Network Science - written exam: pre-call, January 15, 2024

first name	**family name**	**student no.**

Multiple-choice questions (12 points):

1. Which of the following is a bipartite network?

   ☐ a network of people linked by friendship

   ☐ a network of words linked whenever they appear in the same sentence

   ☐ a network of books connected to the authors

   ☐ a network of plants linked if they grow in the same area

2. What information gives a power-law exponent $\gamma = 2.14$?

   ☐ We are dealing with a random network

   ☐ We are dealing with a small world network

   ☐ We are dealing with an ultra-small world network

   ☐ We are dealing with a scale-free regime with no large hubs

3. Under the Bianconi-Barabasi model, what is the information carrried by the dynamic exponent η of growth?

   ☐ It identifies the node degree centrality, the larger η the higher the degree

   ☐ It identifies the attractiveness of a node, the larger η the faster the node growth

   ☐ It identifies the fitness of a node, the smaller η the faster the node growth

   ☐ It identifies a node quality independent on the growth

4. Consider a semantic network of hashtags that are connected if they appear in the same tweet. Which of the following would you use to identify the relevance of a hashtag as authoritative in the online discourse?

   ☐ betweenness

   ☐ pagerank

   ☐ closeness

   ☐ assortativity

Network Science - written exam: pre-call, January 15, 2024

first name	**family name**	**student no.**

5. Which one is the correct formalization of modularity in a directed network with normalized adjacency matrix $\boldsymbol{A}$ and community assignment $\boldsymbol{C}$?

- ☐ $Q = \text{trace}(\boldsymbol{C}(\boldsymbol{A} - \boldsymbol{d}_{\text{in}}, \boldsymbol{d}_{\text{out}}^T)\boldsymbol{C}^T)$, $\boldsymbol{d}_{\text{in}} = \boldsymbol{A}\mathbf{1}$, $\boldsymbol{d}_{\text{out}} = \boldsymbol{A}^T\mathbf{1}$
- ☐ $Q = \text{trace}(\boldsymbol{C}(\boldsymbol{A} - \boldsymbol{d}_{\text{out}}, \boldsymbol{d}_{\text{in}}^T)\boldsymbol{C}^T)$, $\boldsymbol{d}_{\text{in}} = \boldsymbol{A}\mathbf{1}$, $\boldsymbol{d}_{\text{out}} = \boldsymbol{A}^T\mathbf{1}$
- ☐ $Q = \text{sum}(\boldsymbol{C}(\boldsymbol{A} - \boldsymbol{d}_{\text{in}}, \boldsymbol{d}_{\text{out}}^T)\boldsymbol{C}^T)$, $\boldsymbol{d}_{\text{in}} = \boldsymbol{A}\mathbf{1}$, $\boldsymbol{d}_{\text{out}} = \boldsymbol{A}^T\mathbf{1}$
- ☐ $Q = \text{trace}(\boldsymbol{C}(\boldsymbol{A} - \boldsymbol{d}_{\text{in}}, \boldsymbol{d}_{\text{out}}^T)\boldsymbol{C}^T)$, $\boldsymbol{d}_{\text{in}} = \boldsymbol{A}^T\mathbf{1}$, $\boldsymbol{d}_{\text{out}} = \boldsymbol{A}\mathbf{1}$

6. What is the normalized mutual information (NMI) measuring in a topic assignment on a word-to-document network?

- ☐ How well documents are described by the words they contain
- ☐ How well words are interconnected inside each topic
- ☐ How well documents inside a topic are connected
- ☐ How well words are able to clearly identify the topics

7. What is the main drawback of stochastic block models (SBMs)?

- ☐ They cannot capture the most common network structures
- ☐ They are only applicable to binary adjacency matrices
- ☐ They They generally too complex to be run on large networks
- ☐ They cannot handle distributions other than the Gaussian one

8. What is HDBSCAN?

- ☐ It is an hierarchical agglomerative approach, providing a dendrogram structure
- ☐ It is an hierarchical divisive approach, based on distances
- ☐ It is a community detection algorithm based on an empirical simplification of modularity
- ☐ It is an optimization tool for community detection

Network Science - written exam: pre-call, January 15, 2024

first name	**family name**	**student no.**

9. Consider ForceAtlas2. Which of the following is true?

- ☐ It is a community detection algorithm based on nodes distances
- ☐ It is a force directed layout algorithm where the equilibrium distance is proportional to the desired distance
- ☐ It is a force directed algorithm taking into account the node degree, for a better visual layout
- ☐ It is based on a logarithm mapping of nodes distances

10. How to properly build a semantic network from occurrences of words in documents, $\boldsymbol{N}_{wd}$?

- ☐ By building a document-to-document probability matrix $\boldsymbol{P}_{dd}$
- ☐ By exploiting the TF-IDF matrix linking words to documents
- ☐ By doing a projection onto the words network, i.e., counting how many times two words appear in the same document
- ☐ By linking words only if they appear in a sufficient number of documents

11. How does non-negative matrix factorization (NMF) works?

- ☐ It assumes a Dirichlet distribution for topics
- ☐ It considers a binary document-to-topic map $\boldsymbol{C}$
- ☐ It does not take into account for the document-to-topic map $\boldsymbol{C}$
- ☐ It expresses the word probability in each document as the weighted collection of the distribution inside each topic contained in the document

Network Science - written exam: pre-call, January 15, 2024

first name	**family name**	**student no.**

12. How can we identify the presence of an echo chamber in a network?

- ☐ When the individual leaning of nodes is similar to the individual leaning of their neighbors
- ☐ When the degree of nodes is similar to the degree of their neighbors
- ☐ When the individual leaning of nodes is always positive
- ☐ When the nodes are organized in communities

Network Science - written exam: pre-call, January 15, 2024

first name	**family name**	**student no.**

Open question #1 (10 points, 1 page):
Explain and comment the PageRank algorithm when applied on an undirected network whose adjacency matrix is $\boldsymbol{A}$. Consider discussing the following aspects: Which is the PageRank equation? Which are the parameters of interest? What is their meaning? How can these be used to infer interesting analytics from the network? Any further (mathematical) insight is welcome.

Network Science - written exam: pre-call, January 15, 2024

first name	**family name**	**student no.**

Open question #2 (10 points, 1 page):
List, describe in some detail, and discuss the most interesting evaluation metrics for assessing the quality of a community detection algorithm. Any insight on mathematical equations, meaning, reliability, generality, etc. is welcome. Assume that the adjacency matrix $\boldsymbol{A}$ is normalized (elements summing up to 1).