

Coursework 2 – Network Analysis

| | |
|------------------|---|
| Student: | Petur Einarsson |
| ID: | |
| Date: | 1. March 2017 |
| Dataset: | http://konect.uni-koblenz.de/networks/sociopatterns-infectious |
| Network name: | Infectious |
| Number of nodes: | 410 |
| Number of edges: | 17,298 (multiple edges between two nodes) |
| Unique Edges: | 2,765 |
| Network type: | Undirected |
| Software: | R (igraph package) |

Analysis of network characteristics

The dataset chosen was the INFECTIOUS dataset, which describes the face-to-face behaviour of people during the exhibition “INFECTIOUS: STAY AWAY” in 2009 at the Science Gallery in Dublin. Nodes are exhibition visitors and edges are face-to-face contacts active for at least 20 seconds. The network is partially connected, as all nodes are connected to at least one other node. Multiple edges between two nodes were possible, but those were removed so only the unique contacts remain.

A random undirected graph with the same number of nodes and edges was then generated following an Erdos-Renyi model, which is binomial.

Network visualization

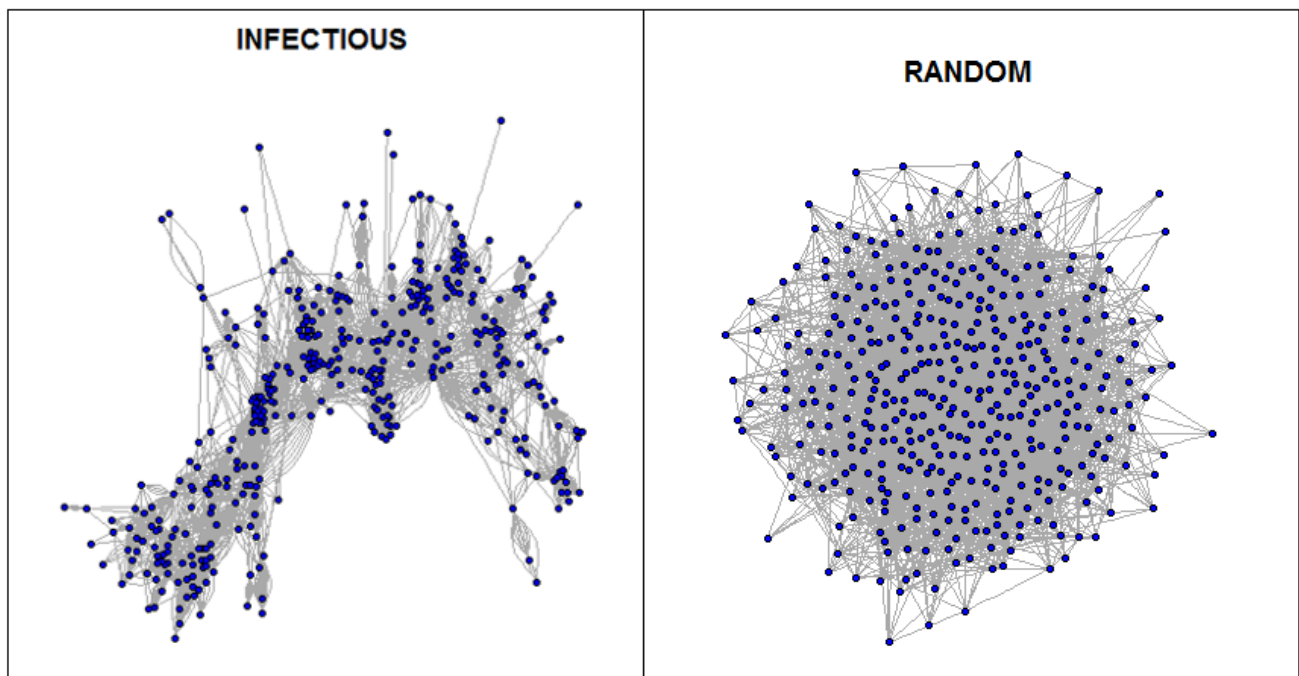


Figure 1: Network visualization of "Infectious" dataset and an equally sized randomly generated dataset

