Table of Contents

List of Figures

1. Introduction

We have decided to do a network analysis on the dataset on wild birds interaction which is collected over 6 days. The dataset is an edge list which gives us the edges in the birds graph that we are going to analyze. The first two columns represent the birds' names or numbers and their interaction with each other. We also have a weightage on our edges with a third column as half weighted index. The weight basically shows how strong the interaction/edge strength is between the birds. Higher the half weight index, higher is the time of interaction between birds. The fourth column represents the day of the week and it is a numerical number ranges between 1 and 6. Later in the analysis, we will be splitting the data by making subsets of the whole graph to analyze the interactions over each day separately.

The reason for choosing this dataset is because I have always been fascinated by the patterns the birds fly in over the sky as a child. The perfection in those patterns always made me wonder if they all communicate with each other. By doing some research about the patterns of communication of birds, we could bring some insights to how different species of birds behave and communicate with each other on a daily basis just like humans.

After importing our dataset to R environment in jupyter notebook, we converted it into a matrix and plotted and igraph of the whole network over six days as shown in fig. 1. As seen from the graph, we can not gain any useful insights other than some of the vertices/birds on the outside of the graph have not had that significant interaction as others.

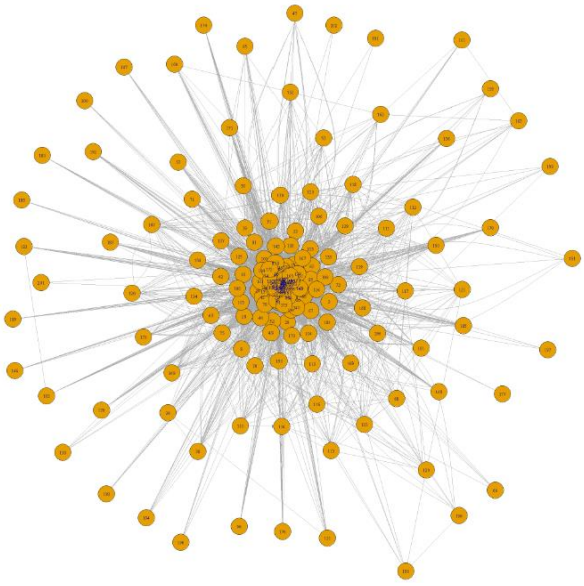

Figure 1: wild birds' network over 6 days

1.1. Questions to answer

To gain some insights from the graph, we will be splitting and making subsets of the graph according to interactions on each day. We will also be calculating some important graph metrics along the way. The questions that we are hoping to answer after our network analysis are:

1. The overall pattern of change in interactions between the birds over 6 days.
2. The highest and lowest degree of the overall graph and the birds' with highest and lowest degree in a span of 6 days.

3. The bird with the most interactions each different day.
4. Day with the most and lowest interactions between the birds.

2. Methods

What we are trying to achieve in this network analysis is to gain some insights on the frequency of communication between birds each day. The edges in our graph are undirected, which means that edges do not have a direction and birds can have a two-way interaction with each other. We will see in subsets of the graph for each day that there are certain birds which communicate within their own clusters.

2.1. Summary of graph statistics

```
IGRAPH 6090840 U--- 202 11900 --
+ attr: Day (e/n), Halfweight_Index (e/n)
```

- The graph is undirected
- There are 202 vertices or nodes. Each distinct vertex represents a bird.
- There are 11900 edges. The edges represent the relationship or interaction between the birds with neighboring birds.
- We assign two attributes to our graph. One being the half weight index (column 3) which is the strength of interaction between the birds. Second being the day (column 4) of interaction.

2.2. Graph Metrics

According to the questions we are trying to answer, we will be dividing our bird network in 6 different subsets. Each subset of day from 1 to 6 represents the interaction between birds on a specific day. After our analysis, we will find valuable insights on which day the birds interacted the most and how did the interaction changed over days.

