श्रद्धावान् लभते ज्ञानम्

AMRITA VISHWA VIDYAPEETHAM

DEEMED TO BE UNIVERSITY UNDER SECTION 3 OF THE UGC ACT, 1956

DEPARTMENT OF MATHEMATICS

Int. M.Sc. DATA SCIENCE

COIMBATORE CAMPUS (INDIA)

(APRIL 2021)

18CSC312
GRAPH ANALYTICS AND ALGORITHM
CASE STUDY REPORT

Dolphin Social Network

Done by

Linci Felicita P - CB.SC.I5DAS18022

Parameshwari S - CB.SC.I5DAS18026

Outline

Abstract

The study of social networks is a new but quickly widening multidisciplinary area involving social, mathematical, statistical, and computer sciences. To better address questions concerning animal sociality, animal behaviorists and behavioral ecologists are increasingly turning to the suite of analytical techniques known as social network analysis (SNA). SNA allows for the quantification of multi-actor interactions, thereby providing a more realistic representation of social patterns and relationships. Here, we provide a detailed analysis of the social network of bottlenose Dolphins using some measuring techniques like Centralities.

Introduction

Dolphins have long-lasting friendships and form cliques while shunning other groups. Ecological constraints, such as the availability of prey, could explain the inclusion of older-looking dolphins into the social groups. These animals may possess long-term knowledge needed to tackle such constraints and thus play a key role in their community, when it comes to making friends, it appears dolphins are just like us and form close friendships with other dolphins that have a common interest. Bottlenose dolphins have close bonds that last for years based on common interests. These dolphins are unique in their use of marine sponges as foraging tools; they are the only ones to ever have been observed doing this. The method is taught by mothers to calves and helps the "spongers," as the ones who do it are called, find food in deeper water. Both male and female dolphins can be spongers. Spongers spent more time with other spongers and that the bonds were based on similar foraging techniques and not relatedness or other factors.

Data Description

The Dolphin social network dataset used in this case study was provided by David Lusseau and contains an, “undirected social network of frequent associations between 62 dolphins in a community living off Doubtful Sound, New Zealand.” The nodes in the network each represent a member of a bottlenose dolphin community living off Doubtful Sounds in New Zealand. An edge exists between two nodes if there is a frequent association between the dolphins represented by those nodes. The observations were gathered between 1994 and 2001.

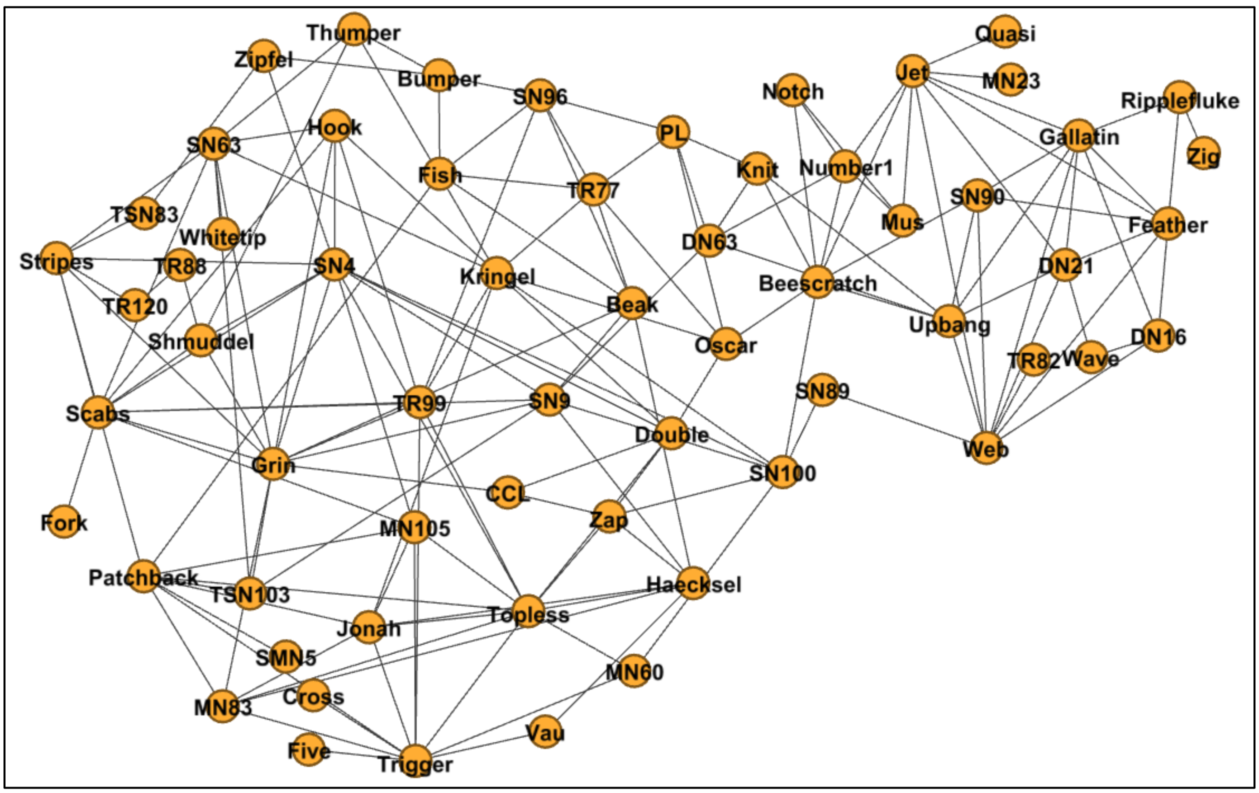

Figure 1: Social network of bottlenose dolphins

Importance of this graph

This graph depicts the social network of the bottlenose dolphin community. Community structure detection is of great importance because it can help in discovering the relationship between the function and the topology structure of a network. Interactions between individuals generate a social environment at the population level which in turn selects for behavioral strategies at the individual level. A social network is often a perfect means by which to represent heterogeneous relationships in a population. This graph helps to monitor the relationship among bottlenose dolphins.

Motivation

There are several studies on human social networks, but there is still an amusement when it comes to animal social networks. Dolphins also have a social network that is similar to that of human beings. Studies have proved that Dolphins learn new skills from their fellow dolphins and not just from their mothers. Dolphins not only acquire new ways to catch prey but are also motivated to learn from peers. Learning from others allows for a rapid spread of novel behaviors across populations, and it has been suggested that species with the capacity for learning from others in this way may be better able to survive. This fact seemed to be quite inquisitive and hence our decision to analyze the social network of bottlenose dolphins.

Network Statistics

Nodes - 62

Edges - 159

Maximum degree - 12

Minimum degree - 1

Average degree - 5.12

Density - 0.084

Chromatic number - 6

Eccentricity

Eccentricity of a node is the maximum of the distance of any other node in the graph from that node. Maximum eccentricity is the diameter and minimum eccentricity is the radius of the graph. Centers are the nodes with minimum eccentricity.

