Preliminary draft 07:11 5 July 2019

5 July 2019
student@hse.edu.ru

HSE Notes

Author Name
HSE, MAGoLEGO "Social Network Analysis", SNA, Russia, Moscow

Keywords: Bending Magnet, path length, pole face angle, beam trajectory, survey, GEODE

Summary

This document shows how to prepare for MAGoLEGO "Social Network Analysis" final exam. There are few sections in the exam and might be approximately 10-12 questions with different points assign. Time for exam - 120 minutes. Russian answers are allowed, but not advised. You are allowed to bring and use simple calculator. Computers, mobiles and materials are forbidden. Cheating or breaking exam policy leads to grade 0 for the exam.

Contents

1 Task 1 **2**

2 Task 2 **2**
- 2.1 Task 2.1 3
- 2.2 Task 2.2 3

3 Task 3 **3**

4 Task 4 **3**

5 Task 5 **4**
- 5.1 Task 5.1 4
- 5.2 Task 5.2 4
- 5.3 Task 5.3 5

5.4 Task 5.4 5
5.5 Task 5.5 5

6 Task 6 **5**

7 Task 7 **5**

8 Task 8 **6**

9 Task 9 **7**

10 Task 10 **7**

11 Task 11 **7**

1 Task 1

Construct the adjacency matrix for the following direct network:

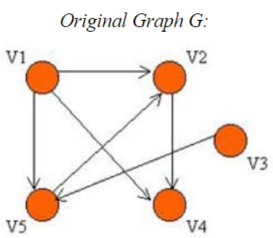
Solution:

An adjacency matrix is defined as follows:

Let G be a graph with "n" vertices that are assumed to be ordered from v1 to vn. The n x n matrix A, in which (the existence of an edge between two vertices vi and vj is shown by an entry of 1 in the ith row and jth column of the adjacency matrix. This entry represents a path of length 1 from vi to vj.)

aij= 1 if there exists a path from vi to vj

aij = 0 otherwise.



Matrix representation of path of length 2

	v1	v2	v3	v4	v5
v1	0	1	0	1	0
v2	0	0	0	0	0
v3	0	1	0	0	0
v4	0	0	0	0	0
v5	0	0	0	1	0

Figure 1: Task 1

2 Task 2

Suppose you are studying the spread of a rare disease among the set of people pictured in Figure 21.22 2. The contacts among these people are as depicted in the network in the figure, with a time interval on each edge showing when the period of contact occurred. We assume that the period of observation runs from time 0 to time 20. (This task is similar to the task in the book 21.9 [2])

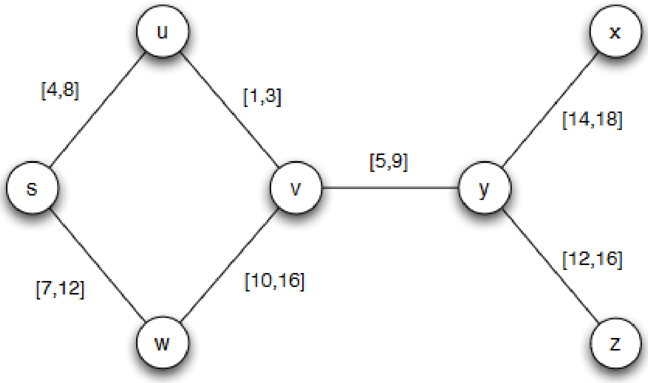


Figure 2: Task 2

2.1 Task 2.1

if node S is the only source of decease at time 0, which nodes could potentially acquire the decease by the end of observation time?

Solution: S, U W, V

2.2 Task 2.2

if it is known, that everyone got infected, but there were no external sources of infection except node S, could you correct a single number (start or end of an interval) so it is possible for infection to spread over all network?

Solution: 15

3 Task 3

Consider an undirected network of size N in which every node has a degree k = 1. Draw possible network. How many connected components does it have? What condition should N satisfy? Write labels of nodes that will adopt behaviour A.

