BIG DATA STATISTICS FOR BUSINESS

AY 2023-24

Giovanni De Luca
Parthenope University of Naples

1

Introduction

- The basic idea of a social network is very simple.
- A social network is a set of actors (or points, or nodes, or vertices) that may have relationships (or edges, or ties) with one another.
- Networks can have few or many actors, and one or more kinds of relations between pairs of actors (e.g. students in a classroom might like or dislike each other, they might play together or not, they might share food or not, ecc.).
- Network data are defined by actors and by relations (or nodes and ties, etc.).
- Network analysis focuses on the relations among actors.

2

Graphical techniques

- For the calculation of indexes describing networks, it is most useful to record information as matrices.
- For visualizing patterns, graphs are often useful.
- One reason for using mathematical and graphical techniques in social network analysis is to represent the descriptions of networks compactly and systematically.

3

Network data

- «Conventional» data consists of a rectangular matrix.
- The rows of the array are the cases, or subjects, or observations. The columns consist of (numerical or categorical) variables.
- «Network» data (in their purest form) consist of a square matrix of measurements.
- The rows of the matrix are the subjects (or actors or observations).
- The columns of the matrix are the same set of subjects.
- Each cell of the matrix describes a relationship between the subjects.

4

Network data

- The major difference between conventional and network data is that conventional data focuses on actors and variables, while network data focuses on actors and relations.

5

Example

- Suppose we were describing the structure of close friendship in a group of four people: Bob, Carol, Ted, and Alice.
- We could describe this pattern of liking ties with an actor-by-actor matrix where the rows represent choices by each actor (network matrix).
- We will put in a "1" if an actor likes another and a "0" if he doesn't.

6

Example

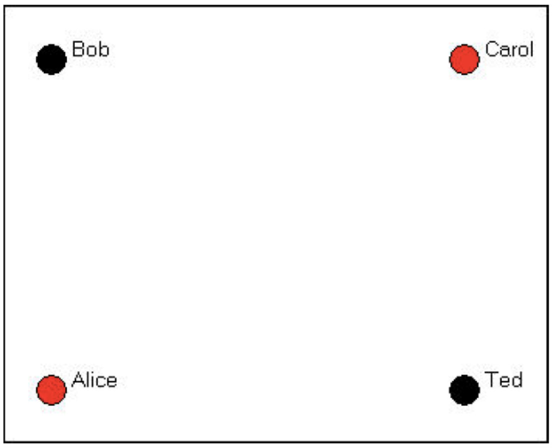
	Bob	Carol	Ted	Alice
Bob	---	1	1	0
Carol	0	---	1	0
Ted	1	1	---	1
Alice	0	0	1	---

7

Graphs

- Network analysis uses (primarily) one kind of graphic display that consists of points (or nodes) to represent actors and lines (or edges) to represent ties or relations.
- These graphic displays are known as «undirected graphs» (undirected network) or «directed graphs» (directed network) according to the indication of the relations.

8

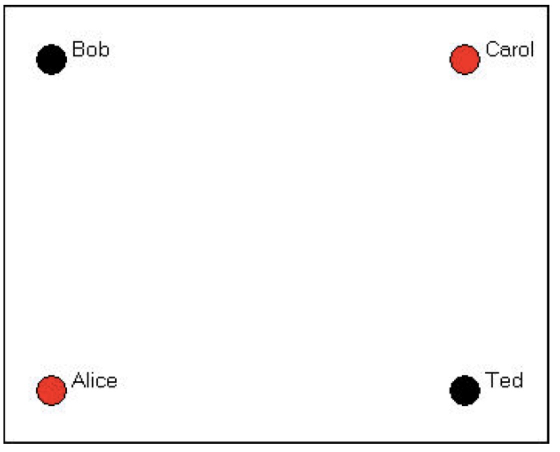


9

	Bob	Carol	Ted	Alice
Bob	---	1	1	0
Carol	0	---	1	0
Ted	1	1	---	1
Alice	0	0	1	---

10

Undirected network

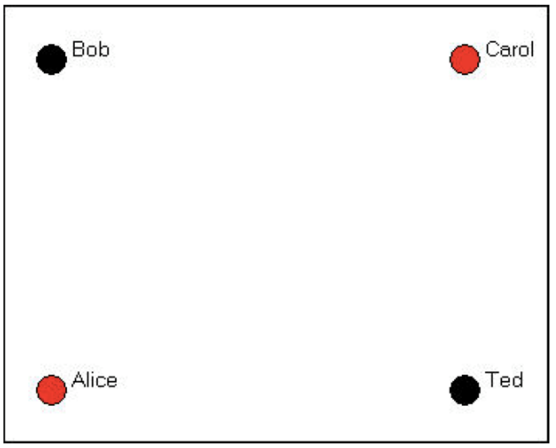


11

	Bob	Carol	Ted	Alice
Bob	---	1	1	0
Carol	0	---	1	0
Ted	1	1	---	1
Alice	0	0	1	---

12

Directed network



13

Popularity of the nodes

- If we sum the elements of the column vectors in this example, I would be measuring how "popular" each node was (in terms of how often they were the target of a directed friendship tie).