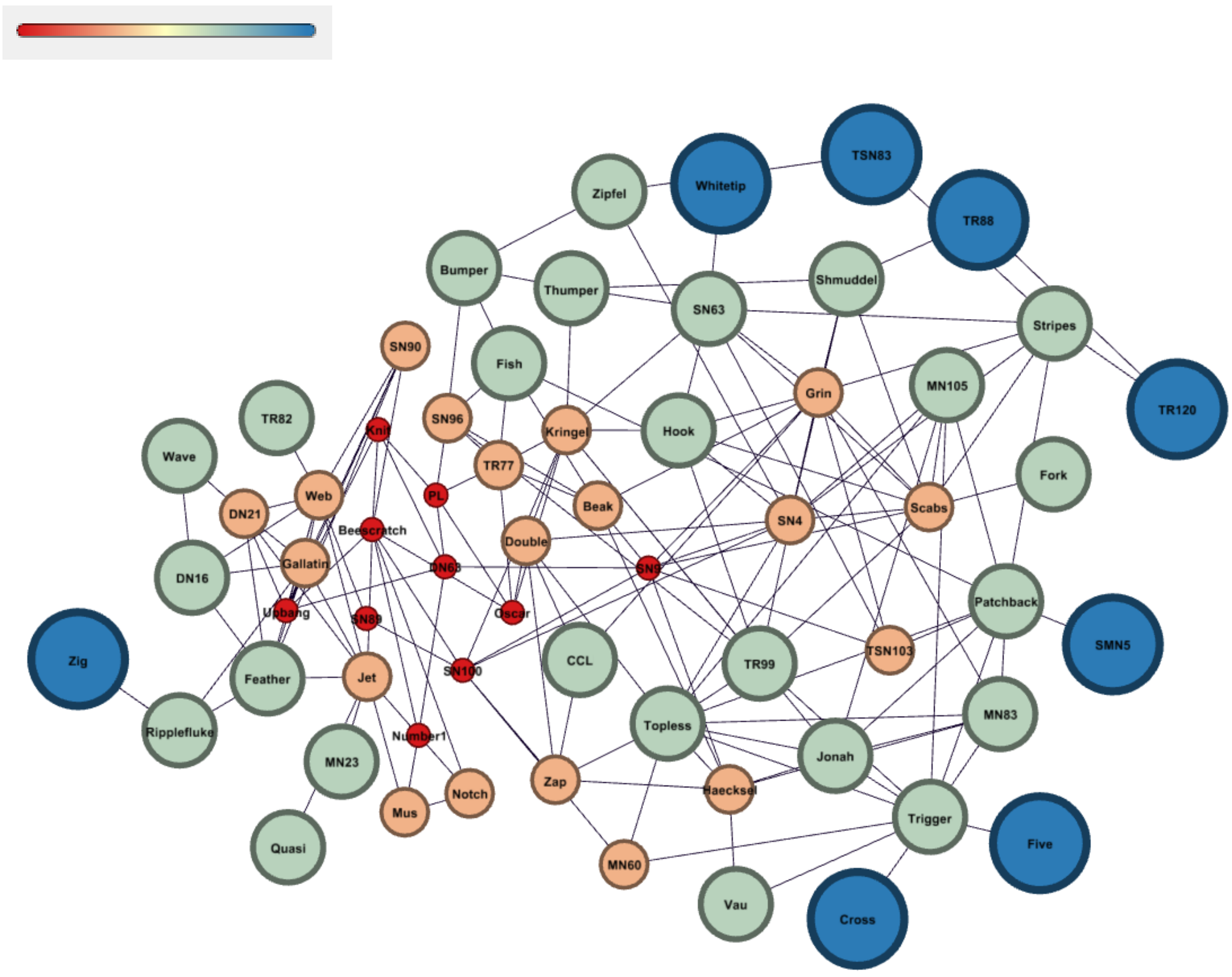

Figure 2: Eccentricity graph of bottlenose dolphin social network - node color and size according to the value of eccentricity.

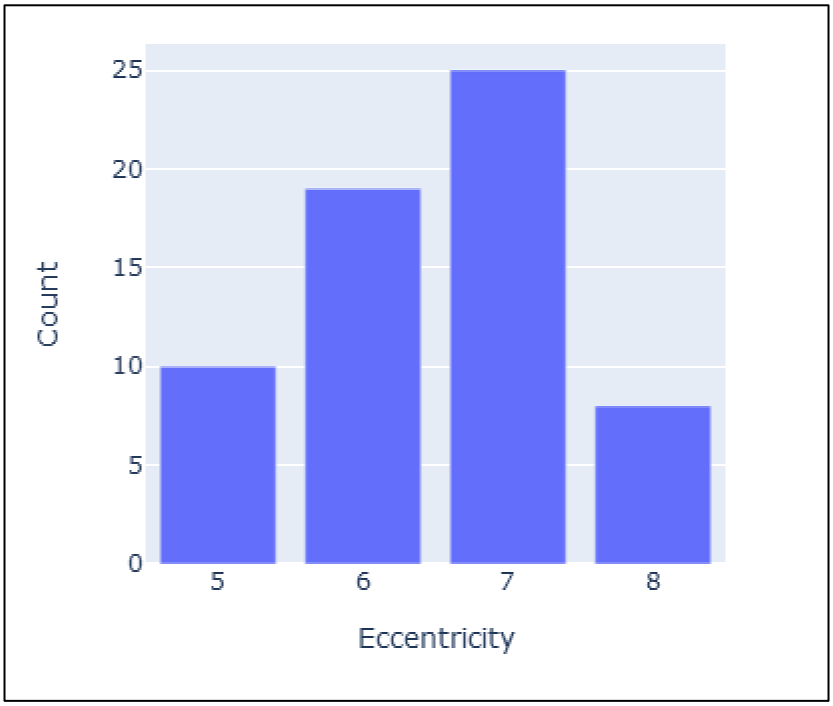

Figure 3: Bar graph representing the distribution of eccentricity

Eccentricity of each node -

- *Beescratch, DN63, Knit, Number1, Oscar, PL, SN100, SN89, SN9 and Upbang* – **5**
- *Beak, DN21, Double, Gallatin, Grin, Haecksel, Jet, Kringel, MN60, Mus, Notch, Scabs, SN4, SN90, SN96, TR77, TSN103, Web, Zap* – **6**
- *Bumper, CCL, DN16, Feather, Fish, Fork, Hook, Jonah, MN105, MN23, MN83, Patchback, Quasi, Ripplefluke, Shmuddel, SN63, Stripes, Thumper, Topless, TR82, TR99, Trigger, Vau, Wave, Zipfel* – **7**
- *Cross, Five, SMN5, TR120, TR88, TSN83, Whitetip, Zig* – **8**

Radius - 5

Diameter - 8

Centre - *Beescratch, DN63, Knit, Number1, Oscar, PL, SN100, SN89, SN9, Upbang*

Community

Many animals are forming communities. Communities have the advantage that they offer protection against predators, a place for finding a sexual partner or as well they offer the ability to apply hunting techniques which just work in a group. From the visual impression we can think about to divide our network in at least two communities.

For finding communities we apply the greedy algorithm. The basic idea is that we start with n communities, where n is equal to the number of nodes in the network. In each step we merge two communities together which have a maximum modularity. The greedy algorithm converges when there is just one single community. Then we take the value where the modularity was at the maximum and we know the number of communities.