Solution: N/2 == 0 (even), connected components (1)—-(2) (3)——(4)

4 Task 4

Consider the model from Chapter 19 in Figure 19.29 [1] for the diffusion of a new behavior through a social network. Recall the model of influence propagation. All nodes follow behaviour B and have threshold q = 2/5. At some point nodes C and D have switched to behaviour A. Write labels of nodes that will adopt behaviour A.

Solution: C, D, E, F, H, I

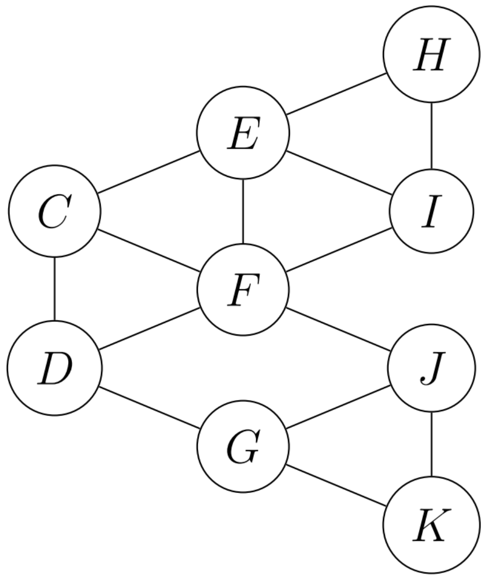


Figure 3: Task 4

5 Task 5

For the given network find

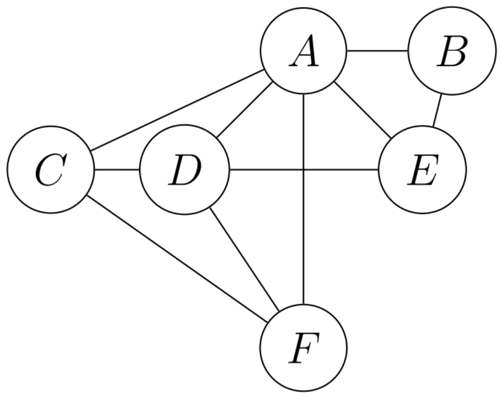


Figure 4: Task 5

5.1 Task 5.1

Network diameter **Solution:** 2

5.2 Task 5.2

Max degree **Solution:** 4

5.3 Task 5.3

Min degree **Solution:** 2

5.4 Task 5.4

The node with the smallest clustering coefficient **Solution:** B

5.5 Task 5.5

The node with the largest clustering coefficient **Solution:** A, C, E

6 Task 6

Given the following graph Plot degree distribution (histogram) and answer the question whether it satisfies Power-Law

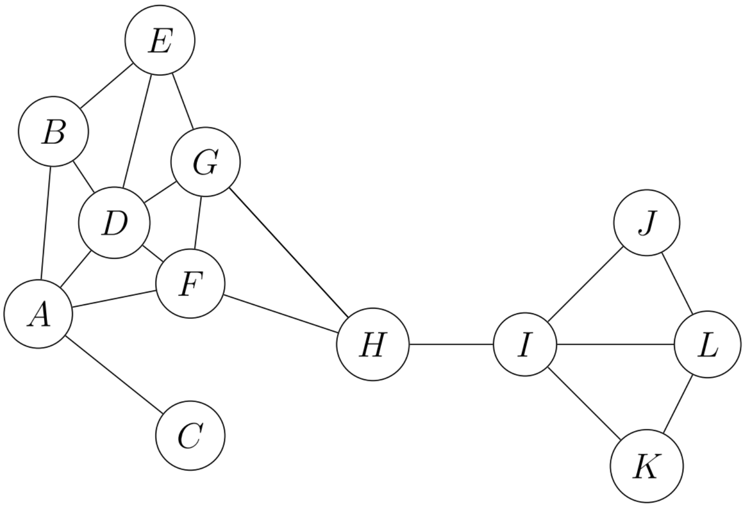


Figure 5: Task 6

Solution: Make a diagram - node degree (horizontal) and number of nodes (vertical) - as a result, you will get a distribution, which would not satisfy Power Law Task 6 - Solution

Name of node	A	B	C	D	E	F	G	H	I	J	K	L
Degree	4	3	1	5	3	4	4	3	4	2	2	3

7 Task 7

Find a 3-core of the given network

Solution: (1, 2, 8, 9)

