Solutions to Week2-Assignment1

1. A

The nodes in the citation network represent the research papers and an edge from say, a node A to a node B indicates that the paper A cites the paper B. This sort of information being directional, makes a citation network a directed network.

2. B

The nodes in the co-authorship network represent the authors and an edge between say, a node A and node B indicates that the author A and the author B have collaborated for some research work. The symmetrical nature of this information makes a co-authorship network an undirected network.

3. C

The full form of CSV is 'Comma Separated Values'.

4. B

read_weighted_edgelist() function in networkx can be directly used to read a network dataset where the edges have been assigned weights, however, this function can only be used to read the weights if they are numeric such as 5, 3.5 etc, and not in any other format such as string.

5. B

GML stands for Graph Modeling Language and it is different from a GraphML format. Also, it is possible to store the attributes of nodes and edges in both the formats. Therefore, the options A, C and D are not the correct answers. However, GML does not store the data in the form of tags like XML unlike GraphML, hence the option B is false and hence the correct answer.

6. B

Gephi is written in Java language.

7. B

GEXF format was created in 2007 by the people working for the Gephi project. This new format was created to avoid the exchange issues associated with the existing formats and to provide better extensibility and ease of storing various details regarding the networks.

8. C

Gephi provides features for the analysis as well as the visualization of networks.

9. D

nx.read_edgelist() function reads a graph data set where each edge is listed in a separate row. nx.read_adjlist() function reads a graph where the data is listed in the form of an adjacency list such that in a row, the first node is taken as the source and the subsequent nodes in the row are taken as the targets of the source node. Since the edgelist formatted data also carries the same property, i.e. the first node should be taken as the source node and the subsequent node (i.e. the only node) should be taken as the target of the source node, the data in the given format can be read using both the functions, nx.read_edgelist() as well as nx.read_adjlist().

10. C

GEXF stands for **G**raph **E**xchange **X**ML **F**ormat.

11. C

.NET extension is used for graphs in Pajek format; .txt is used for edgelist or adjlst formats; .GDF is a format for graphs which is similar to CSV format. However, .NTF is a format for images, and not for networks.

12. D

Adjlist format carries limited details about a network and is relatively less flexible when it comes to assigning attributes to graph, nodes and edges.

13. A

GEXF format uses tags to store network details similar to XML.

14. B

In a complete network, all the neighbors of *any* given node are friends to each other, hence, the clustering coefficient of *every* node will be 1. Therefore, the average clustering coefficient will also be 1.

15. C

Degree distribution of most real-world networks follows power law, which means there are very few nodes with very high degrees and there are a lot of nodes with very less degrees.

16. D Diameter of a network is the number of edges on the shortest path between the two most distant nodes in the network.

17. B

In a star graph with 10 nodes, there will be one central node and 9 periphery nodes, which are directly connected to the central node, but not to each other. Clustering coefficient of a given node tells the extent of friendships amongst the neighbors of a given node. Since all the neighbors of the central node (i.e. all

the 9 nodes) are not connected to each other, the clustering coefficient of central node will be 0.