

Hospital Network Analysis Report:

A) Sample Selection:

- We took major cities of Indian and collected name of all the hospitals in those cities. We also collected the data such as number of beds and number of Doctors in each hospital. These parameters were used to classify hospitals and we were able to select a sample size of top 140 hospitals of India. First assumption taken was all the country's major hospital lies in these major 16 metropolitan cities. After obtaining the data of all the hospitals, we sorted the list of hospitals on the basis of number of doctors. We set our threshold to a minimum of 15 doctors and 100 beds. Ranking of the selected hospitals was done on the basis of number of doctors solely.
- About 140 hospitals remained by the provided sorted process. We further assigned the hospitals with the unique hospital id based on their ranks. Both the minimum number of doctors criteria and beds criteria provided us with a clean sample space. Our assumption of having a lower limit on the number of doctors if justified as follows, hospitals with less doctors and more number of beds were most likely to be a general ward type hospital with less facilities and hence, should not be classified among the top hospitals of the country. On the other hand, a hospital with large number of doctors but less beds were clearly lacking the infrastructure to compete among the best.
- Our assumption to consider only 16 major cities is justified by the fact that these were all the cities who made the hospital data to be public online. All the other efforts to find any such data for other cities resulted in no data at all or violating the above criteria. Hence, it is safe to assume the validity of our sample data.

B) Data Collection:

- We wrote a script on python to collect the data of the hospitals for these 16 cities. For the source, we used the website: www.practo.com. After collecting the data of number of doctors and bed, we sorted them using the above criteria. Now, we have created our sample size.
- Another script on python helped to scrap all the details like Name, Age, Degree etc. of the doctors of the hospitals belonging to our sample size. The website www.practo.com was enough for both tasks. We manually confirmed the validity of the data for various hospitals from their individual websites.
- Despite all the sources, we were not able to find data for the executives in a common center such as data for the doctors. We searched multiple sources such as Bloomberg, Zaubas Corp etc. and even to the individual hospital websites in many sources. Any hospital data listed in multiple data was compared and then added to ensure data integrity.
- The use of Electronic Medical Record (EMR) was another parameter that we investigated. The hospitals using EMR generally require UHID for patients to register. Hospitals providing online report (the CDA CCD), maintaining the medicine database or disease classification (WHO ICD 10) generally implied the presence of some sort of EMR. These factors were searched for on the websites of each hospital manually to judge the presence of EMR.

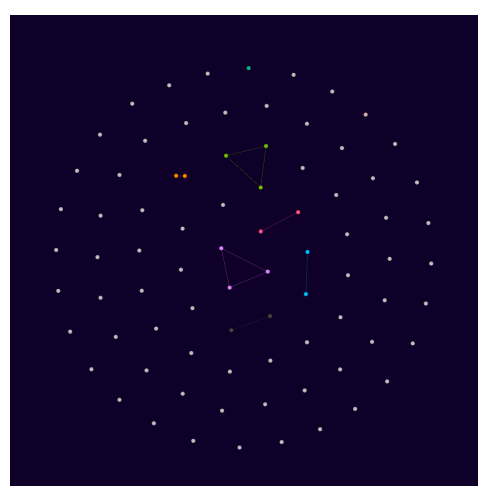
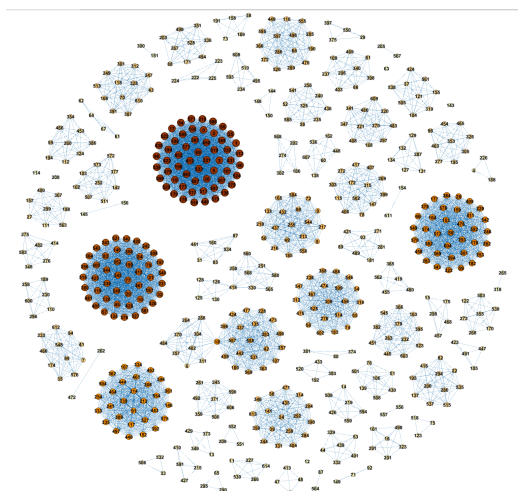
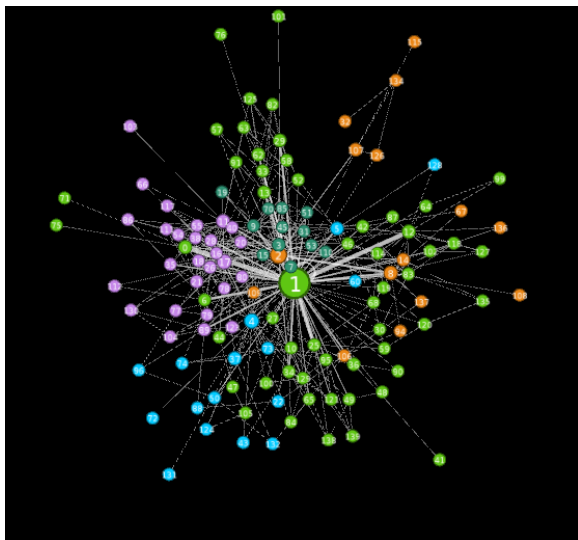
C) Methods used for the analysis:

- First, we needed to give every unique executive, hospital and doctor name with a unique ID which could be used to create networks. We used NetworkX library on Python3, and created four graphs using all the data. First graph consisted of network between hospitals as its node and doctors as its edge. Second graph consisted of network between doctors as its node and hospitals as its edge. Similarly, third and fourth graph with hospital and executives as nodes respectively and vice-versa for the edges were used for the networks. All the variables calculated are listed below.
Note that some of the variables are derived from the standard variable while others are derived from the standardization made in David paper (2003).

Parameters	Formulae Used
Degree Centrality	$C_D(u) = \frac{\deg(u)}{(n-1)}$
Closeness Centrality	$C_C(u) = \frac{(n-1)}{\sum_{v=1}^{n-1} d(v,u)}$
Betweenness Centrality	$C_B(v) = \sum_{s,t \in V} \frac{\sigma(s,t v)}{\sigma(s,t)}$
K (interlocks per nodes)	$\frac{\text{No. of Edges}}{\text{No. of Nodes}}$
L_{Actual}	Average short path length between largest connected component.
L_{Random}	$\frac{\ln(n)}{\ln(k)}$
C_{Actual}	Average degree of the local clustering of the largest connected set.
C_{Random}	$\frac{k}{n}$
SW	$\frac{C_{\text{Actual}}}{L_{\text{Actual}}} \times \frac{L_{\text{Random}}}{C_{\text{Random}}}$

D) Results:

The Visualization of all the graphs in the listed order is given below:



Graph 1	
Degree Centrality(*)	0.25, 0.164, 0.128, 0.128, 0.128
Closeness Centrality(*)	0.406, 0.383, 0.369, 0.365, 0.35
Betweenness Centrality(*)	0.157, 0.106, 0.085, 0.074, 0.058
K (interlocks per nodes)	2.688
L_{Actual}	3.137
L_{Random}	5.005
C_{Actual}	0.022
C_{Random}	0.019
SW	1.852

Graph 2	
Degree Centrality(*)	0.124, 0.109, 0.104, 0.102, 0.101
Closeness Centrality(*)	0.4, 0.38, 0.379, 0.377, 0.376
Betweenness Centrality(*)	0.157, 0.106, 0.085, 0.074, 0.058
K (interlocks per nodes)	64.91
L_{Actual}	3.488
L_{Random}	2.114
C_{Actual}	0.971
C_{Random}	0.009
SW	61.5

(*): Top 5 values of the data provided.

Both Table one and Table 2 shows that both the network has SW value greater than 1. Hence, we come to the conclusion that Small world applies to both the networks. The analysis for the last two networks is not done as it is evident from the graph that the networks lack connections for being called in-fact a network. Hence, to conclude:

1. The hospitals network is small world in reference to the doctors but lacks a proper network in terms of executives.
2. The doctors network is small world in reference to the hospitals but executives lack a proper network in terms of hospitals.

E) Future Directions:

We could try to relate the data obtained for the EMR to the network created. Network could also be used to find any relation between the revenue of the hospitals and the doctors in common, as the same doctors with high status and experience are likely to work in a bigger hospital. A lack of the network through executives could be used to understand the difference in the management decisions made by the hospitals. Other variables such as number of beds and number of doctors could be used further strengthen the claims made.