14

	Bob	Carol	Ted	Alice
Bob	---	1	1	0
Carol	0	---	1	0
Ted	1	1	---	1
Alice	0	0	1	---

15

Rearranging the matrix

- It is also helpful, sometimes, to rearrange the rows and columns of a matrix so that we can see patterns more clearly.
- Let's rearrange (permute) the matrix so that the two males and the two females are adjacent in the matrix.

16

	Bob	Ted	Carol	Alice
Bob	---	1	1	0
Ted	1	---	1	1
Carol	0	1	---	0
Alice	0	1	0	---

17

- If we calculate the proportion of all ties within a block that are present, we can create a block density matrix.

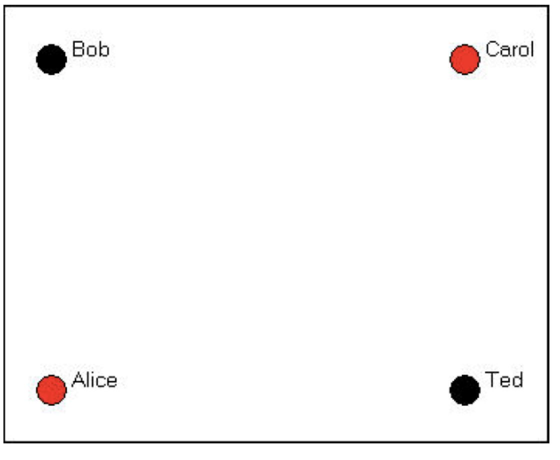
	Male	Female
Male	1.00	0.75
Female	0.50	0.00

18

Network measures
Degree

- In a network analysis, the **degree** of a node is the number of all its connections.

19

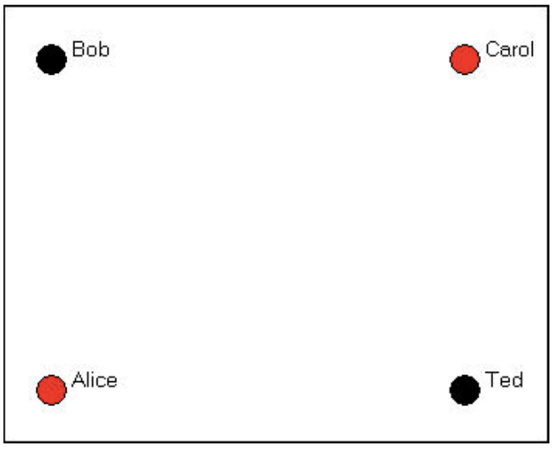


20

Network measures
Degree

- In directed network, we can distinguish between in-degree and out-degree: in-degree is a count of the number of ties directed to the node, and out-degree is the number of ties that the node directs to others.

21

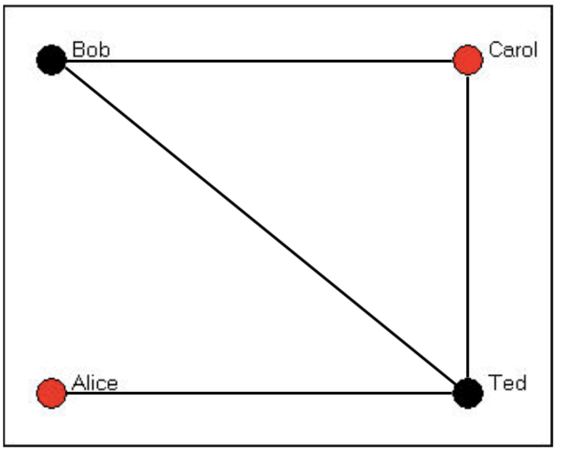


22

Network measures
Diameter

- In network analysis the **diameter** of a network is the shortest distance between the two most distant nodes in the network.
- In other words, once the shortest path length from every node to all other nodes is calculated, the diameter is the longest of all the calculated path lengths.

23

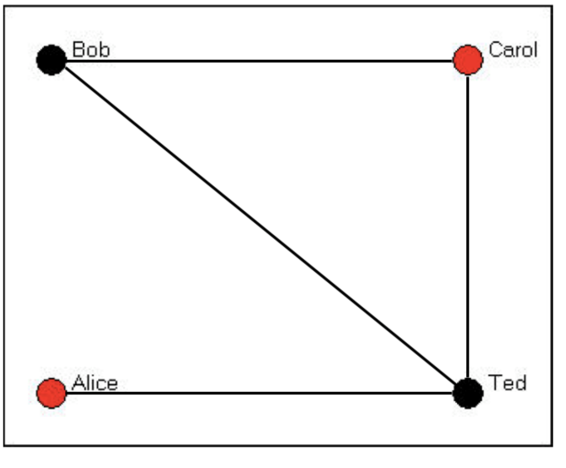


24

Network measures
Average distance

- In network analysis the **average distance** calculates the average path length in a graph, by calculating the shortest paths between all pairs of vertices.