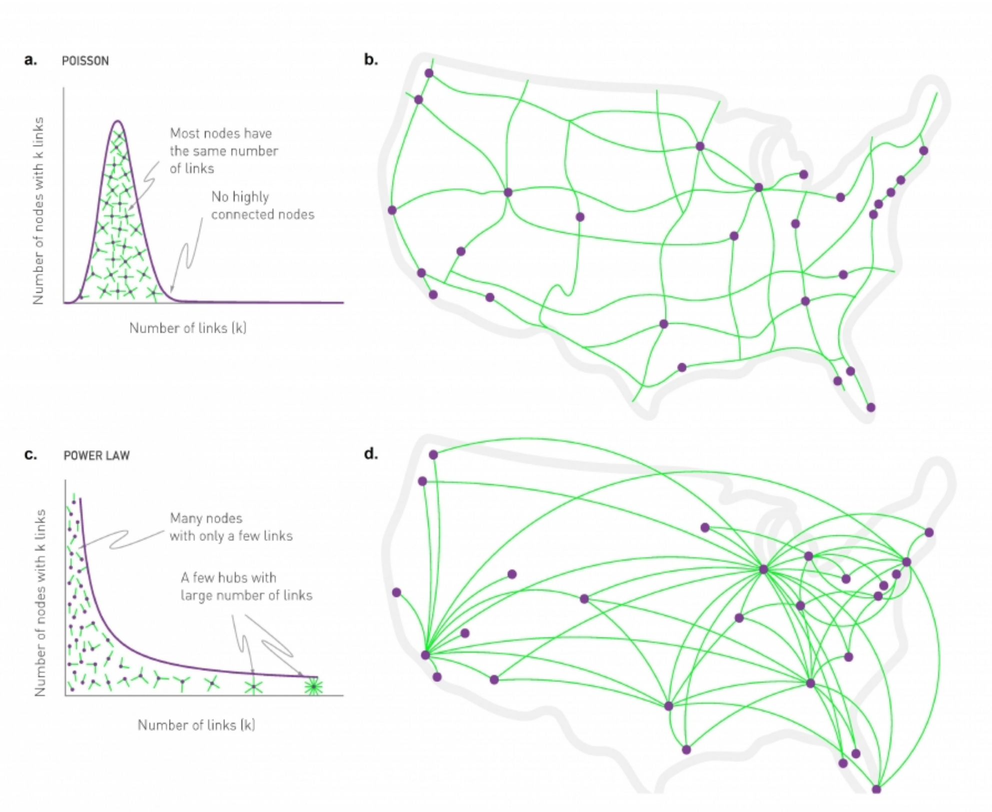


Figure 6: Task 6 - Solution

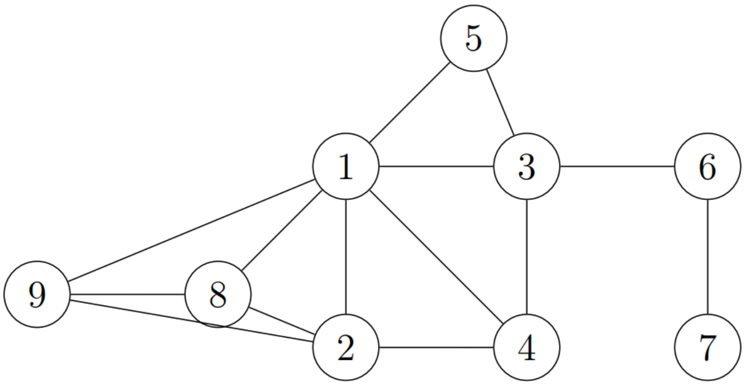


Figure 7: Task 7

8 Task 8

Compute Hubs and Authorities values (HITS) for the given directed graph G=($\{A, B, C\}$, $\{(C, A), (C, B), (A, B), (B, A)\}$). Normalize the results using Euclidean norm.