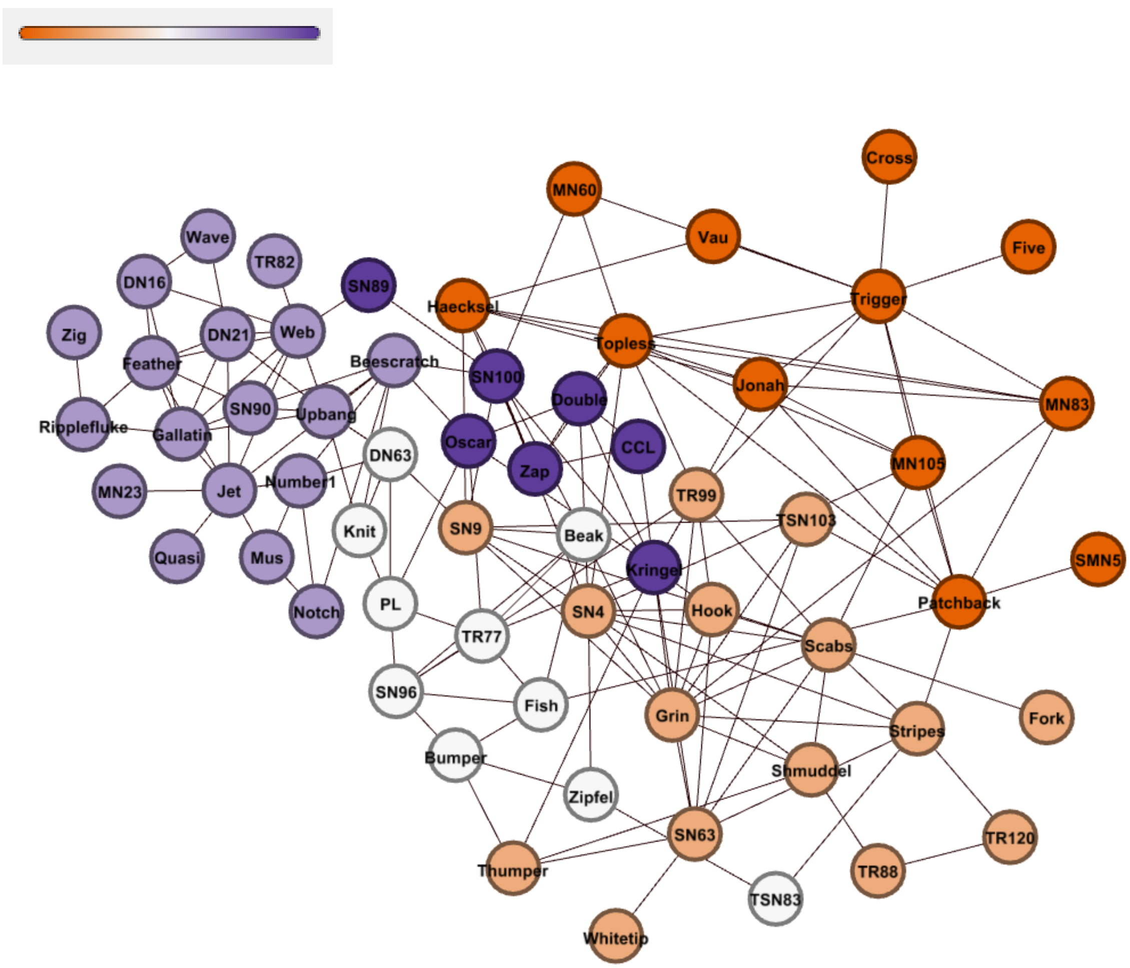

Figure 4: Community network graph of the bottlenose dolphin social network - color represents the community where each individual belongs

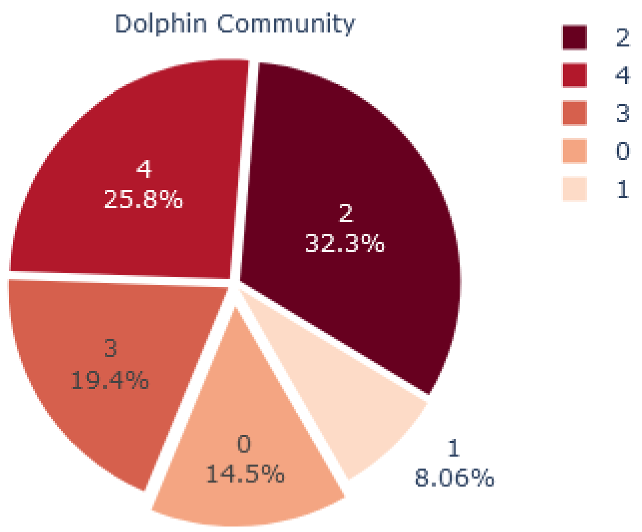

Figure 5: Pie chart representing the percentage of dolphins present in each community

Inference:

- There are 5 communities in this Dolphin social network where Dolphins exhibit similar behavior within communities and less similar behavior between communities.
- Community 2 has a greater number of dolphins which is 20 (32.3%) followed by community 4 which has 16 (25.8%) dolphins.
- Community 1 has the least number of dolphins which is 5 (8.06%)

Centrality

In graph analytics, Centrality is a very important concept in identifying important nodes in a graph. It is used to measure the importance (how central a node is) of various nodes in a graph. There are various centrality measures:

1) Degree Centrality

Degree centrality is defined as the number of links incident upon a node. If the network is directed, then two separate measures of degree centrality are defined, namely, indegree and outdegree.

Id	Label	Degree Centrality	Normalized degree centrality
14	Grin	12	0.196721311
37	SN4	11	0.180327869
45	Topless	11	0.180327869
33	Scabs	10	0.163934426
51	Trigger	10	0.163934426
17	Jet	9	0.147540984
20	Kringel	9	0.147540984
29	Patchback	9	0.147540984
57	Web	9	0.147540984
1	Beescratch	8	0.131147541
13	Gallatin	8	0.131147541
38	SN63	8	0.131147541
40	SN9	8	0.131147541
9	Feather	7	0.114754098

15	Haecksel	7	0.114754098
18	Jonah	7	0.114754098
36	SN100	7	0.114754098
43	Stripes	7	0.114754098
50	TR99	7	0.114754098
54	Upbang	7	0.114754098
0	Beak	6	0.098360656
6	DN21	6	0.098360656
8	Double	6	0.098360656
16	Hook	6	0.098360656
21	MN105	6	0.098360656
24	MN83	6	0.098360656
42	SN96	6	0.098360656
47	TR77	6	0.098360656
7	DN63	5	0.081967213
10	Fish	5	0.081967213

27	Number1	5	0.081967213
28	Oscar	5	0.081967213
30	PL	5	0.081967213
34	Shmuddel	5	0.081967213
41	SN90	5	0.081967213
59	Zap	5	0.081967213
2	Bumper	4	0.06557377
5	DN16	4	0.06557377
19	Knit	4	0.06557377
44	Thumper	4	0.06557377
52	TSN103	4	0.06557377
3	CCL	3	0.049180328
23	MN60	3	0.049180328
25	Mus	3	0.049180328
26	Notch	3	0.049180328
32	Ripplefluke	3	0.049180328

61	Zipfel	3	0.049180328
39	SN89	2	0.032786885
46	TR120	2	0.032786885
49	TR88	2	0.032786885
53	TSN83	2	0.032786885
55	Vau	2	0.032786885
56	Wave	2	0.032786885
4	Cross	1	0.016393443
11	Five	1	0.016393443
12	Fork	1	0.016393443
22	MN23	1	0.016393443
31	Quasi	1	0.016393443
35	SMN5	1	0.016393443
48	TR82	1	0.016393443
58	Whitetip	1	0.016393443
60	Zig	1	0.016393443

