MLRD Supervision 4

Neelu Saraswatibhatla (srns2)

Graph algorithms

Choose a vertex on the graph at random, and run a breadth-first search with this vertex as the start vertex, keeping track of the vertex furthest away from this one. If there are any vertices in the graph that haven't been visited then there are one or more disconnected vertices, so the diameter is infinity. If all vertices are in the graph, then find the first one that is furthest away from the start vertex. Then run another breadth-first search starting from this vertex, and the distance to the furthest away vertex is the diameter of the graph.

This is because in the first BFS, if a vertex that is the start or end point of longest shortest path on the graph was picked, then one on the other end will be found to be the furthest away, and the second BFS will find its way back to the first one (or one equally distant), finding the longest shortest path. If the vertex isn't on the longest shortest path, then the vertex found by the first BFS will be an endpoint of a longest shortest path, and the second BFS will once again find the distance of this path.

This does not work on graphs with cycles. For example, consider a diamond-shaped graph with 4 vertices (1, 2, 3, 4) and edges (1, 2), (1, 3), (2, 3), (2, 4), and (3, 4). Let the first BFS start at 2. All other the vertices are at a distance of 1, so let 3 be the one selected as the furthest away. Then the second BFS will find that all vertices are 1 away from vertex 3, so will incorrectly conclude that the diameter of the graph is 1, when in actual fact it is 2.

Betweenness centrality and Newman-Girvan method examples

1. (a) The betweenness centralities are as follows:

Node	**Betweenness centrality**
1	1.0
2	1.0
3	1.0
4	1.0
5	1.0

   (b) All 5 of the edges in this graph have the same betweenness centrality (3.0).

   (c) You get a completely disconnected graph, i.e. the 5 vertices on their own without any edges between them.

2. (a) The betweenness centralities are as follows:

Node	Betweenness centrality
1	6.0
2	0.0
3	0.0
4	0.0
5	0.0

(b) All 4 of the edges in this graph have the same betweenness centrality (4.0).

(c) You get a completely disconnected graph, i.e. the 5 vertices on their own without any edges between them.

3. (a) The betweenness centralities are as follows:

Node	Betweenness centrality
1	5.0
2	0.0
3	0.0
4	0.0
5	0.0

(b) The (1, 4) and (1, 5) edges have highest betweenness centrality (5.0). 4 ?

(c) We end up with one cluster and two lone vertices: [1, 2, 3] connected by the edges (1, 2), (2, 3), and (1, 3), and two lone vertices (4 and 5) disconnected from the rest of the graph, i.e. with no edges to/from either of them.

4. (a) The betweenness centralities are as follows:

Node	Betweenness centrality
1	0.0
2	3.0
3	0.0
4	3.0
5	0.0

(b) The (1, 2), (2, 4) and (4, 5) edges have highest betweenness centrality (4.0).

(c) We end up with one cluster and two lone vertices: [2, 3, 4] connected by the edges (2, 3) and (3, 4), and two lone vertices (1 and 5) disconnected from the rest of the graph, i.e. with no edges to/from either of them.

5. (a) The betweenness centralities are as follows:

Node	Betweenness centrality
1	0.0
2	3.0
3	4.0
4	0.0
5	0.0

(b) The (2, 3) edge has highest betweenness centrality (6.0).

(c) We end up with two clusters: [1, 2] (connected by the (1, 2) edge), and [3, 4, 5] (connected by the (3, 4), (4, 5), and (3, 5) edges).

6. (a) The betweenness centralities are as follows:

Node	**Betweenness centrality**
1	0.0
2	5.0
3	0.0
4	0.0
5	0.0

(b) The (1, 2) and (2, 5) edges have highest betweenness centrality (4.0).

(c) We end up with one cluster and two lone vertices: [2, 3, 4] connected by the edges (2, 3), (3, 4), and (2, 4), and two lone vertices (1 and 5) disconnected from the rest of the graph, i.e. with no edges to/from either of them.

7. (a) The betweenness centralities are as follows:

Node	**Betweenness centrality**
1	3.0
2	1.0
3	0.0
4	1.0
5	0.0

(b) The (1, 5) edge has highest betweenness centrality (4.0).

(c) We end up with one cluster and one lone vertex: [1, 2, 3, 4] connected by the edges (1, 2), (2, 3), (2, 4), (3, 4), and (1, 4), and one lone vertex (5) disconnected from the rest of the graph, i.e. with no edges to/from it.

Random graphs and metrics

Distribution of degree of nodes

In an Erdős–Rényi graph, every pair of edges has an equal probability of having an edge connecting it, and no node is any more likely to have an edge than another, so the degree of nodes in an Erdős–Rényi graph is expected to be uniformly distributed.

In a Watts-Strogatz graph, nodes start of connected around a circle and then may have these connections replaced by connections to other nodes, but each node has an equal likelihood of having an edge rewired to it, so the degree of nodes in a Watts-Strogatz graph is expected to be uniformly distributed.

In a collaboration network, there are some people who have collaborated with a lot of other people, and some people will have collaborated with far fewer people, and there is less likely to be an in between. Also, people are probably more likely to have collaborated with fewer people than more people. Therefore, the expected distribution of degree of nodes is a high number of nodes with low degree, very few with mid-range degree, and somewhere between these for high degree.

Number of connected components

In an Erdős–Rényi graph, there are likely to be several connected components a there are likely some pairs of nodes that have no paths between them. In a Watts-Strogatz graph, there are likely to be fewer connected components as it starts with a single connected component and then performs rewiring, which is less likely to result in broken paths than just randomly assigning edges. If $k >> ln(n)$ then it is guaranteed that the graph will be connected, i.e. there will be a single connected component. In a collaboration network, there is likely a single connected component as people in the scientific community likely collaborate with a widge range of people who collaborate with other people, and so on.

Distribution of shortest paths

In an Erdős–Rényi graph, the distribution of lengths of shortest paths is likely to be normally distributed as there will be very few pairs of edges that are directly connected and very few that are very distantly connected but most will be in between. This is also the case for Watts-Strogatz graphs. In a collaboration network, the normal distribution is likely skewed to the right as fewer people will have collaborated with a lot of people, as mentioned before.

low average

Extent of clustering

In an Erdős–Rényi graph, since all edges are equally likely to exist, you wouldn't expect any particular clustering. Of course some clustering would occur by chance but not a huge amount. This is also the case for a Watts-Strogatz graph, although slightly more clustering is expected as the graph starts off as a single connected component with no clustering but the first reqired edge introduces a small amount of clustering, which may increase but may also decrease back to nothing. In a collaboration network, a large amount of clustering is expected, as people are likely to collaborate with other people within their own university, country, or field, with some links to scientists abroad or in other fields.

no triadic

Watts-Strogatz → high clustering