Katholieke
Universiteit
Leuven

**Master in
Artificial Intelligence**

DOLPHIN SOCIAL NETWORK ANALYSIS

Analysis of Large Scale Social Network: Project Report

Sarah Binta Alam Shoilee (r0729856)

Academic year 2018–2019

1 Introduction

We have a dataset which contains undirected social network of frequent associations between 62 dolphins in a community living off Doubtful Sound, New Zealand. This dataset is connected, undirected and unweighted. While analyzing this graph dataset, we tried to find out which centrality measures best describe our dataset and how. This centrality measures directs us to label the node with more influence within the community in the community and their role in persisting the network cohesion. We tried to illustrate, in case of death of any dolphin what is the likelihood that the community still remains connected.

2 Result

Our chosen dataset contains 62 nodes of bottlenose dolphin and they have 159 edges. According to David Lusseau, the edges represent as preferred companionship of dolphins. By analyzing the centrality measure, we can say that these bottlenose dolphins lives in a very tightly knit group with an average connectivity of 5.129032.

To understand degree distribution of given dataset, we tried to analyze it's similarity both with Poisson Distribution and power-law distribution of degree(k). Our result suggests that we cannot rule out the possibility of the degree of individual follows a Poisson distribution. The result from goodness of fit test with a Poisson distribution is 16.35 with p value 0.13. We also run this same test on power-law distribution for the tail of the degree distribution$whendegree\ \bar{7}$ **as suggested by David Lusseau in his paper[1]. However, we do not get any acceptable p value for over a range of exponential value.**

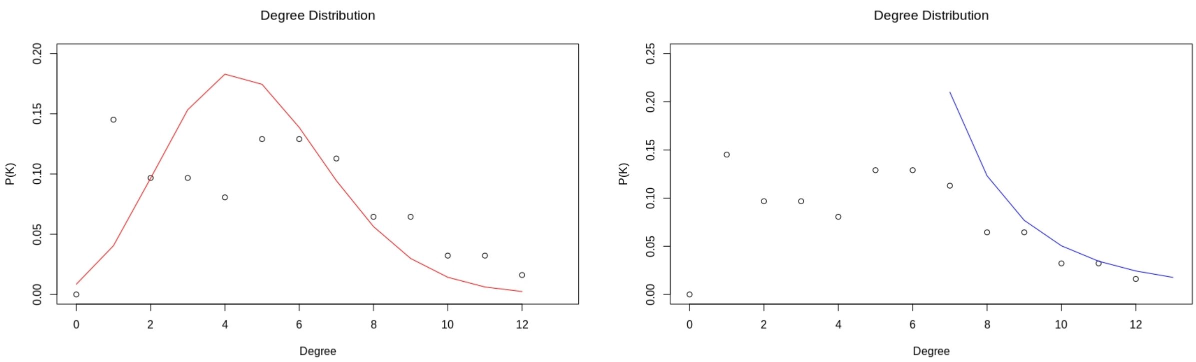
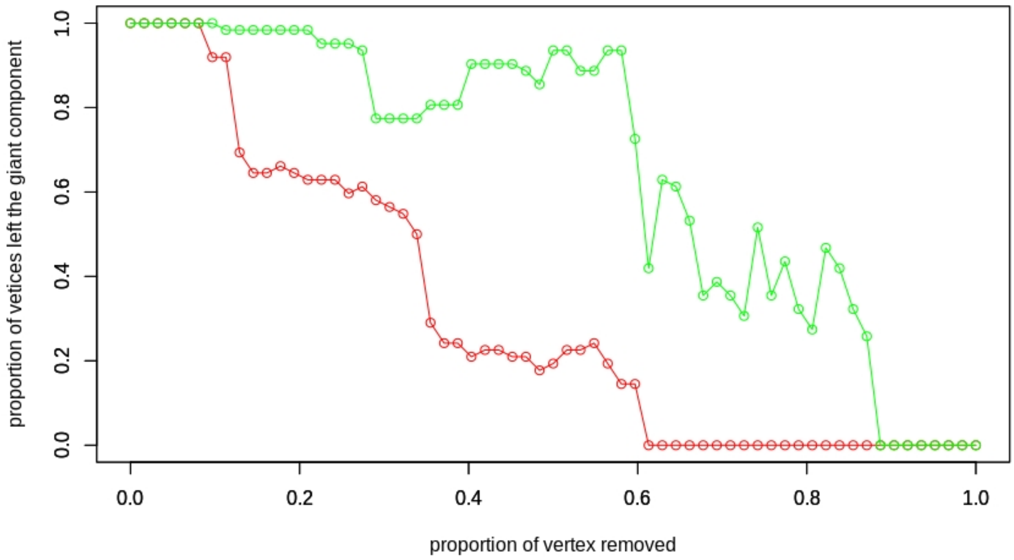