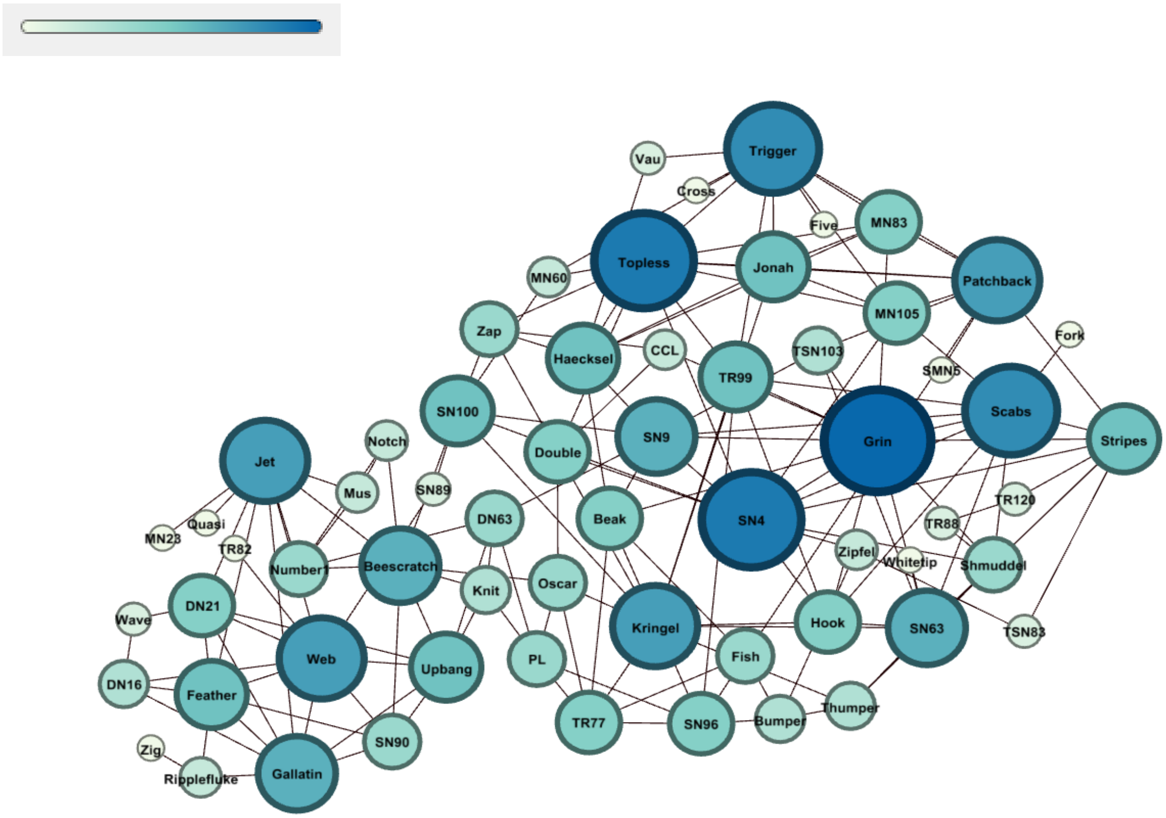

Figure 6: Degree centrality graph of dolphin social network - node color and size according to number of node incidents on it.

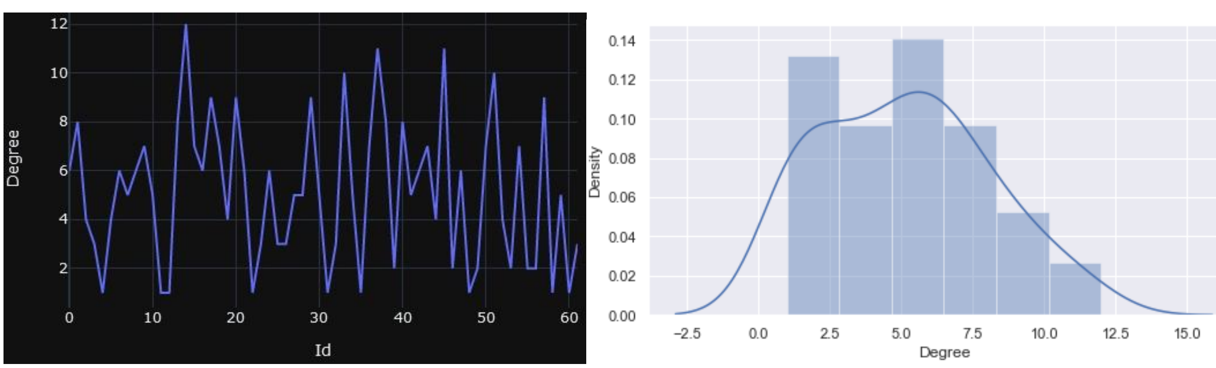

Figure 7: Line chart and histogram for degree centrality

Inference:

- By looking at its size and color we can say that *Grin* has the highest degree, meaning that it is connected to many other dolphins, it has more friends compared to other dolphins. We can also observe that Grin has a very good social relationship.
- Following Grin, we have *SN4, Topless, Scabs, Trigger, Kringel* and *Jet* who have higher degrees, meaning that these dolphins also have good social relationships as they are connected to many other dolphins.
- *Cross, Five, Fork, MN23, Quasi, Whitetip, Zig, SMN5* and *TR82* have very less node size and lighter shade which means their degree is very low. These dolphins are not well connected with other dolphins. They have poor social relationships.
- Degree is not normally distributed and by looking at the histogram we can say degree is a bimodal function that is they have more than one mode.
- It has two separated classes or intervals equally representing the maximum density of the distribution.

2) Betweenness Centrality

Betweenness centrality is a way of detecting the amount of influence a node has over the flow of information in a graph. It is often used to find nodes that serve as a bridge from one part of a graph to another. The betweenness of a node is measured by counting how often that node is frequented when travelling between all possible pairs using shortest paths.

Id	Label	Betweenness centrality
36	SN100	454.274069
1	Beescratch	390.383717
40	SN9	261.963619

37	SN4	253.582713
7	DN63	216.376673
17	Jet	209.169298
20	Kringel	187.841704
54	Upbang	181.392614
51	Trigger	154.959376
57	Web	154.094571
39	SN89	129.045705
28	Oscar	122.165227
29	Patchback	119.918587
43	Stripes	114.980006
14	Grin	113.408769
33	Scabs	104.614585
13	Gallatin	96.708781
38	SN63	82.994597
23	MN60	77.194498

45	Topless	74.426906
50	TR99	61.142194
15	Haecksel	60.924764
30	PL	60.482343
32	Ripplefluke	60
34	Shmuddel	59.83141
6	DN21	53.751742
27	Number1	53.503455
42	SN96	53.359052
41	SN90	42.550429
47	TR77	42.458701
8	Double	40.9293
9	Feather	38.236716
59	Zap	37.208978
52	TSN103	35.198851
0	Beak	34.921151

10	Fish	29.448398
18	Jonah	27.184466
61	Zipfel	25.976818
19	Knit	24.365341
21	MN105	23.242197
44	Thumper	22.029185
2	Bumper	16.603247
24	MN83	13.51097
5	DN16	8.015949
26	Notch	7.983333
16	Hook	6.047619
46	TR120	5.505495
3	CCL	4.344048
25	Mus	3.00873
53	TSN83	2.183333
49	TR88	1.7

55	Vau	1.605769
56	Wave	0.25
4	Cross	0
11	Five	0
12	Fork	0
22	MN23	0
31	Quasi	0
35	SMN5	0
48	TR82	0
58	Whitetip	0
60	Zig	0