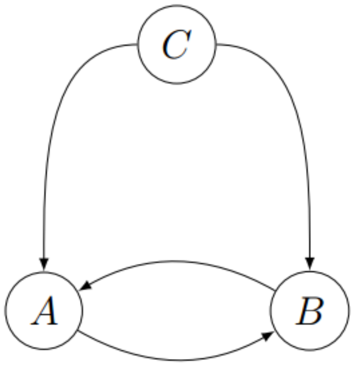


Figure 8: Task 8

9 Task 9

For the four networks indicate assortative and disassortative cases with respect to degree and closeness centrality metrics

Solution: A. Disassortative network by degree - uniform (low degree adjacent to high degree) (high degree nodes connected to low degree nodes, star-like structure)

B. Assortative network by degree - core of high degrees and a periphery of low degrees (interconnected high degree nodes - core, low degree nodes - periphery)

Disassortative and assortative networks. Schematic illustration of a disassortative network (A) and an assortative network (B). C. The 15 best connected proteins and their direct links to other proteins of yeast protein network constructed by proteins localized in nucleus. D. The rest of network after removal of the 15 best connected nodes. Nodes disconnected to the largest component are not shown. A predominant feature of B and D is the over-abundance of links between low connected nodes.

10 Task 10

Calculate the number of nodes reachable in d steps from the central node of a Cayley tree (every node has a degree k). Example of a tree with k = 3 is shown below:

Solution: Task 10

Cayley Tree

A Cayley tree is a symmetric tree, constructed starting from a central node of degree k. Each node at distance d from the central node has degree k, until we reach the nodes at distance P that have degree one and are called leaves (see Image 3.16 for a Cayley tree with k = 3 and P = 5.)

11 Task 11

In which case (a,b,c,d) modularity will be the largest? Why? Nodes are colored according to the communities they are assigned to. Compute Modularity