Figure 1: Degree distribution of Dolphin dataset. (a)Compares the distribution with Poisson Distribution; (b)Compares the distribution function with power-law exponential function where degree$k$ is greater or equal to 7

By analyzing different centrality measures like degree, betweenness, average connectivity, clustering-coefficient and path length, we presume the "betweenness" best describes out data set. We compare and analyze our finding with respect to [1],[2]. Since betweenness is more influential than other centrality measure for our given situation, not surprisingly, Eigen value clustering also turns out to provide best community clustering and efficiently manage to identify the brokerage nodes.

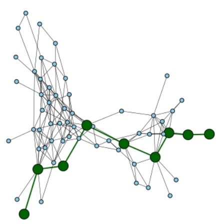
Our result suggests that the dolphin network is very persistent towards the random removal of nodes. It tends to persist cohesion up to 13% removal of vertices following it's betweenness decreasing order . If we remove 30% percent nodes based on their decreasing order of degree to lose it's cohesion with entire community. Figure 2, illustrates the graphical representation of correlation between cohesion persistence and vertex removal.

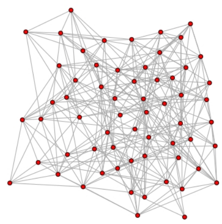


removal.jpeg

Figure 2: Represents the correlation between cohesion persistence and vertex removal. The red line symbolizes the correlation with respect to vertex removal in decreasing order of betweenness and the green line represents the removal of vertices in degree decreasing order

3 Discussion





world.jpeg

Figure 3: (a)Dolphin Community representation with Fruchterman-Reingold vizualization; The green line spot the diameter of the network (b)Small-world network with degree=6

Our dolphin community is described by an undirected and unweighted graph containing 62 Bottlenose Dolphins. Each edge of this graph indicated the connection between two dolphin represented as "source" and "destination" is data set. Like friendship in human community network the direction of dolphin companionship is not existent. The connected community of dolphins ensures its information flow within the whole network with a diameter 8 and the average shortest path for information traversal 3.3. By observing its average connectivity(5.13) and its average neighbourhood size(6.13), we presume this network to be well-connected.

A common assumption is graph theory is that the nodes with higher degree tends to form bond with most other nodes and eventually ends up creating a long tail in degree distribution(Barabasi-Albert model). Figure 1(b) suggests no such sign of power-law degree distribution. Though it creates a short tail after degree 7, by comparing its correlation with exponential power-law function, we end up by rejecting the hypothesis of being a scale-free network. Rather, it suggests more to maintain a small word phenomenon with "8 degree of separation"