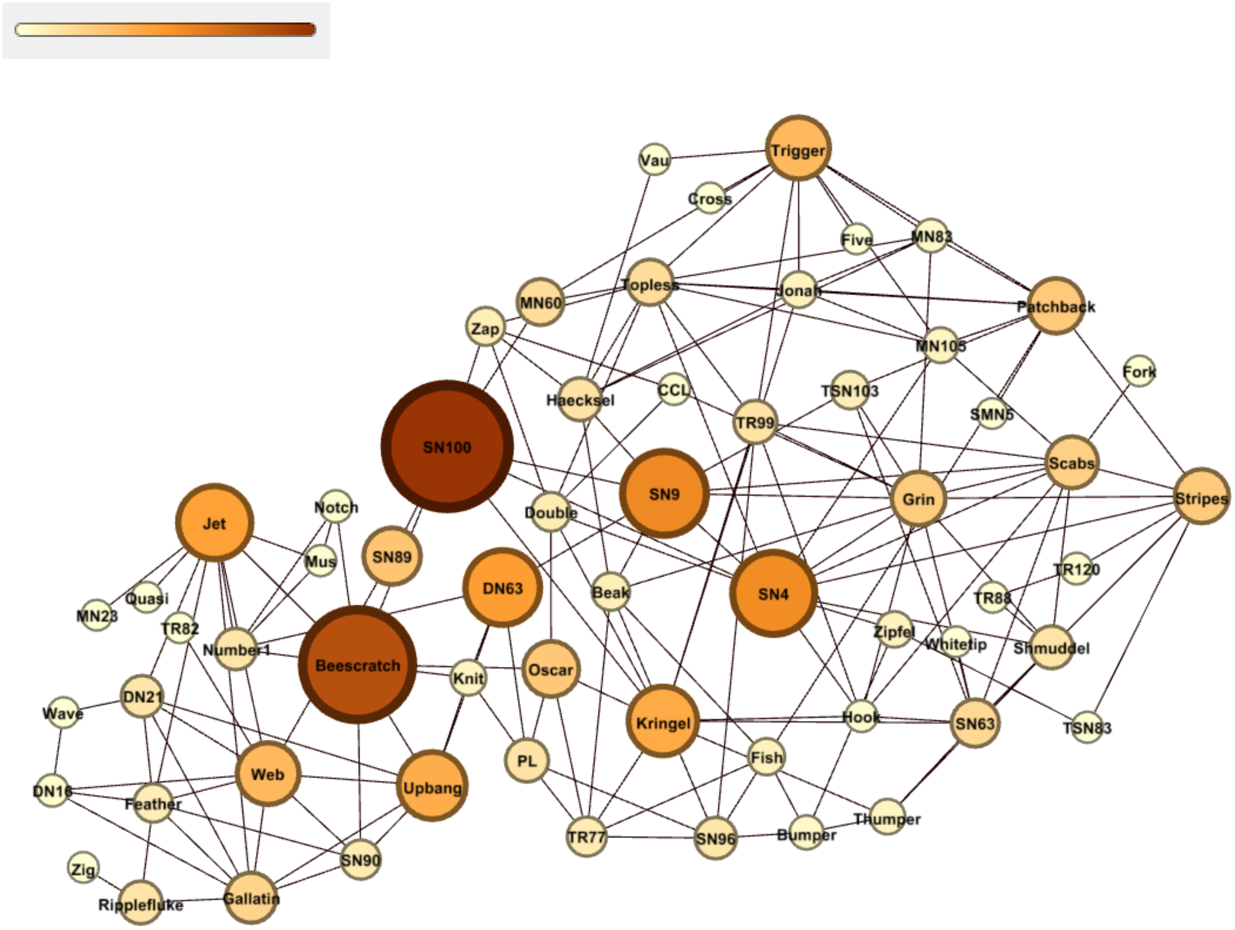

Figure 8: Betweenness centrality graph of dolphin social network - node color and size according to betweenness

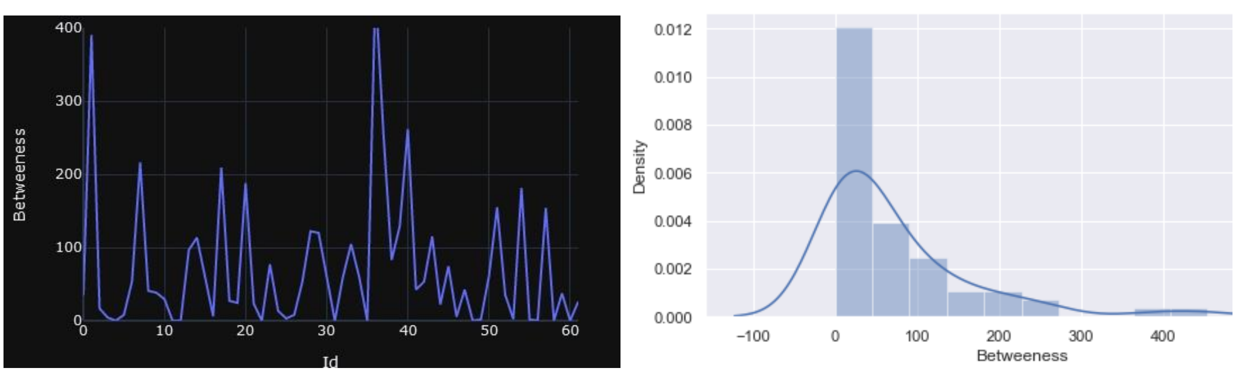

Figure 9: Line chart and histogram for betweenness centrality

Inference:

- The higher the betweenness, tend to be information brokers in human societies and potentially in bottlenose dolphin's societies as well. It was introduced as a measure for quantifying the control of a dolphin on the communication between other dolphins in this social network. It is defined using the shortest path length.
- By observing the size and color of the graph we can say that *SN100* has the highest betweenness, meaning that it acts as the shortest path for communication between many dolphins.
- Followed by *Beescratch, SN9, SN4, DN63* and *Jet* have highest betweenness centrality which means they have also been the shortest path for communication between many other dolphins.
- *Cross, Five, Fork, MN23, Quasi, SMN5, Whitetip, Zig and TR82* have very low betweenness as they are found in lighter shade in the graph, meaning that they are not often found in between any communication between other dolphins.
- Betweenness centrality has a right-skewed distribution. This is a unimodal data set, with the mode closer to the left of the graph and smaller than either the mean or the median. This shape indicates that there are a number of data points, perhaps outliers, that are greater than the mode.

3) Clustering Centrality

Clustering coefficient is a measure of the degree to which nodes in a graph tend to cluster together. The **local clustering coefficient** of a vertex (node) in a graph quantifies how close its neighbours are to being a clique (complete graph).

Id	Label	Clustering centrality
25	Mus	0.666667
26	Notch	0.666667

16	Hook	0.6
41	SN90	0.6
6	DN21	0.533333
21	MN105	0.533333
24	MN83	0.533333
9	Feather	0.52381
18	Jonah	0.52381
5	DN16	0.5
13	Gallatin	0.5
19	Knit	0.5
10	Fish	0.4
27	Number1	0.4
47	TR77	0.4
54	Upbang	0.380952
57	Web	0.361111
0	Beak	0.333333

3	CCL	0.333333
23	MN60	0.333333
32	Ripplefluke	0.333333
33	Scabs	0.333333
42	SN96	0.333333
52	TSN103	0.333333
45	Topless	0.309091
28	Oscar	0.3
30	PL	0.3
34	Shmuddel	0.3
59	Zap	0.3
38	SN63	0.285714
8	Double	0.266667
14	Grin	0.257576
29	Patchback	0.25
40	SN9	0.25

51	Trigger	0.244444
15	Haecksel	0.238095
43	Stripes	0.238095
50	TR99	0.238095
37	SN4	0.236364
17	Jet	0.222222
7	DN63	0.2
2	Bumper	0.166667
44	Thumper	0.166667
1	Beescratch	0.142857
20	Kringel	0.138889
36	SN100	0.047619
4	Cross	0
11	Five	0
12	Fork	0
22	MN23	0

31	Quasi	0
35	SMN5	0
39	SN89	0
46	TR120	0
48	TR82	0
49	TR88	0
53	TSN83	0
55	Vau	0
56	Wave	0
58	Whitetip	0
60	Zig	0
61	Zipfel	0