To find the insights stated above, we will need to calculate following graph metrics for the network and day-wise subsets:

1. The length of the overall graph and day-wise subsets of graph
    - Number of *vertices*
    - Number of *edges*
    - *Average path length* which is the mean shortest path averaged over all pair of vertices [1]. It basically represents the average number of steps it takes to get from one node to another node. For an undirected graph, the average path length is calculated by equation 1 shown below:

$$\ell = \frac{1}{N(N-1)} \sum_{i \neq j} d_{ij}$$

    - The *diameter* which is the length of longest shortest path in the graph.
2. The *clustering coefficient* and the average coefficient of each day. It is a measure of degree in which the vertices cluster together in a network [2].
3. The *degree* of the node. For an undirected graph, the degree shows the count of edges that are incident on it. This includes the in degree and the out degree [3]. In our calculations, we will calculate:

- The maximum and average degree of the whole graph and the subsets of graph for each day.
- The bird with the highest degree for each day. (The lowest degree of the graph was only calculated for day 1 graph, the reason is because only the graph for day 1 was connected however for the rest of the days, each graph was disconnected so there were multiple vertices with 0 degree)

4. The *closeness centrality* of the most interactive bird on each day. It is a measure of inverse distance to all other vertices. If the closeness centrality of a vertex is high, it means that it has a shortest distance to other vertices [4]. From this metric, we can gather information about how birds spread information to other birds throughout the graph.
5. The *betweenness centrality* of most interactive bird on each day. This metric calculates the shortest paths in a graph.
6. The *neighboring vertices* of the most interactive bird for each day. This is to discover which birds interact with the bird with the highest degree.

3. Results/Analysis

3.1. Results for overall network

To visualize our overall 6 day network better, we color coded our edge attributes according to the day of the interaction. So, for each day we assigned a vibrant color to the edges and while plotting we made the edge width equal to the strength of the half weight index. The colors assigned are black, red, green, blue, cyan and yellow for days 1, 2, 3, 4, 5, and 6 respectively. The reason for doing so is we would be able to see by the thickness of the edge if the interaction between two birds is strong or weak. As shown in fig. 2 below, we can see that we have three main clusters of birds in our overall network. Each cluster of birds has significant interactions between each other and other clusters.

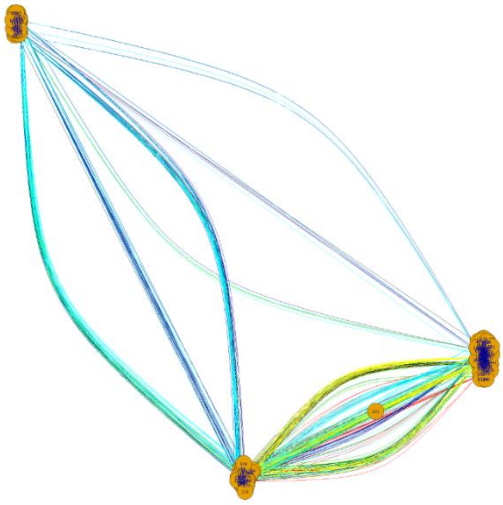

Figure 2: Birds' interaction network

The metrics found for the whole network are:

- Nodes: 202

- Edges: 11900
- Average path length: 2.11
- Diameter: 4
- Transitivity: 0.703
- Highest degree of the vertex: 397
- Lowest degree of the vertex: 2
- Average degree of the vertex: 117.8
- Bird with the highest degree: Bird # 6
- Birds with the lowest degree: Birds # 177, 191, 201, 202

3.2. Results for day 1

To plot the subset graphs for each day, we first calculated the mean half weight index for each day. We then add a new edge attribute 'color' to our graphs and color coded the edges according to the values of half weight index in comparison to the average half weight. If the half weight of a particular vertex is greater than the mean weight index for that day, we color the edge as dark blue. Similarly, dark green if its is equal to half weight index and red if it is less that the index for that day. The following fig. 3 shows the graph of day 1 with layout_with_fr. The graph for day 1 is connected and each bird has communicated with someone else at some point during the day.

