Link Analysis of Dolphin Communities

I. Introduction

This report will perform link analysis operations using Gephi on a data set representing a graph of frequent associations between dolphins in Florida, provided by networkrepository.com. Link analysis methods will be used to predict the possible communities that individual dolphins may belong to. Several graph metrics and network overviews will be employed in this effort to predict community membership such as modularity, betweenness, and network diameter. The goal is to understand the social networks of dolphin communities and how isolated communities tend to behave.

Future sections of this report describe the dataset, the methodology, results accompanied by a discussion, and concluding thoughts. Section II contains a description of the dataset used for this analysis which will be in the form of an adjacency list. The methodology and model choices for link analysis is presented in section III which will reply heavily on Gephi. Section IV contains a report and discuss of the results based on metrics such as modularity and graph betweenness. Finally, section V will provide the conclusion upon these findings and determine and the overall connectivity of dolphin communities.

II. Data Description

There are two columns in total contained in the data set as depicted in Table I. The data set represents an adjacency list which is a list where entries in the same row share a frequent association with each other. An adjacency list of this kind is then used to create an undirected graph of all associations shared amongst the individual dolphins. Both columns in the adjacency list refer to the IDs of the individual dolphins. Each unique ID represents a vertex in the graph and each row connecting one dolphin to another represents an edge. The data set therefore represents an undirected graph with sixty-two vertices, individual dolphins, and one hundred and fifty-nine edges, associations.

TABLE I.

Attribute	Type	Example Value	Description
Source	Integer (natural)	57	Dolphin Id
Target	Integer (natural)	7	Dolphin Id

Figure I shows this undirected graph of individual dolphins with all sixty-two vertices and one hundred and fifty-nine edges. In Section V, the table in Table I will have addition rows which will be appended linked analysis of this undirected graph is performed. A Fruchterman Reingold layout algorithm was performed to produce Figure I.

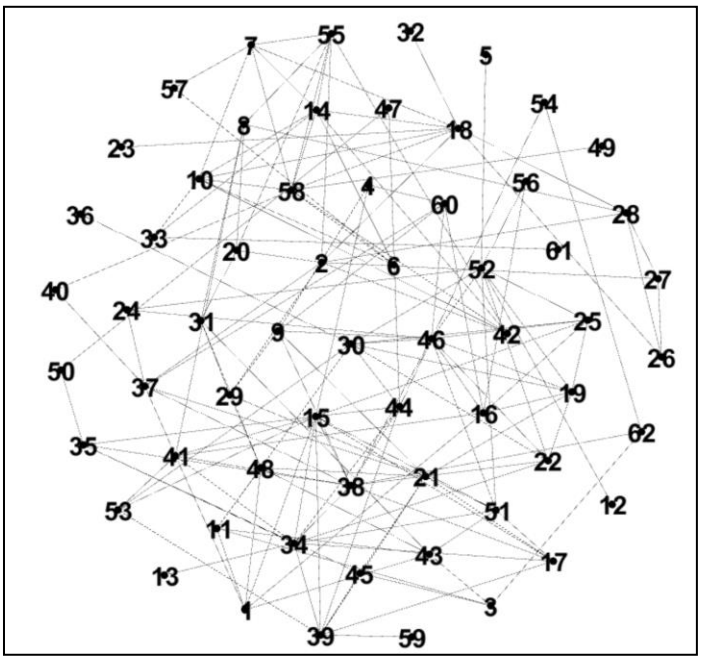

Fig. 1. Undirected Graph constructed
from adjacency list

III. METHODOLOGY

The adjacency list was graphed and analyzed using the Gephi tool. The first analyzes that was conducting using the Gephi tool was creating a modularity report. The modularity report, as shown in Figure II, displays the total number of modularity classes as a 2-D graph where the Modularity Class number is the horizontal axis and the number of vertices that belong to each class is the vertical axis.

Fig. 2. Modularity Classes of
adjacency list

Each of the Modularity Classes were then assigned a color and used to color the graph from Figure I. The new colored graph is represented in Figure III and helpfully visualizes the connections of each node within their respective communities.