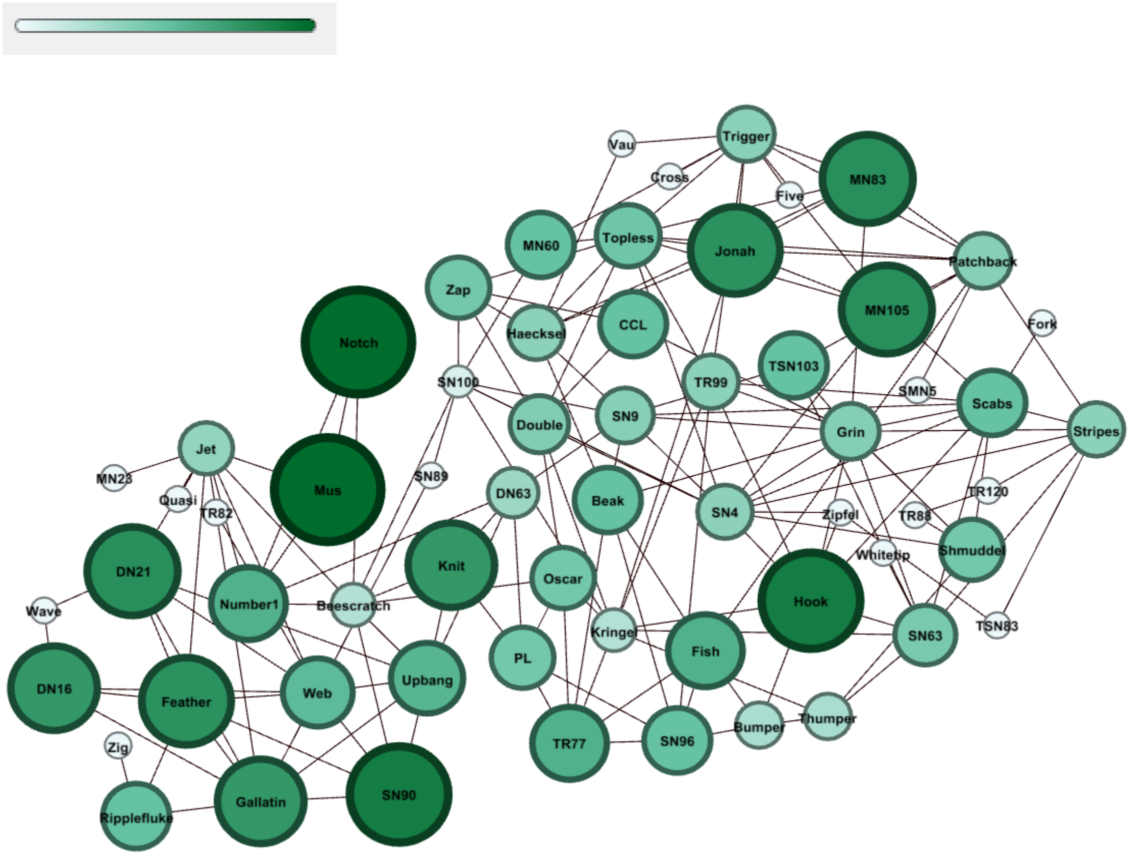

Figure 10: Clustering centrality graph of dolphin social network - node color and size according to the clustering centrality rate.

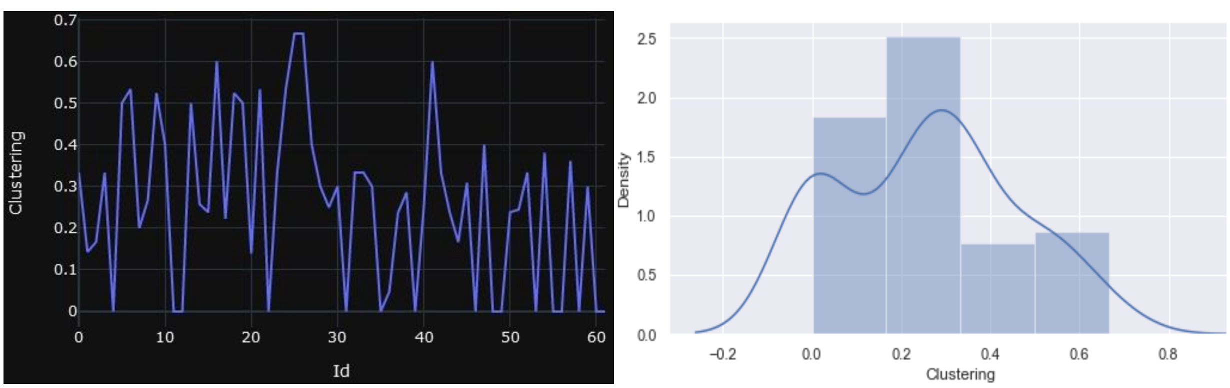

Figure 11: Line chart and histogram for clustering centrality

Inference:

- Dolphins who have higher clustering coefficient signifies how strong is their connection and also, they belong to a common community.
- By observing the size and color of the graph we can say that *Mus and Notch* have the highest clustering centrality coefficient followed by *Hook and SN90* which means their neighborhood is strongly connected. This also means these dolphins have more mutual friends than others.
- *Cross, Five, Fork, MN23, Quasi, SMN5, SN89, TR120, TR82, TR88, TSN83, Vau, Wave, Whitetip, Zig and Zipfel* have the lowest clustering centrality coefficient as they are found in lighter shade on the graph meaning, their friends are not connected which they have less mutual friends.
- Clustering centrality also has two peaks which means it is a bimodal function, where it has two separated intervals representing highest density.

4) Closeness Centrality

Closeness centrality measures how fast the flow of information would be through a given node to other nodes. It indicates how close a node is to all other nodes in the network.

Id	Label	Closeness centrality
36	SN100	0.417808
40	SN9	0.403974
37	SN4	0.398693
20	Kringel	0.391026
14	Grin	0.376543

1	Beescratch	0.371951
7	DN63	0.365269
28	Oscar	0.365269
33	Scabs	0.365269
8	Double	0.363095
50	TR99	0.350575
0	Beak	0.346591
45	Topless	0.346591
52	TSN103	0.342697
59	Zap	0.342697
15	Haecksel	0.338889
47	TR77	0.338889
18	Jonah	0.337017
43	Stripes	0.337017
39	SN89	0.335165
21	MN105	0.333333

23	MN60	0.333333
16	Hook	0.32973
38	SN63	0.32973
42	SN96	0.32973
51	Trigger	0.32973
54	Upbang	0.324468
29	Patchback	0.322751
30	PL	0.322751
19	Knit	0.316062
27	Number1	0.316062
34	Shmuddel	0.316062
10	Fish	0.312821
24	MN83	0.312821
44	Thumper	0.311224
17	Jet	0.309645
3	CCL	0.308081

57	Web	0.30198
61	Zipfel	0.30198
41	SN90	0.297561
2	Bumper	0.282407
26	Notch	0.277273
13	Gallatin	0.271111
55	Vau	0.271111
12	Fork	0.268722
6	DN21	0.267544
53	TSN83	0.259574
46	TR120	0.256303
25	Mus	0.25523
9	Feather	0.252066
4	Cross	0.24898
11	Five	0.24898
58	Whitetip	0.24898

49	TR88	0.246964
35	SMN5	0.24498
5	DN16	0.238281
22	MN23	0.237354
31	Quasi	0.237354
48	TR82	0.232824
32	Ripplefluke	0.216312
56	Wave	0.213287
60	Zig	0.178363

