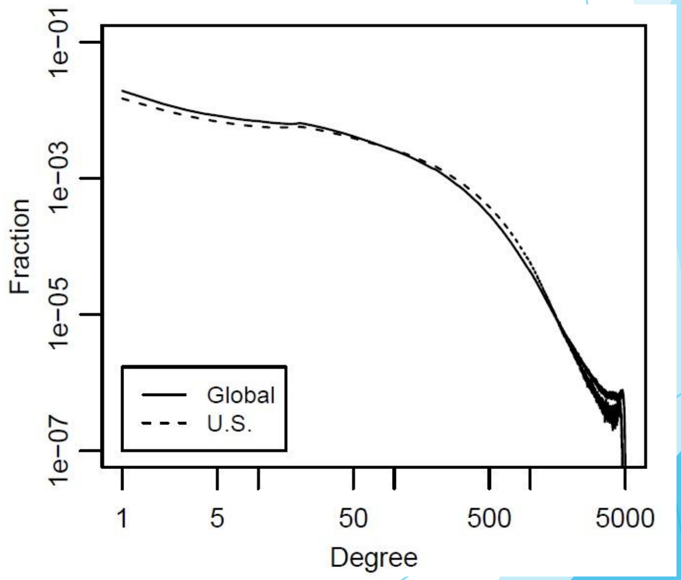
Degree Distribution Plot

The x-axis represents the degree and the y-axis represents the fraction of nodes having that degree

- On social networking sites

  There exist many users with few connections and there exist a handful of users with very large numbers of friends.

  (**Power-law degree distribution)**



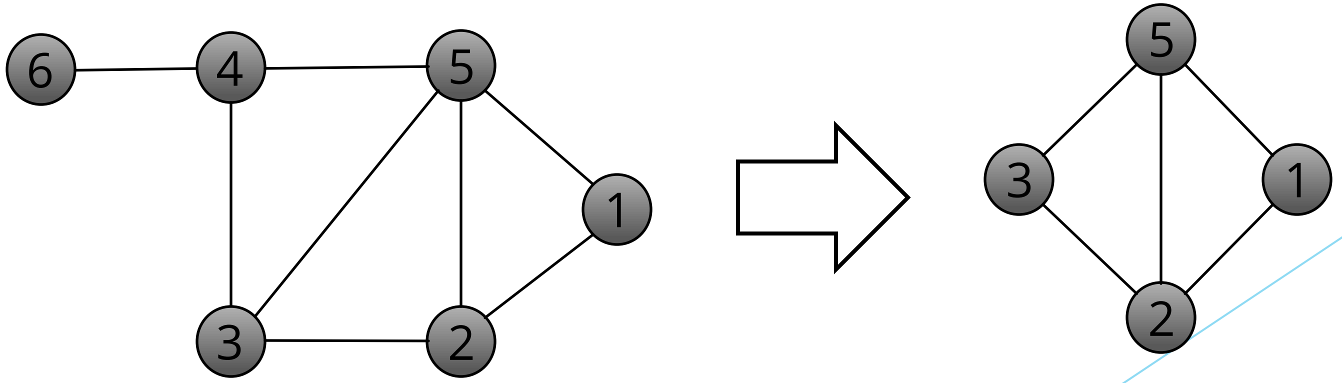
Facebook Degree Distribution

Subgraph

- Graph G can be represented as a pair $G(V, E)$ where V is the node set and E is the edge set

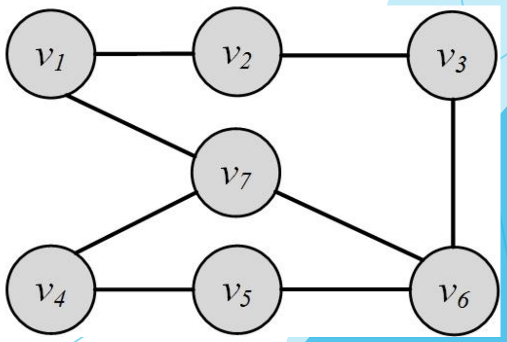
- $G'(V', E')$ is a subgraph of $G(V, E)$