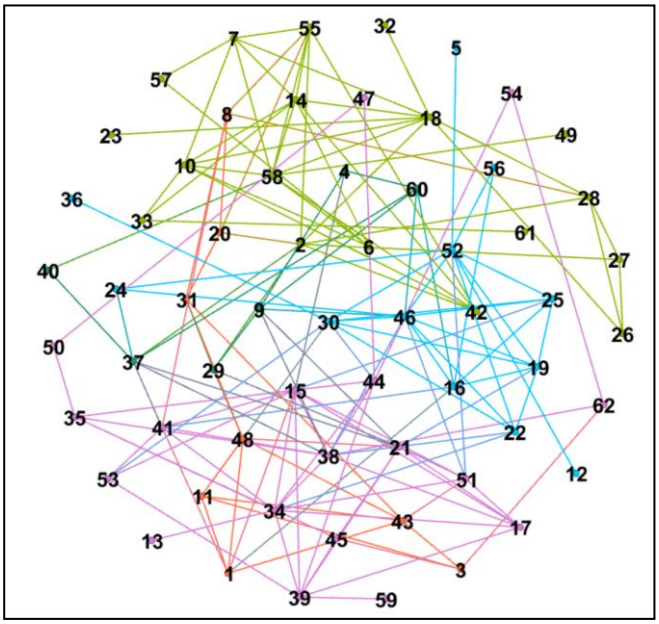

Fig. 3. Undirected Graph colored
using Modularity Classes

The Gephi tool was also used to measure the betweenness of the nodes in the graph. Betweenness centrality was chosen over closeness centrality because betweenness was a better measure of determining which dolphins socialize the most and as a result which dolphins are more likely to create connections between other Modularity Classes. The betweenness centrality of each individual dolphin was then applied to Figure III to create Figure IV where the mass of each node represents the betweenness centrality score of that dolphin.

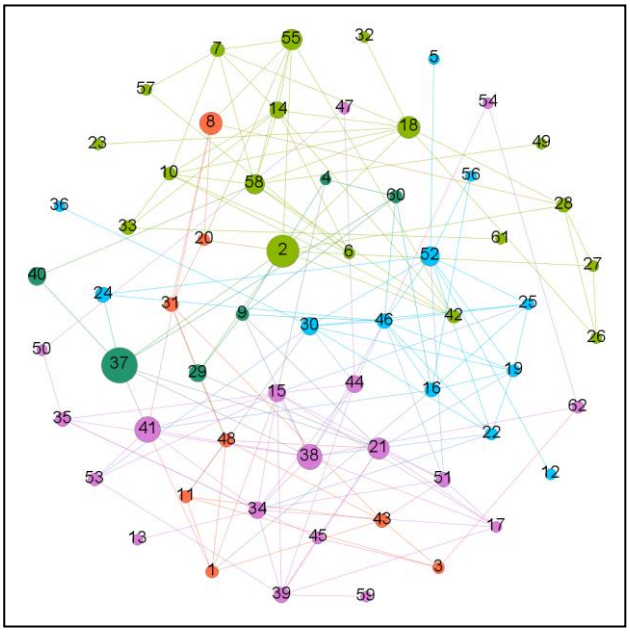

Fig. 4. Undirected Graph sized using
Betweenness centrality

IV. Results and Discussion

The Modularity Classes from Section III created five distinct communities within the graph. These communities, as shown in Table II, ranged in size from roughly ten percent to thirty percent of the entire membership of the graph. Each community of the graph was of a nontrivial size. Furthermore, the graph as a whole has an average path length of 3.35 nodes and a diameter of eight. This implies that the graph contains individuals who are not well associated with each other.

TABLE II.

Modularity Class	**Size (%)**
Class 0	29.03
Class 1	29.03
Class 2	12.9
Class 3	9.68
Class 4	19.35

V. Conclusions

The final results of this linked analysis are recorded in Table III, which is the same as Table II with the columns for modularity and betweenness centrality appended. The modularity column contains values between zero and four and betweenness centrality column contains values between zero and one. From these score it can be determined that the five communities of dolphins in Florida do not strongly associate themselves with other communities of Dolphins forming weak social networks outside of their own communities.

TABLE III.

Table Evaluation Results			
Attribute	**Type**	**Example Value**	**Description**
Source	Integer (natural)	57	Dolphin Id
Target	Integer (natural)	7	Dolphin Id
Modularity Class	Integer (natural)	3	Class (0-4)
Betweenness Centrailty	Interval (0-1)	0.248	Betweenness value of node