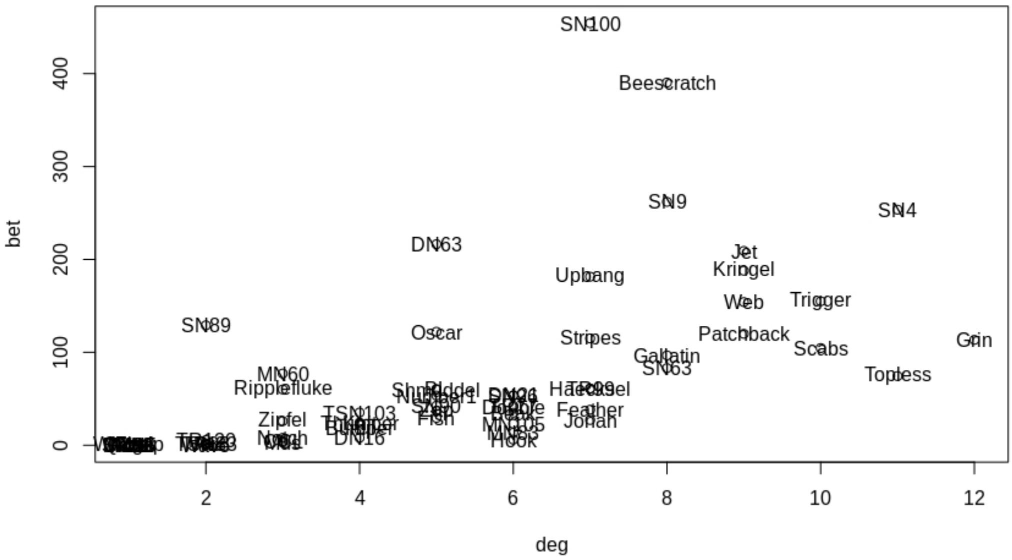


Figure 4: Represents the correlation between degree and betweenness

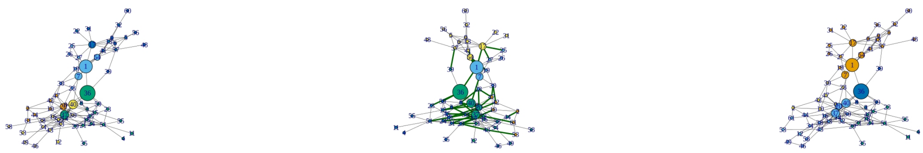


Figure 5: Represent (a)Eigne Value Cluster (b)Louvain Cluster (c)Info Map Cluster; The size of the node is proportional to its betweenness when its betweenness frequency is less than or equal to 13%

If we plot degree distribution of dolphins and analyze it we can notice that dolphin Grin is connected to 12 other dolphins which is the most connected animal in the dataset while around 9 dolphins are connected to only one other dolphin. If we check in terms of betweenness dolphin SN100 has highest betweenness. We find poor correlation between degree and betweenness(0.59). Figure 2 also shows that if we remove vetices with more betweenness it effects more to the community than removing vetices with high degree. From our analysis, we can conclude that in order to measure centrality within the community, we should put more priority on betweenness than degree. To represent our graph, we are using Fruchterman-Reingold approach which is a force directed graph visualization method and also favour betweenness.

We tried three clustering approachs to find out the brokerage node. According to paper[3], the brokerage node within Dolphin community is typically the nodes those fall in the boarder of the sub-community and with high betweenness. Within the dolphin community this nodes acts like a prime source of information within the network. Due to having linkage to more than one community and strong connections within their own community, generally "brokers" have more influence within the sub-community. For spectral clustering, Louvain clustering and clustering using map equation, the modularity is given respectively 0.49, 0.52 and 0.53. Though all three approachs provide highly similar performance in terms of modularity, leading eigen value clustering provides us the best result in reference to brokerage node defination. In figure 5(a) for leading eigen value clustering, we can see that the nodes with higher betweenness falls in the boundary of the sub-community and therefore detected as "broker".

References

[1] D. Lusseau(2003) *Proceedings of the Royal Society of London Series B* 270, S186- S188

[2] D. Lusseau, M. E. J. Newman(2004) *Proceedings of the Royal Society of London Series B* 271, S477-S481.

[3] D Lusseau (2007) *Evolutionary ecology* 2007, Springer,