Degree distribution

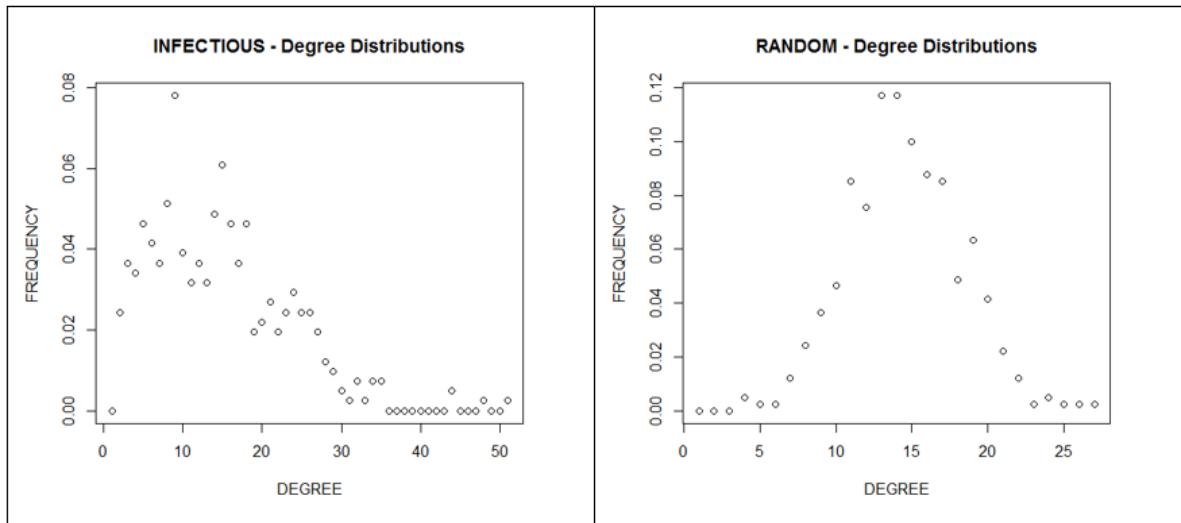


Figure 2: Degree distribution of the two datasets

The degree of a node, shows the number of edges that node has to other nodes.

The INFECTIOUS data has a positive skew, as opposed the Gaussian bell-curve that the random dataset has. This means that most people attending the exhibition made only a few connections while few made many connections. Interestingly the two datasets have the same median and mean number of edges, while the minimum and maximum edges made are very different. This indicates that the INFECTIOUS network is less homogenous than an equally sized random one.

| Infectious | | | Random | |
|------------|-------|--|---------|-------|
| Min. | 1.00 | | Min. | 4.00 |
| 1st Qu. | 7.00 | | 1st Qu. | 11.00 |
| Median | 13.00 | | Median | 13.00 |
| Mean | 13.49 | | Mean | 13.49 |
| 3rd Qu. | 19.00 | | 3rd Qu. | 16.00 |
| Max. | 50.00 | | Max. | 28.00 |

Figure 3: Summary statistics of degree distribution

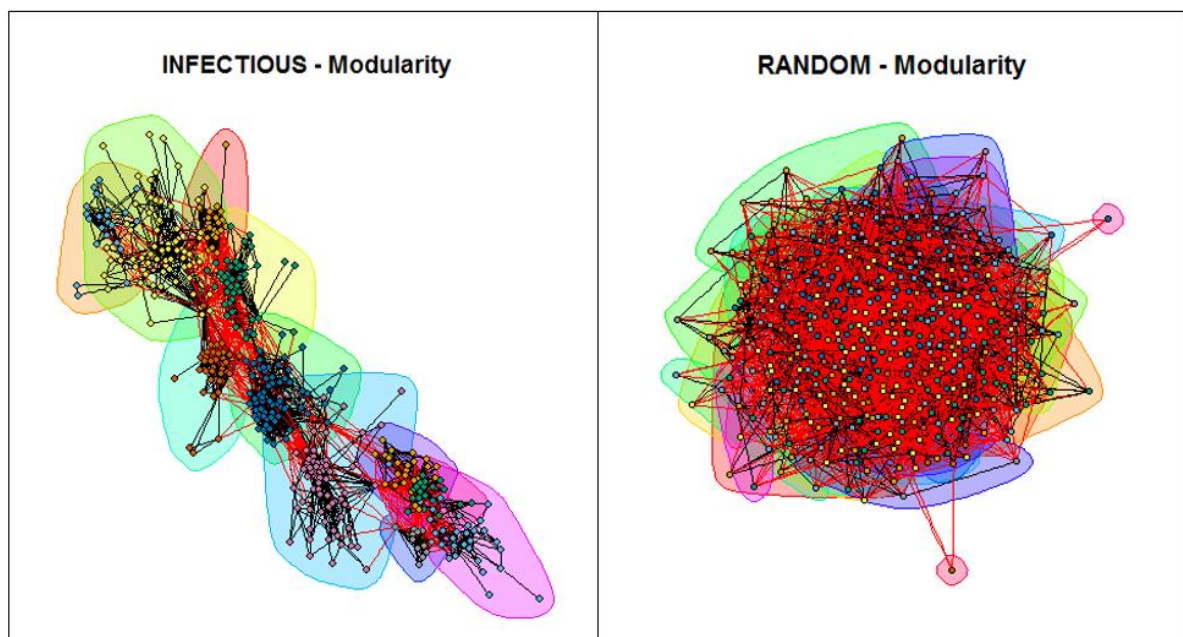
Clustering coefficient

Infectious coefficient: 0.4357

Random coefficient: 0.0334

The clustering coefficient measures the probability to which nodes in a graph tend to cluster together. In other words, how likely is it that nodes that talk to a single node, also talk to each other. There is a 44% chance that nodes cluster together in the INFECTIOUS network. The probability of this connection in the random dataset is much lower, or only 3%.

