Complex networks exam simulation

May 2023

Answer to the following questions in a separate paper. In case of doubt comment and motivate your answer in a concise way, explaining only what is necessary to provide your answer.

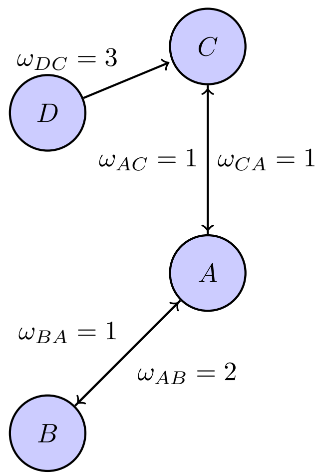
1

Consider the adjacency matrix

$$A = \begin{bmatrix} 0 & 2 & 1 & 0 \\ 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 3 & 0 \end{bmatrix}$$

- Draw the graph
- What type of graph is this?
- What can you tell about the graph connectivity?
- What is the in-out degree of each node?
- Write the strength and degree of each node.

Answer



This is a directed, weighted, weakly connected graph. The in-out degree/strength vectors are

$$
\begin{aligned}
\boldsymbol{k}_{\text{in}} &= (2, 1, 2, 0) \\
\boldsymbol{k}_{\text{out}} &= (2, 1, 1, 1) \\
\boldsymbol{s}_{\text{in}} &= (2, 2, 4, 0) \\
\boldsymbol{s}_{\text{out}} &= (3, 1, 1, 3) \,.
\end{aligned}
$$

The total degree and the total strength is the sum of the in and out degree and strength of each node.

2

Consider an ER random graph with n nodes and connecting probability equal to p. What is the expected clustering coefficient? How does this result relate to the clustering coefficient typically observed in real-world networks?

Answer

The expected clustering coefficient equals the connecting probability p. Since p determines the graph average degree, an ER random graph has a much lower clustering coefficient than a typical real world network with the same average degree.

3

How many edges are there in a Barabasi-Albert network with 1000 nodes and $m = 5$?

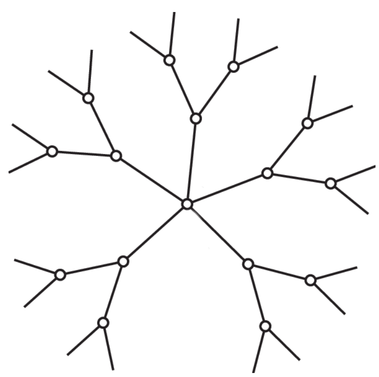
Answer

We assume to start from an initial configuration a clique of m nodes. At each step we then add m edges. We then have L edges

$$L = \frac{m(m-1)}{2} + (n-m)m = 10 + 5000 - 25 = 4985.$$

4

Consider a Cayley tree as the one shown in the picture in which each node has k neighbours (called offsrping). Starting from a central node, we build the tree forming l offspring levels.



- What is the number of nodes as a function of l and k?
- What is the diameter of this graph?
- What property of real world graphs is well captured by this model?

Answer

Remark: the plot is wrong. The graph has n nodes, where

$$n = \sum_{i=0}^{l} k^i = \frac{k^{l+1} - 1}{k - 1} \approx k^l.$$

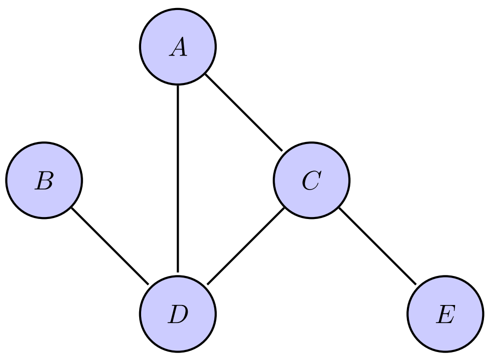
The diameter of this graph is $d = 2l \approx 2\frac{\log(n)}{\log(k)}$ and it is hence small-world, like most real-world networks.

5

Consider the following graph. What is its diameter?

$$A = \begin{bmatrix} 0 & 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 & 1 \\ 1 & 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \end{bmatrix}$$

Answer

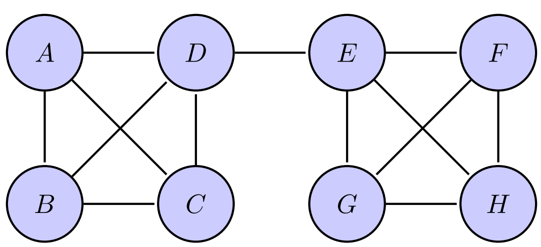


The diameter of this graph is $d = 3$: the longest shortest path is $\{(BD), (DC), (CE)\}$.