25

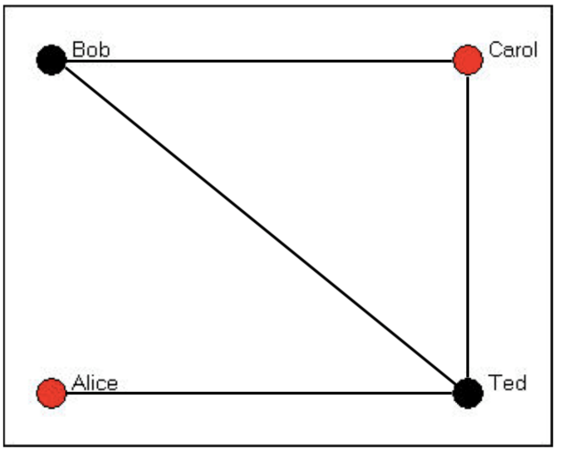


26

Network measures
Density

- In network analysis the density is the ratio of the realized edges and the possible edges.

27

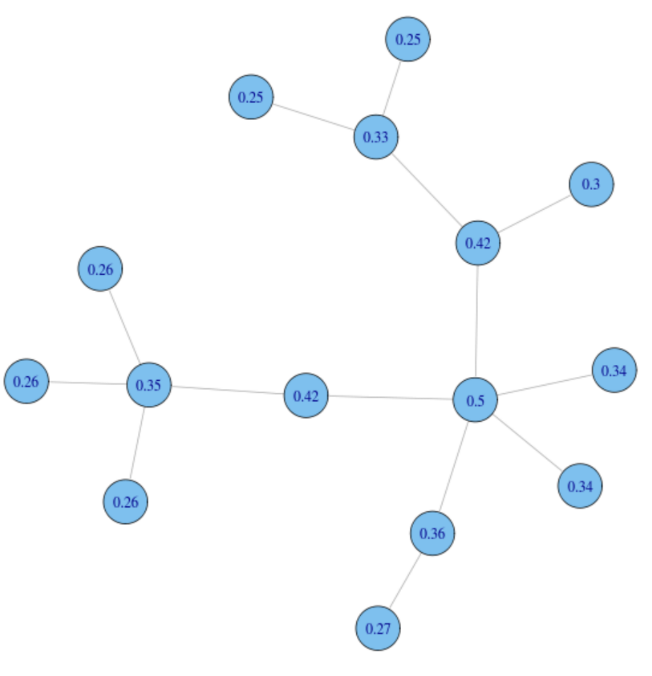


28

Network measures
Closeness

- In network analysis with N nodes, **closeness** centrality (or closeness) of a node is a measure of centrality, calculated as (N-1) divided by the sum of the length of the shortest paths between the node and all other nodes in the graph.
- Thus, the more central a node is, the closer it is to all other nodes.

29

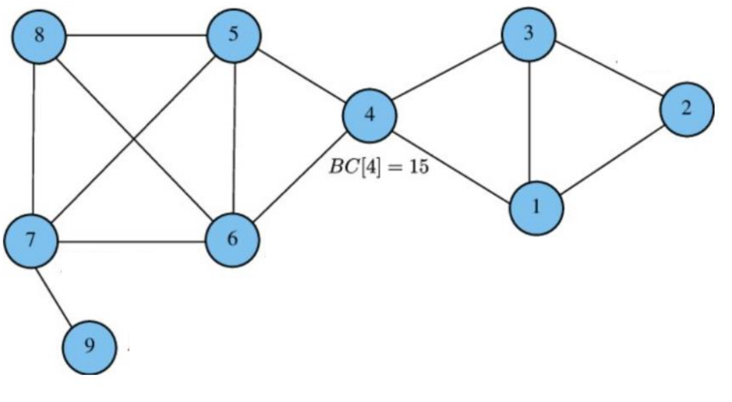


30

Betweenness

- In network analysis, **betweenness** centrality (or betweenness) of each node is how often a node is a bridge between other nodes.

31



32

Weighted network

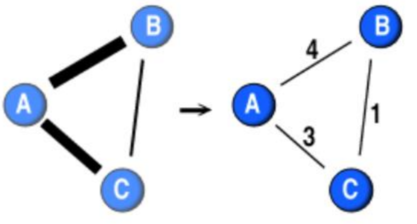
- A **weighted network** is a network with weighted edges.
- Edge weights are often crucial for network analysis and modeling, and many data sets include edge weights. They often represent the strength of a connection, or distance, or some other quantity.
- In many real-world networks, we can observe that not all ties in a network have the same capacity (in terms of their intensity, or capacity).

33

Weighted network

- For example, in social networks, some contacts are friends, whereas others are simply acquaintances.
- Granovetter (1973) argued that the strength of a social tie is a function of its duration, emotional intensity, intimacy, and exchange of services.
- In information networks, variations in the strength of a tie might depend on the flow of information.

34



35