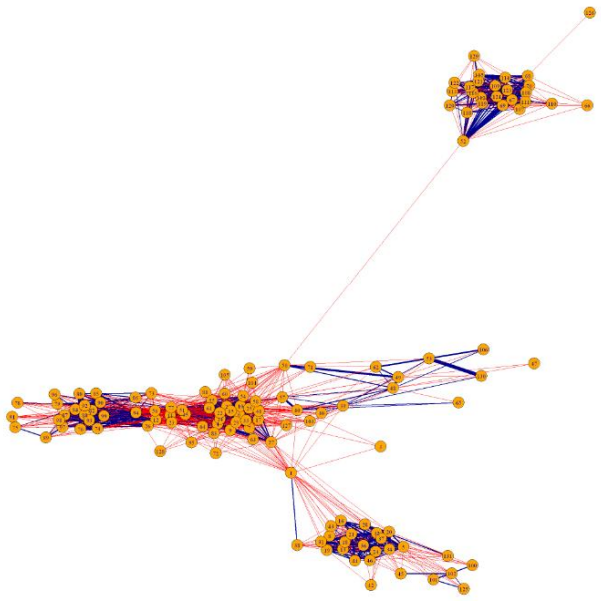

Figure 3: Day 1 Network of bird interaction

The metrics found for day 1 are:

- Vertices: 131
- Nodes: 1444
- Average path length: 2.73
- Vertex with maximum degree: 57
- Average vertex degree: 22.04
- Bird with the highest degree: Bird # 56

- Closeness centrality of most interactive bird # 56: 0.00393
- Betweenness centrality of most interactive bird: 496.49
- Clustering coefficient: 0.68
- Diameter: 6

3.3. Results for day 2

Fig.4 shows the network of day 2. The connected components in the graph are 20, which includes three clusters shown in the figure, so we can say that the graph is disconnected and there are 17 isolated vertices/birds which did not interact with anyone on day 2. For a disconnected graph, the closeness centrality is not well defined.

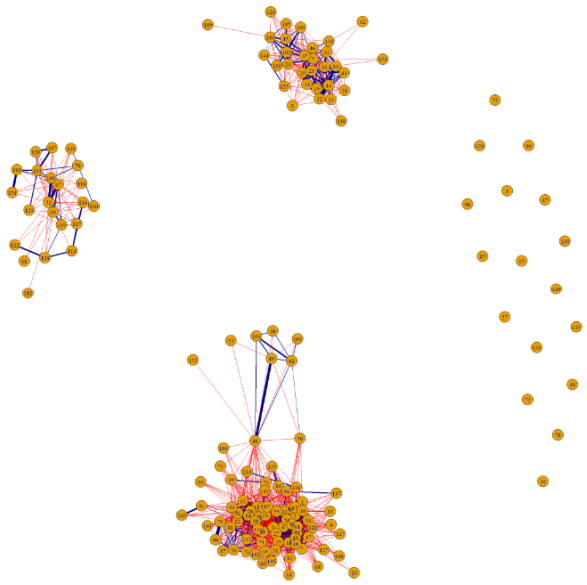

Figure 4: Day 2 network of bird interaction

The metrics found for day 2 are:

- Vertices: 152
- Nodes: 1483
- Average path length: 1.74
- Vertex with maximum degree: 62
- Average vertex degree: 19.51
- Bird with the highest degree: Bird # 6
- Closeness centrality of most interactive bird # 6: 9.31 e^{-05}
- Betweenness centrality of most interactive bird: 169.4
- Clustering coefficient: 0.64
- Diameter: 3