Higher Modularity Implies Better Partition

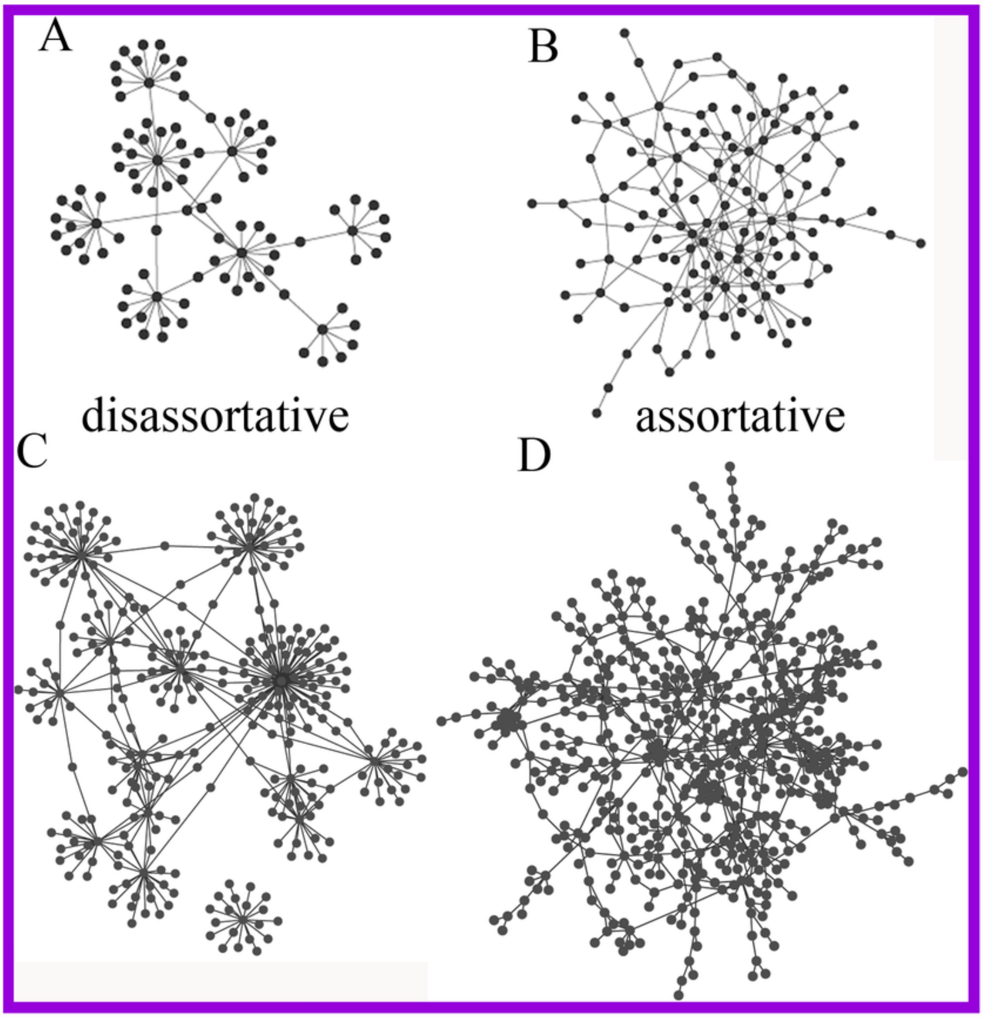


Figure 9: Task 9

The higher is M for a partition, the better is the corresponding community structure. Indeed, in Image 9.16a the partition with the maximum modularity (M=0.41) accurately captures the two obvious communities. A partition with a lower modularity clearly deviates from these communities (Image 9.16b). Note that the modularity of a partition cannot exceed one [31,32].

Zero and Negative Modularity

By taking the whole network as a single community we obtain M=0, as in this case the two terms in the parenthesis of (9.12) are equal (Image 9.16c). If each node belongs to a separate community, we have Lc=0 and the sum (9.12) has nc negative terms, hence M is negative (Image 9.16d).

Modularity - To better understand the meaning of modularity, we show M defined in (9.12) for several partitions of a network with two obvious communities.

a. Optimal Partition - The partition with maximal modularity M = 0.41 closely matches the two distinct communities.

b. Suboptimal Partition - A partition with a sub-optimal but positive modularity, M

Section 3.12

Advanced Topic 3.A Deriving the Poisson Distribution

To derive the Poisson form of the degree distribution we start from the exact binomial distribution (3.7)

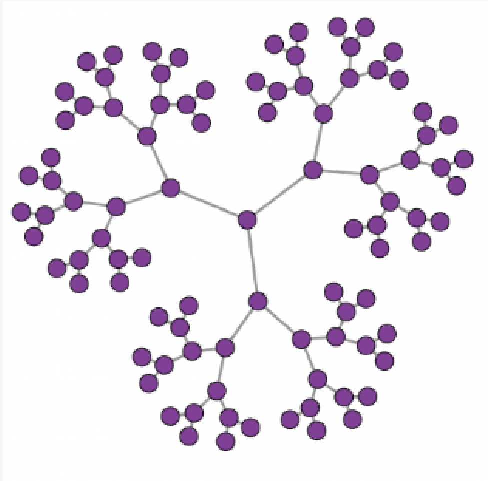


Image 3.16

Cayley Tree

A Cayley Tree With k = 3 and P = 5.

$$p_k = \binom{N-1}{k} p^k (1-p)^{N-1-k} \qquad (3.22)$$

that characterizes a random graph. We rewrite the first term on the r.h.s. as

$$\binom{N-1}{k} = \frac{(N-1)(N-1-1)(N-1-2)\ldots(N-1-k+1)}{k!} \approx \frac{(N-1)^k}{k!} \qquad (3.23)$$

where in the last term we used that $k \ll N$. The last term of (3.22) can be simplified as

$$\ln\left[(1-p)^{(N-1)-k}\right] = (N-1-k)\ln\left(1-\frac{\langle k\rangle}{N-1}\right)$$

and using the series expansion

$$\ln(1+x) = \sum_{n=1}^{\infty}\frac{(-1)^{n+1}}{n}x^n = x - \frac{x^2}{2} + \frac{x^3}{3} - \ldots, \forall|x| \le 1$$

we obtain

$$\ln\left[(1-p)^{(N-1)-k}\right] \approx (N-1-k)\frac{\langle k\rangle}{N-1} = -\langle k\rangle\left(1-\frac{k}{N-1}\right) \approx -\langle k\rangle$$

which is valid if $N \gg k$. This represents the *small degree approximation* at the heart of this derivation. Therefore the last term of (3.22) becomes

