1. What is a Subgraph?
   a. A graph within a graph
2. What is an Induced Subgraph?
   a. A subgraph that is created from an initial selection of nodes
3. What does the half-hop induced subgraph tell you about selected nodes on a graph?
   a. Whether the selected nodes are connected / are one hop away
4. What does the one-hop induced subgraph tell you about selected nodes on a graph?
   a. Whether the selected nodes have common neighbours / are two hops away
5. What is an ego-network?
   a. A subgraph of a node's neighbours and their connections
6. Why are ego-networks useful?
   a. They identify triangles within a node's local network
7. Why are triangles important in social network analysis?
   a. They are indicative of close bonds / communities
8. How many communities is the blue blazed node connected to?
   a. 2
9. What sort of words emerge in these communities?
   a. One had supermarket terms, one has news/newspaper terms
10. How many communities is the red blazed node connected to?
   a. 2
11. What sort of words emerge in these communities?
   a. One is tied to Kathy Arberg, the other to words around the anthrax investigation in 2001.
12. How many communities is the yellow blazed node connected to?
   a. 3
13. What sort of words emerge in these communities?
   a. One is about Phil Landrigan, one is about asbestos in general, the other is about the health risks.
14. Which node has the highest Common Neighbours score with the green blazed node?
   a. Anthrax
15. How many neighbours do they have in common?
   a. 27
16. How many neighbours do they not have in common?
   a. 256
17. What is Jaccard Index?
   a. The ratio of common features to all features observed
18. Which node has the highest Jaccard Index score with the green blazed node?
   a. biological
19. How many neighbours do they have in common?
   a. 25
20. How many neighbours do they not have in common?
   a. 58
21. What is Levenshtein Distance?
   a. A measure of the distance between two character strings based on the number of changes needed to transform one to the other.
22. Which words in the graph are linked by a Levenshtein Distance of 1 would you consider merging?
   a. flulike and flu-like, us_postal_service and us_postal_services, bioterrorism and bio-terrorism, letter-borne and letter-born, offsite and off-site, mail-born and mail-borne

23. What is Clustering?

    a. Clustering is the process by which we identify interconnected groups of nodes within a larger network, utilising the structure of the graph to find communities.

24. What is a k-Truss?

    a. A k-Truss is a graph structure whereby each edge can be found in k-2 triangles.

25. What is the highest k-Truss value of the entire graph?

    a. 23

26. How many nodes are within this k-Truss?

    a. 39

27. How many departments are represented in this k-Truss?

    a. 18

28. What is Hierarchical Clustering? What is the process of creating clusters?

    a. Another form of clustering by which each node starts in their own cluster, and edges are randomly added to create groups, with a test at each step calculating whether the group that has been created is more interconnected then the network outside that community. The principle is that entities are more likely to be connected to members of their own group than with those outside their group.

29. What is the key difference between Hierarchical Clustering and k-Truss clustering?

    a. K-Trusses are limited by the structure; if there are no triangles there are no groups. Hierarchical Clustering puts every node in a cluster.

30. How many clusters are there at the optimal level for Hierarchical Clustering? How does this compare with the number of departments?

    a. As this is random, it will vary, but in my case, 38 was the optimal level – there are 42 departments in total.

31. Set the number of clusters to 42 and run Arrange by Node Attribute – Clustering.Hierarchical. How well do the clusters match the truth data?

    a. As this is random, it will vary, but in my case, it did a fairly good job with department 14.

32. Run the shortest paths between clusters – which of the selected nodes has the highest NeighbouringDepartmentCount value?

    a. Email 121, with 39.

33. Run Betweenness Centrality on the graph and Scatter Plot Centrality.OutBetweenness with NeighbouringDepartmentCount. What is the relationship?

    a. There is a positive linear relationship between the number of neighbouring departments and OutBetweenness Centrality.

34. Run k-Truss on this network. Which characters make up the highest k-Truss group? (Hint: Use the character guide found on Wikipedia: https://en.wikipedia.org/wiki/List_of_Les_Mis%C3%A9rables_characters)

a. This group is mostly made up of characters from the ABC Society.

35. Which characters make up the next highest k-Truss group?

    a. This group is mostly made up mostly of characters from the Year 1817.

36. What is Dice Similarity?

a. The Dice Similarity is the ratio of common neighbours between two nodes compared to the median number of neighbours of both nodes. The more neighbours two nodes have in common, the more similar they are. The features used to determine similarity with the Dice Similarity are a node's neighbours.

37. What is Cosine Similarity?

a. The Cosine Similarity is a measure of the angle between the feature vectors of two nodes. The features used to determine similarity with cosine similarity are a node's neighbours, including the direction, types and weight of the transactions connecting them. A node's features are converted into vectors, (geometric objects defined by a magnitude and a direction), and similarity is then measured by comparing the direction of the vectors of two nodes. Similar vectors will point in similar directions.

38. What is the Resource Allocation Index?

a. The Resource Allocation Index analytic calculates the fraction of a "resource" that a node can send to another through their common neighbours. It is calculated by summing 1 divided by the degree of each common neighbour node.