3.4. Results for day 3

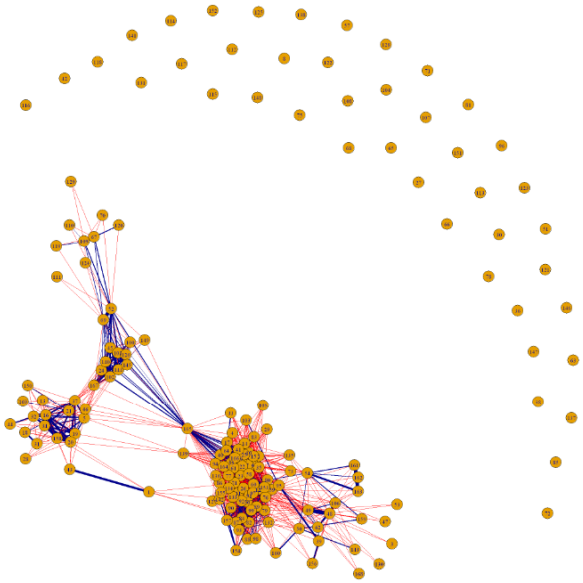

Figure 5: Day 4 network of bird interaction

The metrics found for day 3 are:

- Connected components: 44 (43 isolated vertices/birds)
- Vertices: 169
- Nodes: 1615
- Average path length: 2.38
- Vertex with maximum degree: 58
- Average vertex degree: 19.1
- Bird with the highest degree: Bird # 58
- Closeness centrality of most interactive bird # 58: 0.00013
- Betweenness centrality of most interactive bird: 139.17
- Clustering coefficient: 0.68
- Diameter: 6

3.5. Results for day 4

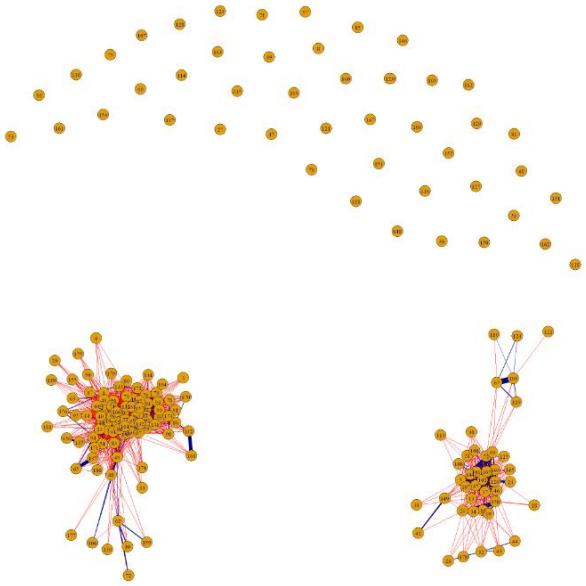

Figure 6: Day 4 network of bird interaction

The metrics found for day 4 are:

- Connected components: 48 (46 isolated birds)
- Vertices: 181
- Nodes: 2009
- Average path length: 1.771
- Vertex with maximum degree: 71
- Average vertex degree: 22.19
- Bird with the highest degree: Bird # 6
- Closeness centrality of most interactive bird # 6: 6.23 e^{-05}
- Betweenness centrality of most interactive bird: 185.84
- Clustering coefficient: 0.71
- Diameter: 4

3.6. Results for day 5

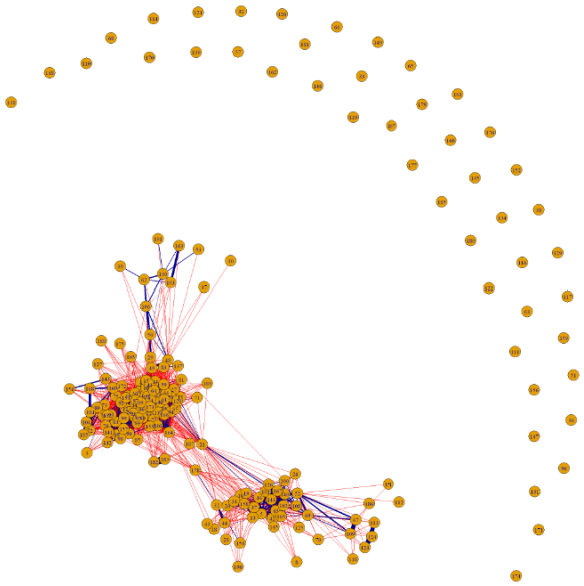

Figure 7: Day 5 network of bird interaction

The metrics found for day 5 are:

- Connected components: 47 (46 isolated birds)
- Vertices: 191
- Nodes: 2512
- Average path length: 2.09
- Vertex with maximum degree: 78
- Average vertex degree: 26.3
- Bird with the highest degree: Bird # 6
- Closeness centrality of most interactive bird # 6: 0.00011
- Betweenness centrality of most interactive bird: 198.7
- Clustering coefficient: 0.71
- Diameter: 4

3.7. Results for day 6

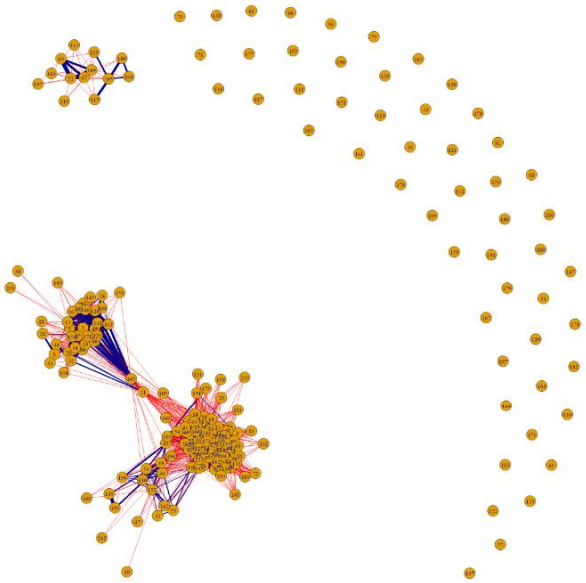

Figure 8: Day 6 network of bird interaction

The metrics found for day 6 are:

- Connected components: 55 (53 isolated birds)
- Vertices: 202
- Nodes: 2837
- Average path length: 2.037
- Vertex with maximum degree: 81
- Average vertex degree: 28.1
- Bird with the highest degree: Bird # 6
- Closeness centrality of most interactive bird # 6: 7.39 e^{-05}
- Betweenness centrality of most interactive bird: 237.18
- Clustering coefficient: 0.76
- Diameter: 4

3.8. Histogram

Finally, we will show the distribution of interactions over the 6 days in a histogram shown below in fig. 9.