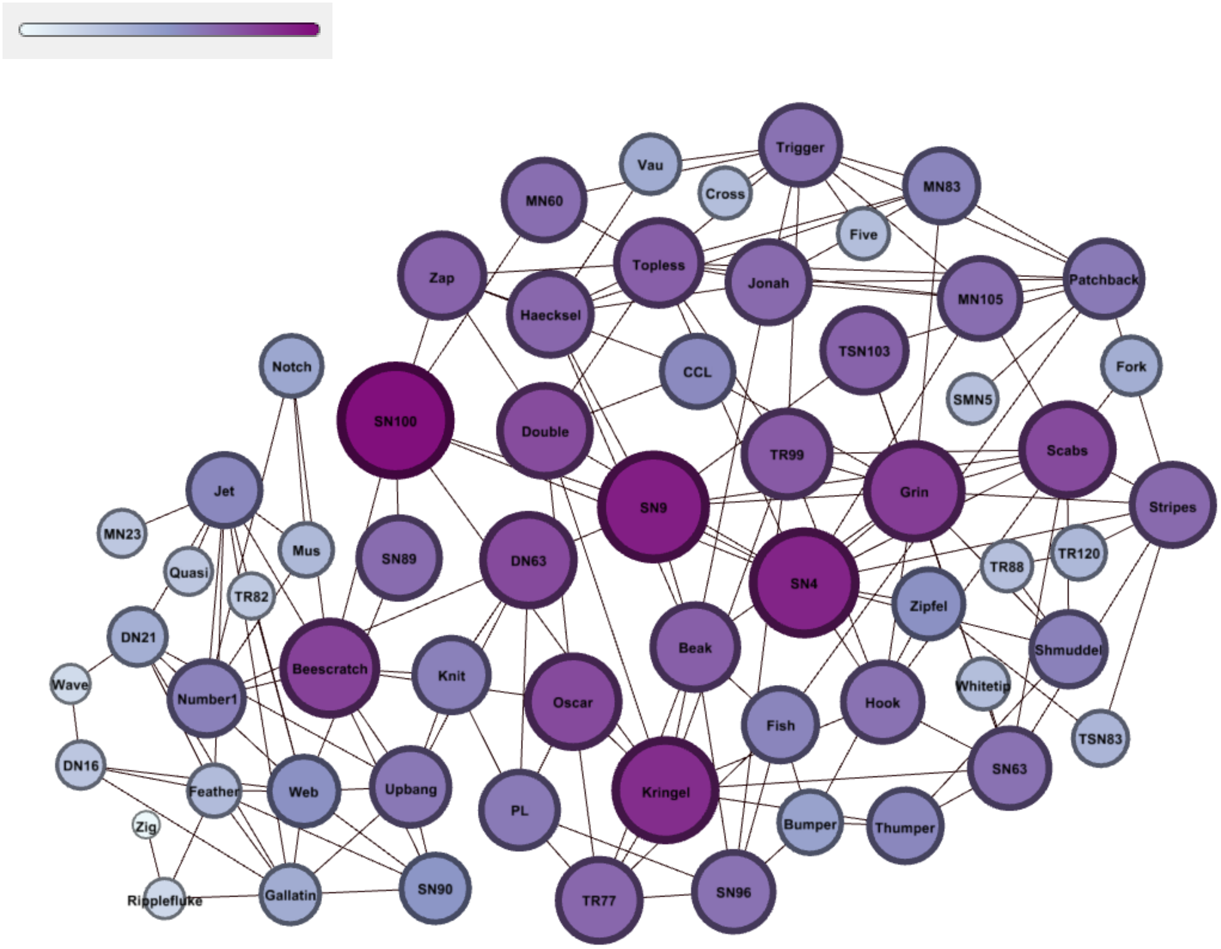

Figure 12: Closeness centrality graph of dolphin social network - node color and size according to the closeness of the node.

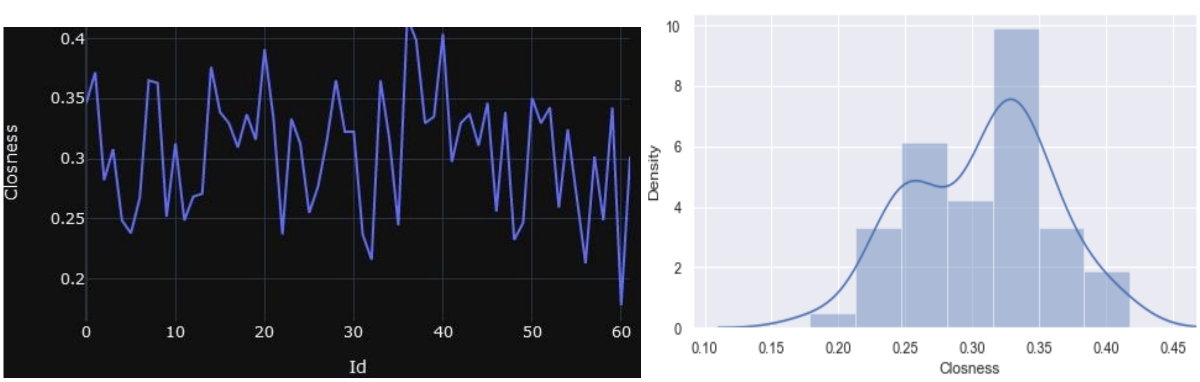

Figure 13: Line chart and histogram for Closeness centrality

Inference:

- Closeness centrality can tell us the individuals who are best placed to influence the entire network most quickly.
- By observing the size and color of the graph we can say that *SN100* and *Beescratch* have the highest clustering centrality coefficient followed by *SN9, SN4* and *DN63.*
- These dolphins are closer to many dolphins than others. Their shortest path between each and every dolphin is very less that they can easily be connected with every one of them.
- *Cross, Five, Fork, MN23, Quasi, SMN5, TR82, Whitetip* and *Zig* have the lowest clustering centrality coefficient, meaning they are not closer to any dolphins. It is difficult for them to connect with other dolphins
- Here the graph represents an undefined bimodal where the shape is not specifically defined. But we can see that it has two peaks which means it has two separated intervals that represent maximum density.

5) Eigenvector Centrality

Eigenvector centrality is a measure of the influence of a node in a network. It assigns relative scores to all nodes in the network based on the concept that connections to high-scoring nodes contribute more to the score of the node in question than equal connections to low-scoring nodes.

Id	Label	Eigenvector centrality
14	Grin	1
37	SN4	0.95912
45	Topless	0.907961

33	Scabs	0.882359
50	TR99	0.697945
40	SN9	0.689558
29	Patchback	0.673294
51	Trigger	0.667463
16	Hook	0.657517
21	MN105	0.650379
18	Jonah	0.645573
20	Kringel	0.627379
38	SN63	0.624166
24	MN83	0.610215
43	Stripes	0.598432
15	Haecksel	0.532595
36	SN100	0.485737
8	Double	0.477527
0	Beak	0.442751

34	Shmuddel	0.437557
52	TSN103	0.411339
57	Web	0.387156
13	Gallatin	0.377444
59	Zap	0.371485
17	Jet	0.35681
1	Beescratch	0.356611
54	Upbang	0.345005
9	Feather	0.329657
47	TR77	0.31867
42	SN96	0.307612
6	DN21	0.305039
28	Oscar	0.294695
41	SN90	0.289033
23	MN60	0.286857
10	Fish	0.276466

44	Thumper	0.260064
3	CCL	0.256029
7	DN63	0.247425
30	PL	0.208857
27	Number1	0.193442
5	DN16	0.191063
19	Knit	0.184724
61	Zipfel	0.172357
55	Vau	0.16706
2	Bumper	0.152172
39	SN89	0.132006
12	Fork	0.119824
32	Ripplefluke	0.119279
25	Mus	0.110748
26	Notch	0.10935
53	TSN83	0.108332

46	TR120	0.094981
35	SMN5	0.093387
4	Cross	0.092841
11	Five	0.092841
58	Whitetip	0.086229
56	Wave	0.080817
49	TR88	0.075563
48	TR82	0.062581
22	MN23	0.058727
31	Quasi	0.058727
60	Zig	0.020519

