Reuters Terror News Network (http://vlado.fmf.uni-lj.si/pub/networks/data/CRA/terror.htm)
The ***Reuters terror news network*** is based on all stories released during 66 consecutive days by the news agency Reuters concerning the September 11 attack on the U.S., beginning at 9:00 AM EST 9/11/01. The vertices of a network are words (terms); there is an edge between two words iff they appear in the same text unit (sentence). The weight of an edge is its frequency. The network has n = 13332 vertices (different words in the news) and m = 243447 edges, 50859 with value larger than 1. There are no loops in the network.

This particular graph is taken from a subset of words identified using hierarchical clustering. Using Reuters_Terror_News_Network_HC_20.star, answer the following questions:

1. What is a Subgraph?
2. What is an Induced Subgraph?
3. What does the half-hop induced subgraph tell you about selected nodes on a graph?
4. What does the one-hop induced subgraph tell you about selected nodes on a graph?
5. What is an ego-network?
6. Why are ego-networks useful?
7. Why are triangles important in social network analysis?
8. How many communities is the blue blazed node connected to?
9. What sort of words emerge in these communities?
10. How many communities is the red blazed node connected to?
11. What sort of words emerge in these communities?
12. How many communities is the yellow blazed node connected to?
13. What sort of words emerge in these communities?
14. Which node has the highest Common Neighbours score with the green blazed node?
15. How many neighbours do they have in common?
16. How many neighbours do they not have in common?
17. What is Jaccard Index?
18. Which node has the highest Jaccard Index score with the green blazed node?
19. How many neighbours do they have in common?
20. How many neighbours do they not have in common?
21. What is Levenshtein Distance?
22. Which words in the graph are linked by a Levenshtein Distance of 1 would you consider merging?

email-Eu-core Network (https://snap.stanford.edu/data/email-Eu-core.html)

The network was generated using email data from a large European research institution. We have anonymized information about all incoming and outgoing email between members of the research institution. There is an edge (u, v) in the network if person u sent person v at least one email. The e-mails only represent communication between institution members (the core), and the dataset does not contain incoming messages from or outgoing messages to the rest of the world.

The dataset also contains "ground-truth" community memberships of the nodes. Each individual belongs to exactly one of 42 departments at the research institute.

This network represents the "core" of the email-EuAll network, which also contains links between members of the institution and people outside of the institution (although the node IDs are not the same).

An attribute has been added indicating the number of departments an email is connected to.

Using email-Eu-core.star, answer the following questions:

23. What is Clustering?
24. What is a k-Truss?
25. What is the highest k-Truss value of the entire graph?
26. How many nodes are within this k-Truss?
27. How many departments are represented in this k-Truss?
28. What is Hierarchical Clustering? What is the process of creating clusters?
29. What is the key difference between Hierarchical Clustering and k-Truss clustering?
30. How many clusters are there at the optimal level for Hierarchical Clustering? How does this compare with the number of departments?
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?
32. Run the shortest paths between clusters – which of the selected nodes has the highest NeighbouringDepartmentCount value?
33. Run Betweenness Centrality on the graph and Scatter Plot Centrality.OutBetweenness with NeighbouringDepartmentCount. What is the relationship?

Les Miserables (https://gephi.org/datasets/lesmiserables.gml.zip)

Les Miserables: coappearance weighted network of characters in the novel Les Miserables. D. E. Knuth, The Stanford GraphBase: A Platform for Combinatorial Computing, Addison-Wesley, Reading, MA (1993).

Using Les_Miserables.star, answer the following questions:

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

Other Analytics

36. What is Dice Similarity?
37. What is Cosine Similarity?
38. What is the Resource Allocation Index?