Modularity

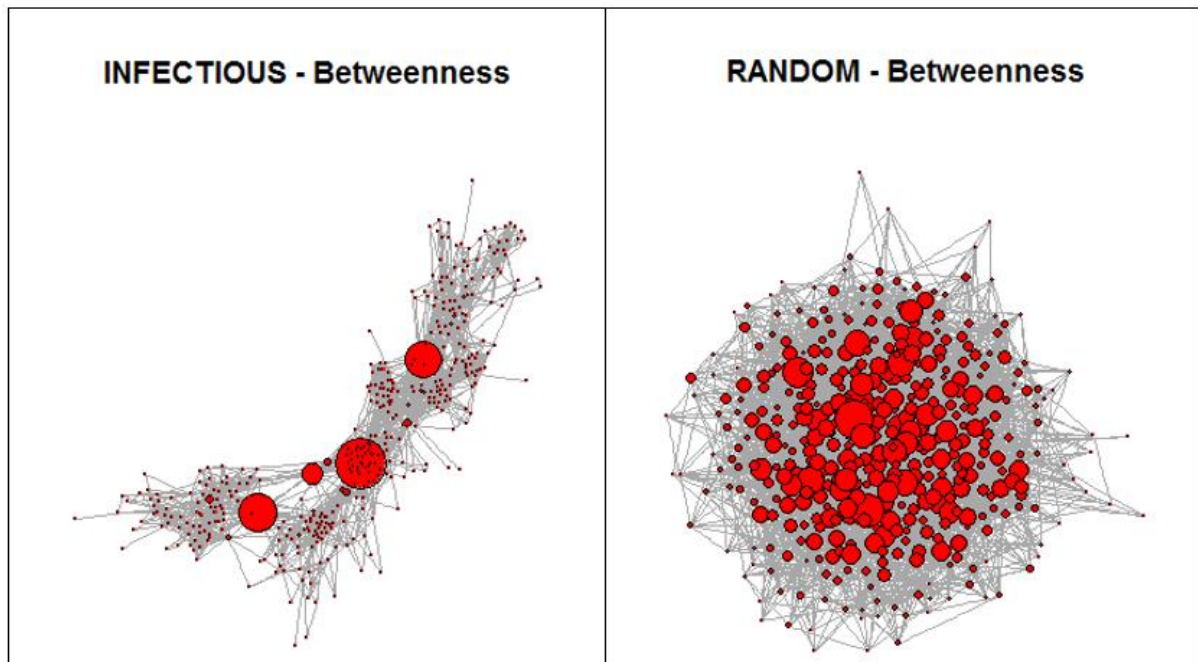


| Infectious | Random |
|------------|--------|
| 0.6848 | 0.2074 |

Modularity of a graph divided in some way measures how good the division is, or how separated are the different vertex types from each other. It measures the structure of that division. A high modularity shows dense connections within the communities of a graph, while a low modularity means it's less dense. Although modularity can easily detect larger communities, it has a hard time detecting smaller ones.

The INFECTIOUS network shows to have much denser community than an equally sized random one.

Centrality (betweenness)



Betweenness metrics INFECTIOUS

| Min. | 1st Qu. | Median | Mean | 3rd Qu. | Max. |
|------|---------|--------|--------|---------|----------|
| 0.00 | 32.31 | 167.60 | 538.00 | 441.70 | 22130.00 |

Betweenness metrics RANDOM

| Min. | 1st Qu. | Median | Mean | 3rd Qu. | Max. |
|-------|---------|--------|--------|---------|---------|
| 10.66 | 201.40 | 296.60 | 326.90 | 429.70 | 1109.00 |

Betweenness of a node, is the number of nodes that have to go through it in order to get to every other node in the network. The higher the betweenness of a node, the more nodes pass through it. A node with a higher betweenness has more control in a network.

The infectious network is dominated by just four nodes that have a much higher betweenness than the rest. These individuals are most likely employees in the art gallery or the artists themselves. The randomly generated data is highly connected and many nodes with a high betweenness.