UNIVERSIDAD POLITÉCNICA DE YUCATÁN

SOCIAL NETWORK ANALYSIS

HOMEWORK 2

Professor:
Ing. Didier GAMBOA

Graph Theory

Student:
Walter VIVES
DATA 8A

may 27, 2021

1 Which word or words from the following list describe each of the five networks below: directed, undirected, cyclic, acyclic, approximately acyclic, planar, approximately planar, tree, approximate tree.

a) The Internet, at the level of autonomous systems.

- Undirected (because is used for communication)
- Cyclic
- Non-planar

b) A food web.

- Directed (Since there is a predator and a prey)
- Acyclic (Since there could be possible that there is no connection between the top predators)
- Non-planar (It can be planar, it depends on the nodes taken)

c) The stem and branches of a plant.

- Undirected
- Cyclic
- Planar

d) A spider web.

- Undirected
- Cyclic
- Non-planar

e) A complete clique of four nodes.

- Undirected
- Cycle
- Bus network

Give one real-life example of each of the following types of networks, not including the five examples above:

a) An acyclic (or approximately acyclic) directed network

- World Cup round of 16.

b) A cyclic directed network

- Fase Regular Guardianes 2021 Liga MX. The nodes are the teams and the links are the matches between them. The edges has weight that is determine by the final score.

c) A tree (or approximate tree)

- Floors of a building

d) A planar (or approximately planar) network

- Street Network

e) A bipartite network

- Analysis of Deep South data, also known as the "Southern Women" data, collected in 1941, representing a set of women attending social events over a period of 9 months [2]

2 Is it possible to have the following degrees in a graph with 7 nodes?

{*4,4,4,3,5,7,2*}
Answer: **No**

Why?

Handshaking theorem states that the sum of degrees of the vertices of a graph is twice the number of edges.

$$\sum_{v \in V} \deg v = 2|E| \tag{1}$$

The sum of the degrees of the vertices of the graph is: 29.

Therefore,

$$E = \frac{29}{2} \tag{2}$$

The number of edges must be an integer. Therefore, the graph cannot exist.

3 A simple network consists of n nodes in a single component. What is the maximum possible number of edges it could have? What is the minimum possible number of edges it could have? Explain briefly how you arrive at your answers.

The maximum possible numbers of edges a simple network could have is given by:

It determine the number of pairs of a given a n number of items.

$$L_{max} = \binom{N}{2} = \frac{n(n-1)}{2} \tag{3}$$

So, N is the total number of vertices and the combination nCr is the total of possible edges.

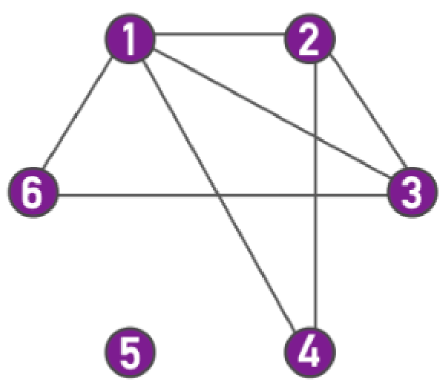
The minimum is given by:

$$(n-1) \tag{4}$$

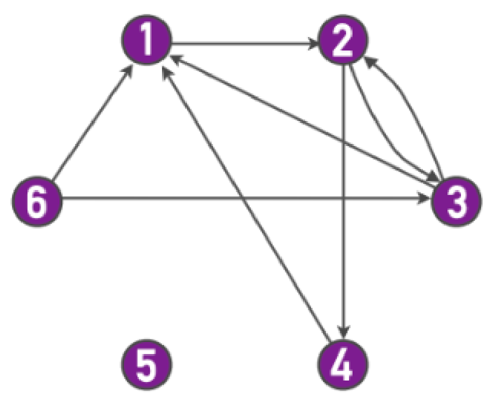
If we fill the fist row of the adjacent matrix we will have *(n-1)* given the upper triangle of the adjacent matrix. It means, that we are going to have at least 1one connection between the nodes.

4 The adjacency matrix is a useful graph representation for many analytical calculations. However, when we need to store a network in a computer, we can save computer memory by offering the list of links in a Lx2 matrix, whose rows contain the starting and end point i and j of each link.

Construct for the networks (a) and (b) in



(a)



(b)

a) The corresponding adjacency matrices.

	1	2	3	4	5	6
1	0	1	1	1	0	1
2	1	0	1	1	0	0
3	1	1	0	0	0	1
4	1	1	0	0	0	0
5	0	0	0	0	0	0
6	1	0	1	0	0	0

Figure 1: Adjacent Matrix (a)

	1	2	3	4	5	6
1	0	1	0	0	0	0
2	0	0	1	1	0	0
3	1	1	0	0	0	0
4	1	0	0	0	0	0
5	0	0	0	0	0	0
6	1	0	1	0	0	0

Figure 2: Adjacent Matrix (b)

b) The corresponding link lists.