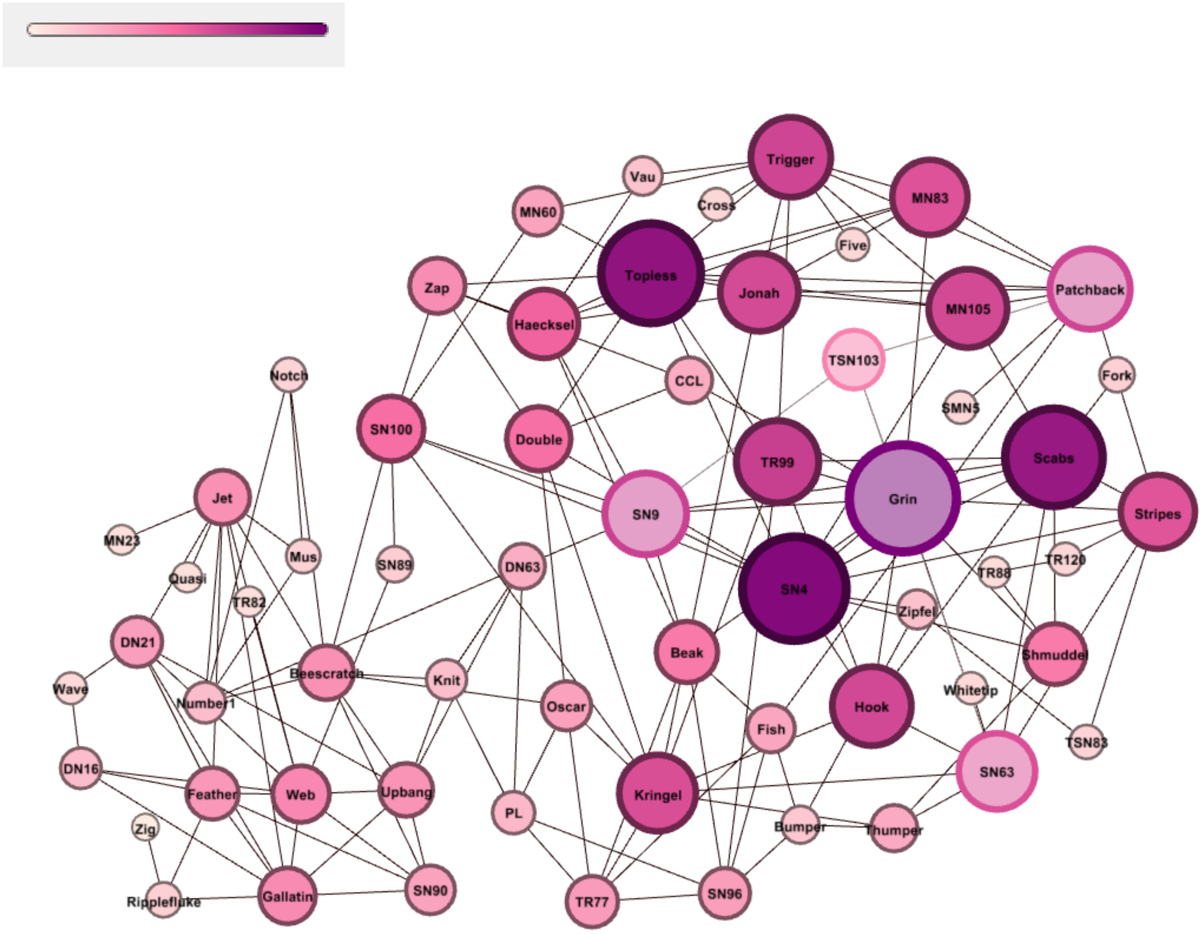

Figure 14: Eigenvector centrality graph of dolphin social network - node color and size according to the eigenvalues.

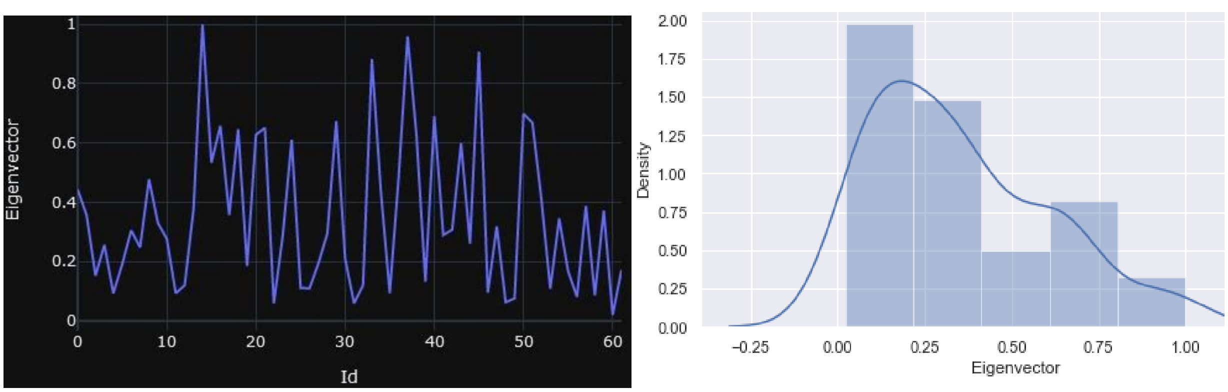

Figure 15: Line chart and histogram for betweenness centrality

Inference:

- Eigenvector centrality is an extension of degree centrality but it also considers the importance of the adjacent nodes.
- By observing the size and color of the graph we can say that *Grin* has the highest eigenvector centrality followed by *SN4, Topless, Scabs, TR99* and *SN9*.
- These dolphins are connected to dolphins which are highly connected. That is, their neighbors are also highly connected. They interact with dolphins who have more friends.
- *TR88, TR82, MN23* and *Quasi* have the lowest eigenvector centrality as they are found in lighter shade and less node size in the graph, meaning they are connected with dolphins who have less connections that is who have less friends.
- Eigenvector centrality represents right-skewed data. And the function is unimodal with mode closer to the left side. This represents there are many outliers that are greater than the mode.

Comparing closeness, eigenvector and clustering centralities:

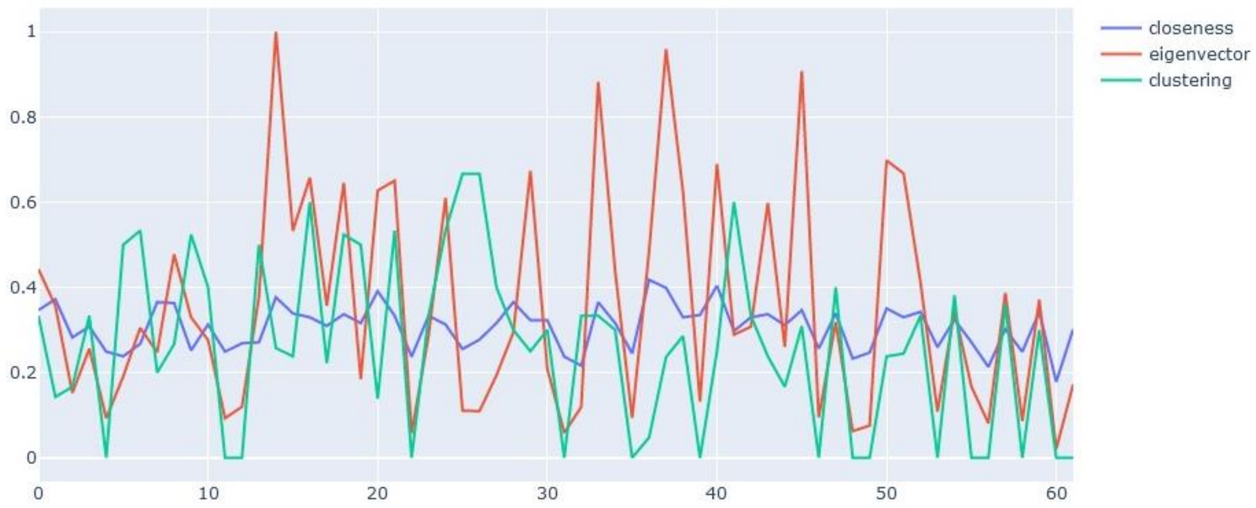

Figure 16: Line chart representing the variation among closeness, eigenvector and clustering centralities.

Inference:

- By looking at this graph we can easily compare three centralities for each node. Obviously, the red line is higher than the other two lines which represents that the eigenvector is more than the other two centralities.
- And also, closeness centrality does not vary much for each node, which means that we can reach every node in the shortest distance from any node. All nodes are almost close to each other.
- We didn't include betweenness and degree centrality as their values are comparatively higher than other centralities.
- We can also observe that there is a slight relationship between eigenvector and clustering centrality. They almost have similar values and have a linear relationship.

Conclusion

The purpose of this study was to apply Social Network Analysis in our own work on bottlenose dolphins. Centrality measures give nodes a scalar score that is easy to interpret and also helps you compare nodes in our networks in terms of importance. Graph centrality algorithms (i.e., degree, clustering, closeness, and eigenvector) were used to analyze the relationship between dolphins and their communication. We have also analyzed some important dolphins that influence most in their network. As a result, social network analysis was used to simplify the study of animal behaviour.

Tools and Software

Gephi - Network analysis software used for calculating analytical graph measures and for producing images of the graph.

Microsoft Excel - Used for creating tables.

Python programming - Used for data visualization.

Reference

- ➢ https://en.wikipedia.org/wiki/Centrality
- ➢ http://www.monkeymiadolphins.org/wp-content/uploads/2019/10/Stanton_Mann_Ch18.pdf
- ➢ http://networkrepository.com/
- ➢ https://www.youtube.com/watch?v=I0S_V9flX4U&t=28s
- ➢ https://www.youtube.com/watch?v=371n3Ye9vVo
- ➢ https://www.sciencedirect.com/topics/social-sciences/social-network-analysis