6

Provide an example of a graph with a structurally weak tie and motivate your answer.

Answer



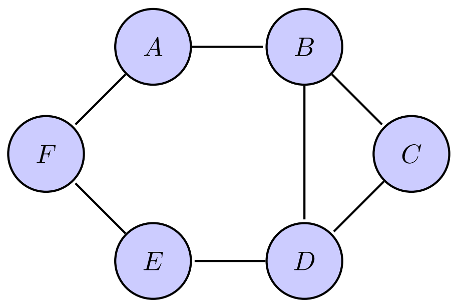
The edge (DE) is a structurally weak tie because their neighbourhoods do not overlap. If it was a strong tie, the absence of triangles to which (DE) participate would violate triadic closure.

7

Draw the following graph and compute the clustering coefficient of every node.

$$\text{edge list} = [(a, b), (b, c), (b, d), (c, d), (d, e), (e, f), (f, a)]$$

Answer

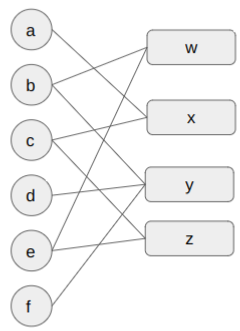


The clustering coefficients are

$$\begin{aligned} C_A &= C_E = C_F = 0 \\ C_B &= C_D = \frac{1}{3} \\ C_C &= 1. \end{aligned}$$

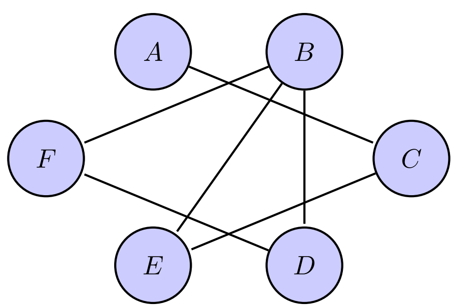
8

Given the bipartite affiliation graph in the figure: (a) Draw the projected graph. (b) Give an example of two different affiliation networks — on the same set of people, but with different foci — so that the projected graphs from these two different affiliation networks are the same. This shows how information can be "lost" when moving from the full affiliation network to just the projected graph on the set of people.

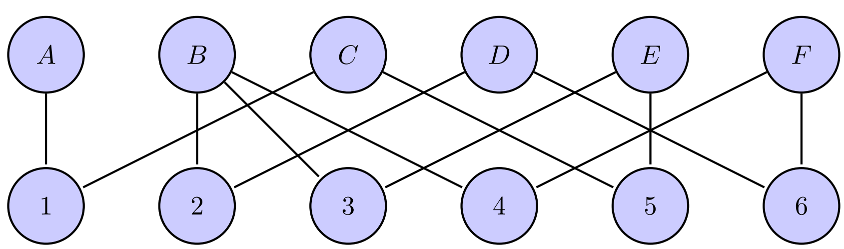


Answer

The projected network is



Another affiliation network that has the same projected network is



9

Suppose you have a valuable piece of information and you want that it spreads quickly on the social network you belong to. Who would you share it with to make the spreading faster and why? (choose between the two for each option)

- A close friend of yours or an acquaintance?
- A very sociable or a moderately sociable person?

Answer

I would share it with an acquaintance because it gives me access to a part of the network I would otherwise be excluded from. I would share it with a sociable person because hubs are super-spreaders.

10

Suppose that some researchers studying educational institutions decide to collect data to address the following two questions. (a) As a function of k, what fraction of Cornell classes have k students enrolled? (b) As a function of k, what fraction of 3rd-grade elementary school classrooms in New York State have k pupils? Which one of these would you expect to more closely follow a power-law distribution as a function of k? Give a brief explanation for your answer, using some of the ideas about power-law distributions.

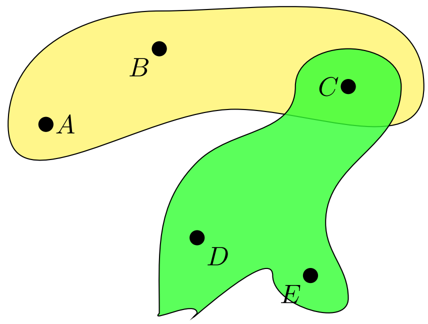
Answer

Considering that university classes are more heterogeneous with respect to the number of students, it is more likely that the distribution associated to Cornell classes follows more closely a power law distribution. On the other hand, 3rd grade classes sizes cannot be arbitrarily large or small, hence farther from a power law distribution.

11

Draw a hyper-graph and its corresponding hyper-edge list representation.

