

Networks and their Structure Assignment

Network Science Topic 3

Note that the networks in this exercise are all undirected.

We recall some definitions and introduce some new ones.

Let d_{ij} be the distance (the length of the shortest path) between vertices i to j .

Then the *closeness centrality* of vertex j is $CC(j) = \frac{1}{\sum_i d_{ij}}$.

The *nearness centrality* of vertex j is $NC(j) = \sum_i \frac{1}{d_{ij}}$. In both these definitions, the sums are over all vertices i , $i \neq j$, in the network.

The *degree centrality* of vertex j is simply its degree (the number of neighbours it has) and is denoted $DC(j)$.

The *adjacency centrality* of vertex j is $AC(j) = \frac{1}{d_j} \sum_i \frac{d_j - d_i}{d_j + d_i}$ where the sum is over all vertices i that are adjacent to j and d_i denotes the degree of a vertex. (So $DC(j)$ and d_j are different notations for the same measure.)

1. [5 marks] Calculate the values of the four centrality measures defined above on each vertex in the network below. (The diagram and the dictionary are two representations of the same network.) Present your answer as four lists --- one for each centrality measure --- that gives the vertices and the calculated values ordered by those values.

```
network = {1: [4],
           2: [4],
           3: [4],
           4: [1, 2, 3, 5, 6],
           5: [4],
           6: [4, 7, 8, 9, 10, 11],
           7: [6, 8, 11],
           8: [6, 7, 9, 11],
           9: [6, 8, 10],
           10: [6, 9, 11, 12],
           11: [6, 7, 8, 10],
           12: [10]}
```



1. [20 marks] Obtain the three datasets in topic3networks.zip (under Topic 3 on Learn Ultra, see the descriptions below). Load these networks. Again, they are all undirected. We wish also to work with connected graphs so find the largest connected component of each and discard other vertices.

For each dataset, for each of the four centrality measures, list, in order, the 20 vertices with the highest values of that measure (include more if the values are tied). Comment on whether you think, based on what you have found, that nearness centrality is a good alternative to closeness centrality and that adjacency centrality is a good alternative to degree centrality.

The datasets:

- london_transport_raw_edges.txt: The network is of London rail and underground stations that are linked if they are adjacent on some line. The second and third items on each line in the file are a pair of nodes that are joined by an edge (the first item describes how they are linked and can be ignored for this exercise).
- Roget.txt: This is a network of words that are linked if they appear together in a thesaurus. At the start of the file is a list of words (the nodes) and their numeric identifiers. Then there are lists (one per line) of words that appear together

in the thesaurus. There should be an edge between any pair of nodes that appear in the same list. For example, the list 3 4 323 325 implies the existence of six edges: (3,4), (3, 323), (3, 325), (4, 323), (4, 325), (323, 325)

- CCSB-Y2H.txt: The network is of interactions amongst proteins in yeast (living cells can be considered as complex webs of macromolecular interactions known as interactome networks). The first two items on each line are a pair of nodes joined by an edge (the rest of the line can be ignored).