$$(1-p)^{N-1-k} = e^{-\langle k\rangle} \qquad (3.24)$$

Combining (3.22), (3.23), and (3.24) we obtain the Poisson form of the degree distribution

$$p_k = \binom{N-1}{k} p^k (1-p)^{(N-1)-k} = \frac{(N-1)^k}{k!}p^k e^{-\langle k\rangle} = \frac{(N-1)^k}{k!}\left(\frac{\langle k\rangle}{N-1}\right)^k e^{-\langle k\rangle}$$

or

$$p_k = e^{-\langle k\rangle}\frac{\langle k\rangle^k}{k!} \qquad (3.25)$$

Section 3.13

Figure 10: Task 10

= 0.22, fails to correctly identify the communities present in the network.

c. Single Community - If we assign all nodes to the same community we obtain M=0, independent of the network structure.

d. Negative Modularity - If we assign each node to a different community, modularity is negative, obtaining M=-0.12.

We can use modularity to decide which of the many partitions predicted by a hierarchical method offers the best community structure, selecting the one for which M is maximal. This is illustrated in Image 9.12f, which shows M for each cut of the dendrogram, finding a clear maximum when the network breaks into three communities.

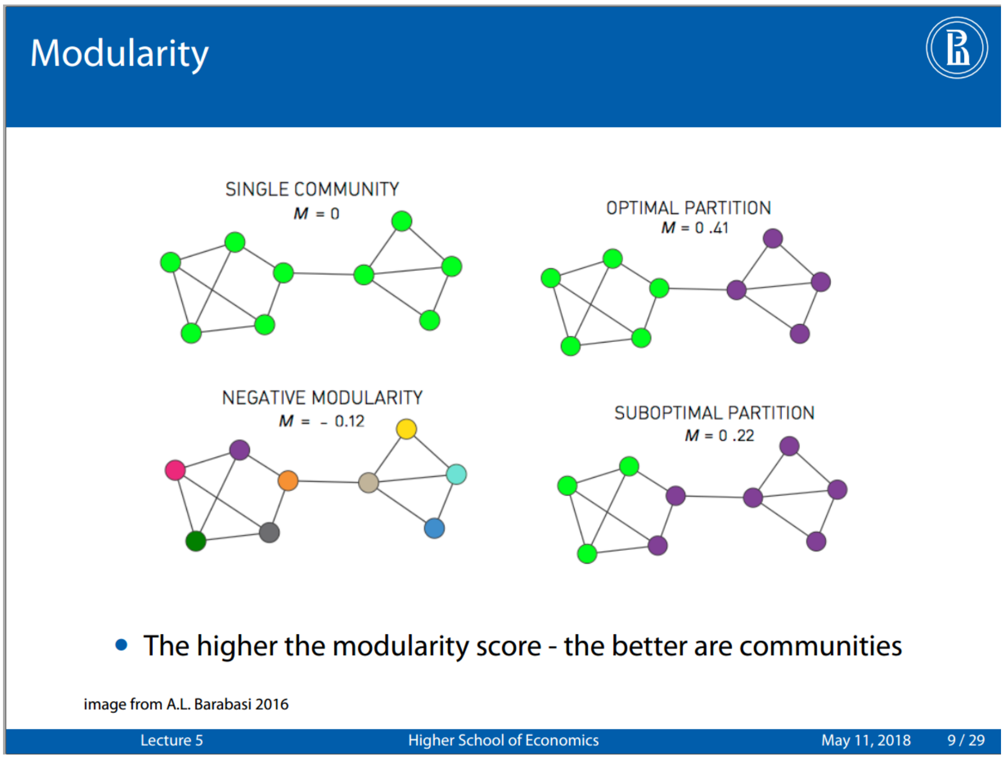


Figure 11: Task 11

References

[1] task4 networks, crowds, and markets: Reasoning about a highly connected world. https://courses.cit.cornell.edu/info204_2007sp/diffusion.pdf, 2008.

[2] David Easley and Jon Kleinberg. task2. https://www.cs.cornell.edu/home/kleinber/networks-book/networks-book-ch21.pdf.