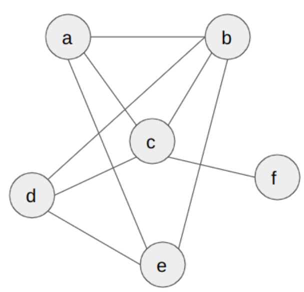
Answer



The hyper-edge list is $[(A, B, C), (C, D, E)]$.

12

Consider the social network represented in the figure. Suppose that this social network was obtained by observing a group of people at a particular point in time and recording all their friendship relations. Now suppose that we come back at some point in the future and observe it again. According to the theories based on empirical studies of triadic closure in networks, which new edge is most likely present? (I.e. which pair of nodes, that do not currently have an edge connecting them, are most likely to be linked by an edge when we return to take the second observation?) Also, give a brief explanation for your answer.



Answer

We must compute the overlap between all pairs of nodes that are not neighbours.

$$O(af) = \frac{1}{3}$$
$$O(ad) = 1$$
$$O(bf) = \frac{1}{3}$$
$$O(ce) = \frac{3}{4}$$
$$O(ef) = 0,$$

thus the most likely edge to appear is (ad).

13

Any network with N nodes and at least N edges must contain a cycle (True/False) while any network with N nodes and less than N edges does not contain any cycle (True/False).

Answer

True, because if you add an edge to a tree you necessarily introduce a cycle. False because the graph may be disconnected.

14

If we consider an epidemic spreading on a graph, only the nodes in the connected component of the first infected individuals can be infected. With this in mind, if we consider an epidemic spreading on an Erdos Renyi graph below the percolation threshold, is it possible that a finite fraction of the population will get infected?

Answer

False: at most all the giant component will get infected. If we are below the percolation threshold, by definition, the size of the giant component is vanishing with respect to the graph size.

15

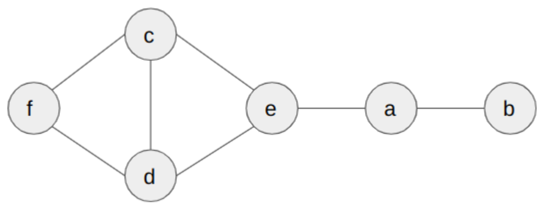
Consider the degree sequence $k = (1, 3, 5, 3)$: is this a reasonable input to create an unweighted graph without self-loops from the configuration model? What about the sequence $k = (1, 3, 4, 3)$

Answer

$k = (1, 3, 5, 3)$ is not a reasonable input because $k_3 = 5 > n = 4$. The vector $k = (1, 3, 4, 3)$ instead is not reasonable because the sum of its entries is odd.

16

Given the graph in the figure. What is/are the node/s with the highest betweenness centrality, and which ones have the highest degree centrality?



Answer

The nodes d, c, e have the highest degree, equal to 3. The node e has the highest betweennes centrality.

17

Consider an ER graph with $n = 5 \cdot 10^4$ nodes and average degree $k = 100$. How can you approximate its diameter?

Answer

The diameter d will be approximately

$$d \approx \frac{\log(n)}{\log(k)} = 4 + \log(5).$$

18

Explain why the Watts-Strogatz model is able to create networks with high clustering coefficients and small diameter.

Answer

The Watts-Strogatz model starts from high clustering and large diameter initial configuration. It is then performed the random rewiring of some edges with a probability p. In a range of values of p, the random rewiring only marginally decreases the clustering coefficient, but it adds “shortcuts” that significantly reduce the graph diameter.

19

Show that any partition of a clique has negative modularity

Answer

$$
\begin{aligned}
Q &= \frac{1}{2L}\sum_{ij}\left(A_{ij} - \frac{k_i k_j}{2L}\right)\delta(\ell_i, \ell_j) \\
&= \sum_{i \neq j}\left(1 - \frac{n-1}{n}\right)\delta(\ell_i, \ell_j) - \sum_i \frac{n-1}{n} \\
&\leq n(n-1)\frac{1}{n-1} - (n-1) = 0.
\end{aligned}
$$

20

True or False, the older nodes of a Barabasi Albert network are those with higher degree.

Answer

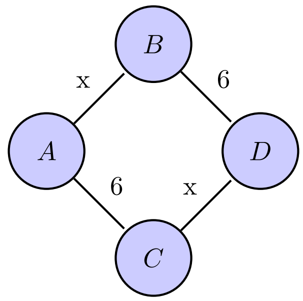
True.

21

Does the creation of a new fast road always decrease the total travelling time? If not, provide a counter-example.

Answer

False. Consider the following network with 5 cars, and the weight equal to the travelling time. x indicates that the travelling time equals the number of cars that go by that route.



Adding an edge (BC) with weight 0 would increase the travelling time.