$$V' \subseteq V$$
$$E' \subseteq (V' \times V') \cap E$$



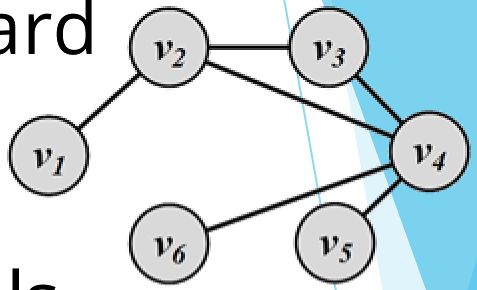
Graph Representation

- Adjacency Matrix
- Adjacency List
- Edge List



Graph Representation

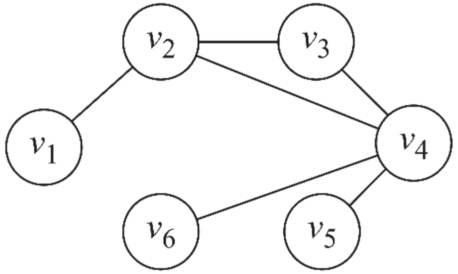
- Graph representation is straightforward and intuitive, but it cannot be effectively manipulated using mathematical and computational tools



- We are seeking representations that can store these two sets in a way such that
  - Does not lose information
  - Can be manipulated easily by computers
  - Can have mathematical methods applied easily

Adjacency Matrix (a.k.a. sociomatrix)

$$A_{ij} = \begin{cases} 1, \text{ if there is an edge between nodes } v_i \text{ and } v_j \\ 0, \text{ otherwise} \end{cases}$$



(a) Graph

	v_1	v_2	v_3	v_4	v_5	v_6
v_1	0	1	0	0	0	0
v_2	1	0	1	1	0	0
v_3	0	1	0	1	0	0
v_4	0	1	1	0	1	1
v_5	0	0	0	1	0	0
v_6	0	0	0	1	0	0

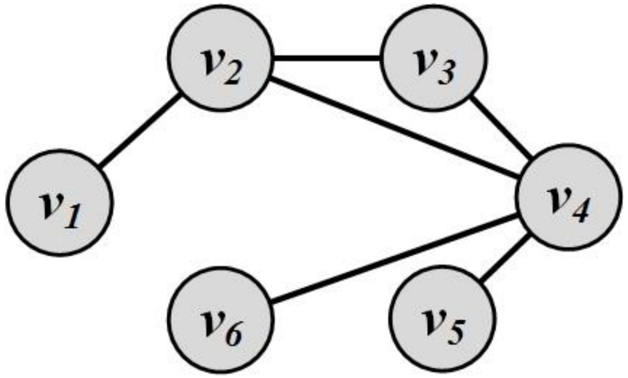
(b) Adjacency Matrix

Diagonal Entries are self-links or loops

Social media networks have very sparse Adjacency matrices

Adjacency List

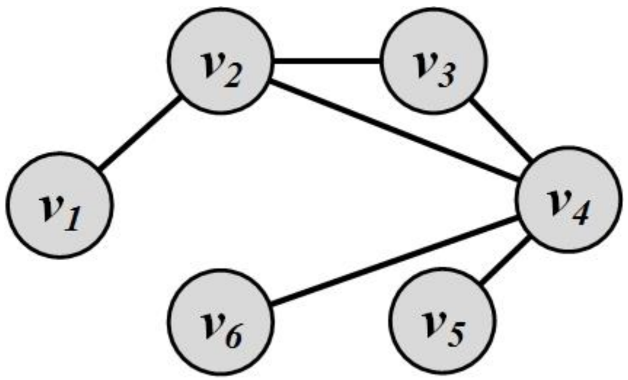
- In an adjacency list for every node, we maintain a list of all the nodes that it is connected to
- The list is usually sorted based on the node order or other preferences



Node	Connected To
v_1	v_2
v_2	v_1, v_3, v_4
v_3	v_2, v_4
v_4	v_2, v_3, v_5, v_6
v_5	v_4
v_6	v_4

Edge List

- In this representation, each element is an edge and is usually represented as (u, v), denoting that node u is connected to node v via an edge



(v_1,v_2)
(v_2,v_3)
(v_2,v_4)
(v_3,v_4)
(v_4,v_5)
(v_4,v_6)