- Network (a):
  - (v1,v2)
  - (v1,v3)
  - (v1,v4)
  - (v1,v6)
  - (v2,v3)
  - (v2,v4)
  - (v3,v6)
  - (v5,*None*)
- Network (b):
  - (v1,v2)
  - (v2,v3)
  - (v2,v4)
  - (v3,v2)
  - (v3,v1)
  - (v4,v1)

  - (v5,$None$)
  - (v6,v1)
  - (v6,v3)

c) Determine the average clustering coefficient of the network show in (a).

   Average Clustering Coefficient Formula:

$$\bar{C} = \frac{1}{n}\sum_{i=1}^{n} C_i. \tag{5}$$

$$C_1 = \frac{2}{4C2} = \frac{1}{3}; C_2 = \frac{2}{3C2} = \frac{2}{3}; C_3 = \frac{2}{3C2} = \frac{2}{3}; \tag{6}$$

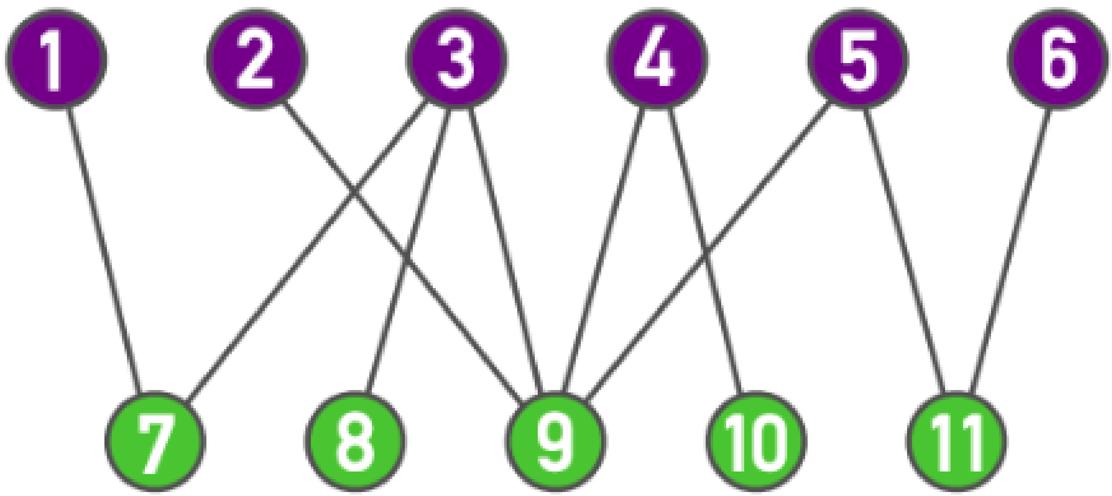
$$C_4 = \frac{1}{2C2} = \frac{1}{1}; C_5 = undefined; C_6 = \frac{1}{2C2} = \frac{1}{1}; \tag{7}$$

$$\bar{C} = \frac{11}{18} \tag{8}$$

d) In the (a) network, how many paths (with possible repetition of nodes and links) of length 3 exist starting from node 1 and ending at node 3? And in (b)?

- Network a: 4
  - 1-2-3
  - 1-3
  - 1-4-2-3
  - 1-6-3
- Network b: 1
  - 1-2-3

5 Consider the bipartite network of the following figure.



a) Construct its adjacency matrix. Why is it a block-diagonal matrix?

	1	2	3	4	5	6	7	8	9	10	11
1	0	0	0	0	0	0	1	0	0	0	0
2	0	0	0	0	0	0	0	0	1	0	0
3	0	0	0	0	0	0	1	1	1	0	0
4	0	0	0	0	0	0	0	0	1	1	0
5	0	0	0	0	0	0	0	0	1	0	1
6	0	0	0	0	0	0	0	0	0	0	1
7	1	0	1	0	0	0	0	0	0	0	0
8	0	0	1	0	0	0	0	0	0	0	0
9	0	1	1	1	1	0	0	0	0	0	0
10	0	0	0	1	0	0	0	0	0	0	0
11	0	0	0	0	1	1	0	0	0	0	0

It is a block-diagonal matrix because there is not connection between the nodes in each part. Therefore:

$$A = \begin{pmatrix} 0 & B \\ B^T & 0 \end{pmatrix} \tag{9}$$

b) Construct the adjacency matrix of its two projections, on the purple and on the green nodes, respectively.

	1	2	3	4	5	6
1	0	0	0	0	0	0
2	0	0	0	0	0	0
3	0	0	0	0	0	0
4	0	0	0	0	0	0
5	0	0	0	0	0	0
6	0	0	0	0	0	0

	7	8	9	10
7	0	0	0	0
8	0	0	0	0
9	0	0	0	0
10	0	0	0	0

c) Calculate the average degree of the purple nodes and the average degree of the green nodes in the bipartite network.

$$AverageDegree = \frac{TotalEdges}{TotalNodes} = \frac{m}{n} \tag{10}$$

- Purple: 10/6
- Green: 10/5

d) Calculate the average degree in each of the two network projections. Is it surprising that the values are different from those obtained in point (c)?

- Network a: 7/6
- Network b: 8/6

It is not surprising given that they are different types of Networks.