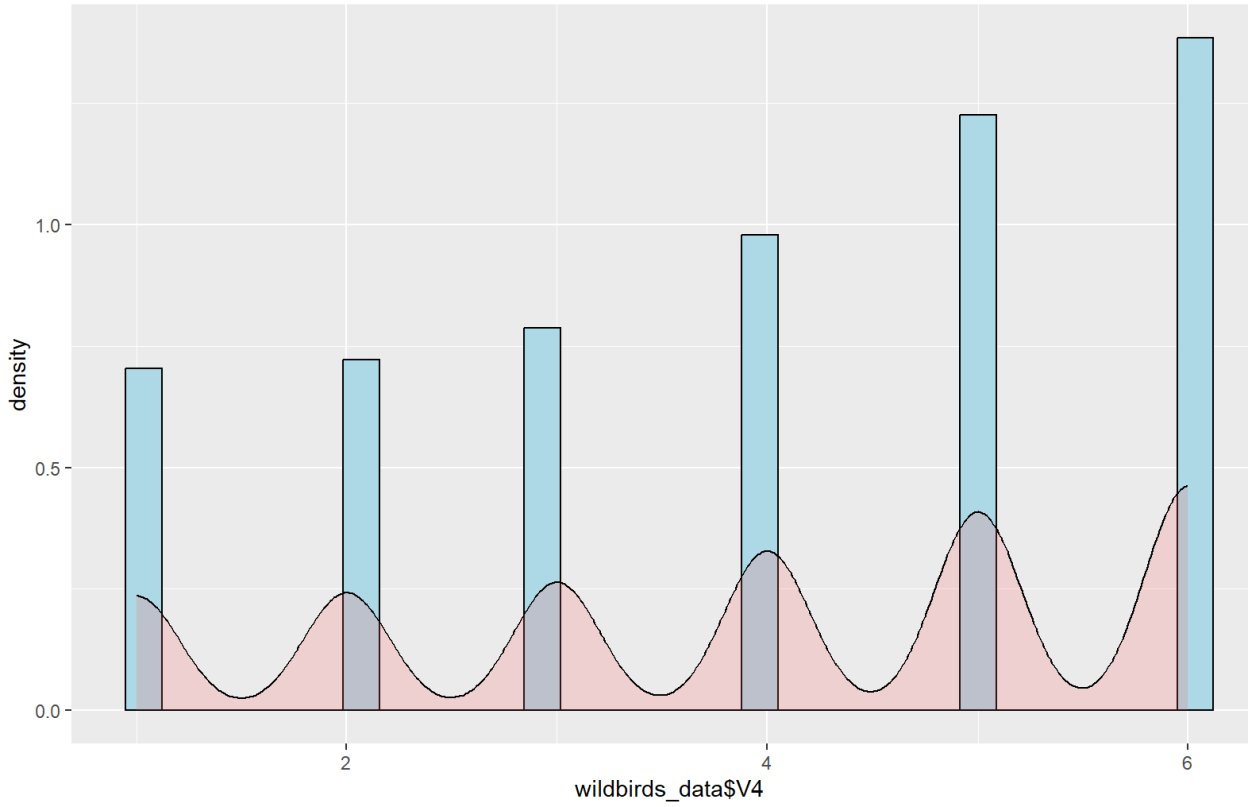

Figure 9: Bird interactions over 6 days

4. Conclusion

After gathering all the insights from the graph metrics and the plots shown above, we are able to answer the questions asked in the beginning.

1. The pattern of change in interactions over 6 days?
    - As we were given the half weight index as an attribute of strength of interactions, the half weight of day 4 and 5 was significantly more than other days.
2. The highest and lowest degree?
    - The highest degree of the overall graph was 397 whereas the lowest was 2.
    - The bird with the highest degree over the span of 6 days was bird # 6 whereas there were multiple birds with the lowest degree (birds # 177, 191, 201, 202).
3. Bird with most interactions on each day?
    - Day 1: 56
    - Day 2: 6
    - Day 3: 58
    - Day 4: 6

- Day 5: 6
- Day 6: 6

Overall we can come to the conclusion that bird # 6 was the most interactive bird throughout the week.

4. Day with most and lowest interaction?
    - As seen from the histogram, day 6 is the most interactive day whereas day 1 had the least interactions.

Conclusively, we did not have enough information to go with at the start. However, once we split the data into subsets of days, we could find different patterns on how the interactions happened each day.

5. References

[1] "Mean path length definition - Math Insight," Mathinsight.org, 2021. https://mathinsight.org/definition/network_mean_path_length

[2] "Clustering Coefficient in Graph Theory - GeeksforGeeks," GeeksforGeeks, Feb. 08, 2018. https://www.geeksforgeeks.org/clustering-coefficient-graph-theory/

[3] "Section 6.1," web.mit.edu. https://web.mit.edu/urban_or_book/www/book/chapter6/6.1.html#:~:text=The%20degree%20of%20a%20node (accessed Dec. 16, 2022).

[4] "Closeness Centrality - Neo4j Graph Data Science," Neo4j Graph Data Platform. https://neo4j.com/docs/graph-data-science/current/algorithms/closeness-centrality/#:~:text=Closeness%20centrality%20is%20a%20way (accessed Dec. 16, 2022).