Network	Nodes	Links	Directed / Undirected	N	L	‹K›
Internet	Routers	Internet connections	Undirected	192,244	609,066	6.34
WWW	Webpages	Links	Directed	325,729	1,497,134	4.60
Power Grid	Power plants, transformers	Cables	Undirected	4,941	6,594	2.67
Mobile-Phone Calls	Subscribers	Calls	Directed	36,595	91,826	2.51
Email	Email addresses	Emails	Directed	57,194	103,731	1.81
Science Collaboration	Scientists	Co-authorships	Undirected	23,133	93,437	8.08
Actor Network	Actors	Co-acting	Undirected	702,388	29,397,908	83.71
Citation Network	Papers	Citations	Directed	449,673	4,689,479	10.43
E. Coli Metabolism	Metabolites	Chemical reactions	Directed	1,039	5,802	5.58
Protein Interactions	Proteins	Binding interactions	Undirected	2,018	2,930	2.90

Table 2.1

N=nodes
L=links
K=average degree

Types of Graphs

- **Null, Empty, Directed/Undirected/Mixed, Simple/Multigraph, Weighted, Signed Graph, Webgraph**

Null Graph and Empty Graph

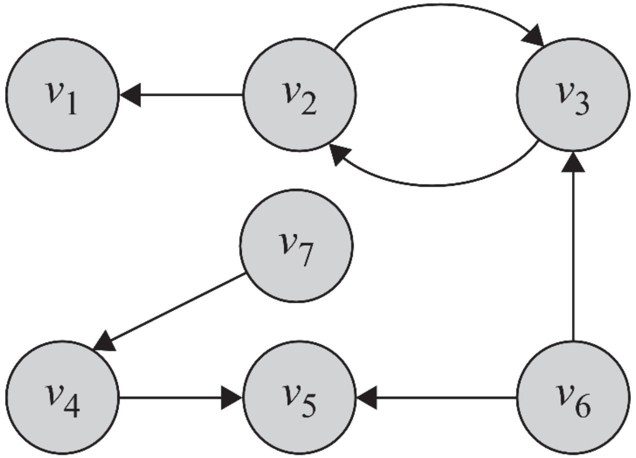
- A **null graph** is one where the node set is empty (there are no nodes)
  - Since there are no nodes, there are also no edges

$$G(V, E), V = E = \emptyset$$

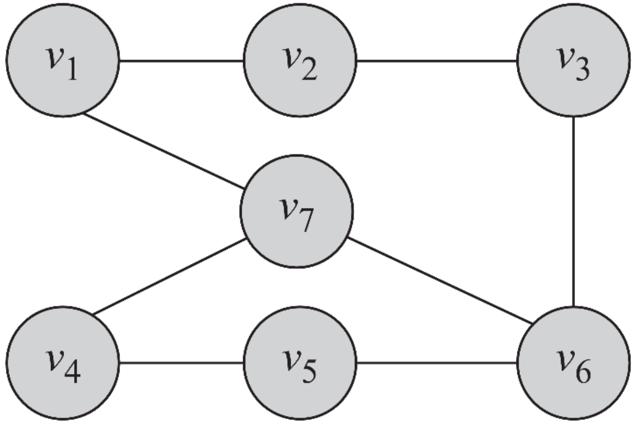
- An **empty graph** or **edge-less graph** is one where the edge set is empty, $E = \emptyset$

- The node set can be non-empty.
  - A null-graph is an empty graph.

Directed/Undirected/Mixed Graphs



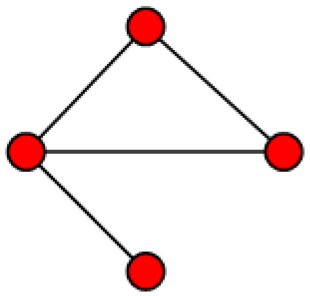
- The adjacency matrix for directed graphs is often not symmetric ($\boldsymbol{A} \neq \boldsymbol{A}^T$)
  - $\boldsymbol{A_{ij}} \neq \boldsymbol{Aji}$
  - We can have equality though



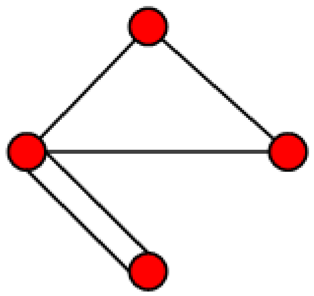
The adjacency matrix for undirected graphs is symmetric ($\boldsymbol{A} = \boldsymbol{A}^T$)

Simple Graphs and Multigraphs

- Simple graphs are graphs where only a single edge can be between any pair of nodes
- Multigraphs are graphs where you can have multiple edges between two nodes and loops



Simple graph



Multigraph

- The adjacency matrix for multigraphs can include numbers larger than one, indicating multiple edges between nodes