6 Cases of Study

1. Netflix keeps data on customer preferences using a big bipartite network connecting users to titles they have watched and/or rated. Netflix's movie library contains approximately 100,000 titles if you count streaming and DVD-by-mail. In the fourth quarter of 2013, Netflix reported having about 33 million users. Assume the average user's degree in this network is 1000. Approximately how many links are in this network? Would you consider this network sparse or dense? Explain.

   - 33,000,000 users
   - AverageDegree: 1,000

   $$TotalEdges = 1,000 \times 33,000,000 = 3.3 \times 10^{10} \tag{11}$$

   It is not a dense network given that the average degree of the users are just 1,000 and there are 33,000,000,000 links approximately.

2. Suppose that from 2013 to 2014 Netflix's library has remained the same size, while the number of users has increased. Further suppose that the average user's degree in this network has remained constant. Has the density of this network increased, decreased, or stayed the same?

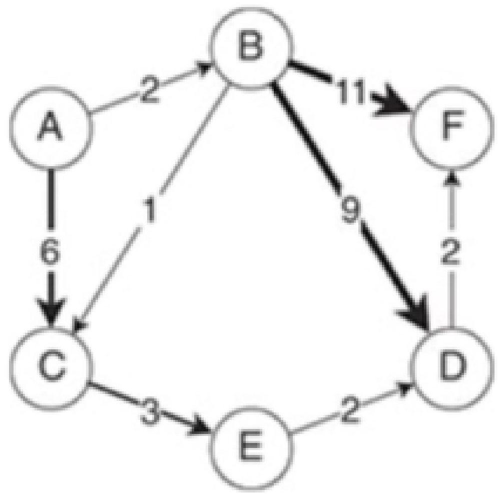
   The density of the network has stayed. The network has more edges but the average user's degree in this network its the same. To be a dense network, the number of links of each node should be close to the maximal number of nodes and in this case the average degree its the same.

7 Consider the following weighted directed network:

a) Which of the following most accurately describes the connectedness of this network?

   i Strongly connected.

   ii **Weakly connected.**



iii Disconnected.

iv None of the above.

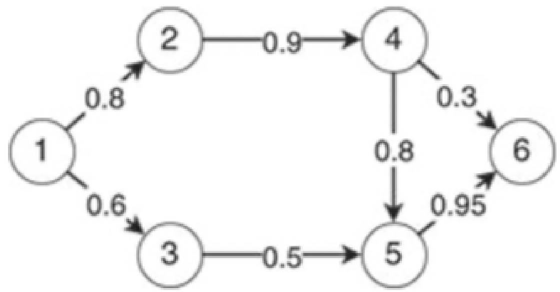
Since there is no way to go from one node to all.

b) What is the in-strength of node D? What is the out-strength of node C?

- in-strength node D: 11
- out-strength node C: 3

c) How many nodes are in the largest strongly connected component? There is not strongly connected components.

8 Consider the following network:



a. Which of the following most accurately describes the connectedness of this network?

   i Strongly connected.
   ii **Weakly connected.**
   iii Disconnected.
   iv None of the above.

b. When discussing path lengths on a weighted graph, one must first define how the weights are related to the distances. The length of a path between two nodes is then the sum of the distances of the links in that path. Consider the previous network and assume that the link weights represent distances. Using this distance metric, what is the shortest path between node 1 and 6?

   - 1-3-5-6: 2.05
   - **1-2-4-6: 2**

c. A common way to define the distance between two nodes is the inverse (or reciprocal) of the link weight. Consider the previous network and assume that the distance between two adjacent nodes is defined as the reciprocal of the link weight. Using this distance metric, what is the shortest path between node 1 and 6?

   - 1-3-5-6: 5.694444444
   - **1-2-4-6: 4.719298246**

References

[1] Wagner, D. K. (2000). Shortest paths in almost acyclic graphs. Operations Research Letters, 27(4), 143-147.

[2] Davis A, Gardner BB, Gardner MR. Deep South: A Social Anthropological Study of Caste and Class. University of South Carolina Press. 2009.

[3] Newman, M. (2018). Networks. Oxford university press.

[4] Barabási, A. L. (2013). Network science. Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences, 371(1987), 20120375.

[5] Chen, K. C., Chiang, M., Poor, H. V. (2013). From technological networks to social networks. IEEE Journal on Selected Areas in Communications, 31(9), 548-572.

[6] Menczer, F., Fortunato, S., Davis, C. A. (2020). A first course in network science. Cambridge University Press.

[7] Network analysis of protein interaction data. An introduction. Available at: https://www.ebi.ac.uk/training/online/courses/network-analysis-of-protein-interaction-data-an-introduction/types-of-biological-networks/

[8] Ricards, N., King, J. (2013). Three Paradoxes of Big Data. Stanford Law Review. Available at: https://www.stanfordlawreview.org/online/privacy-and-big-data-three-paradoxes-of-big-data/

[9] Zafarani, R., Abbasi, M. A., Liu, H. (2014). Social media mining: an introduction. Cambridge University Press.