COMPLEX SYSTEM PROJECT

SONGBIRD-SOCIAL DATA ANALYSIS

Submited by

HANIE HATAMI(99100614)

ALIREZA ASTANE(99100574)

MOHAMMAD GHORAISHI(99100788)

Supervisor

PROF. ROUHANI

DEPT. OF PHYSICS

SHARIF UNIVERSITY OF TECHNOLOGY

1402/4

CONTENTS

Chapter 1

INTRODUCTION

Graph theory is a branch of mathematics that studies graphs, which are mathematical structures used to model pairwise relations between objects. A graph in this context is made up of vertices (also called nodes or points) which are connected by edges (also called arcs or lines).

There are many types of graphs, including undirected graphs, directed graphs (also known as digraphs), weighted graphs, bipartite graphs, and others. Each type of graph has its own unique properties and uses.

Network analysis is a method of analyzing complex systems by representing them as graphs. In this context, a "network" is simply another term for a graph. The nodes in the network represent entities (which could be anything from people in a social network, to computers in a computer network, to neurons in a neural network), and the edges represent relationships or interactions between these entities.

By representing a complex system as a network, we can use the tools and techniques of graph theory to analyze the system. For example, we can use graph theory to identify the most important nodes in the network (a process known as centrality analysis), to detect communities of nodes that are more densely connected with each other than with the rest of the network (community detection), to measure the network's robustness to failures or attacks, and so on.

In summary, graph theory provides the mathematical foundation for network analysis, allowing us to analyze complex systems in a rigorous and systematic way.

In this article, we are looking to investigate one of the most interesting networks, which is bird song, and draw its graph, explain and calculate some of its parameters such as density, transitivity, degree assortativity coefficient, and ext.

Chapter 2

EXPLANATION OF SONGBIRD-SOCIAL DATA AND WHY IT IS IMPORTANT

Songbirds are known for their complex vocal behaviors and their ability to communicate through elaborate songs. The study of songbirds-social behavior often involves collecting and analyzing data on their vocalizations, interactions, and social structures.

Researchers interested in songbird social data may use various techniques to gather information. These can include audio recordings of songbird vocalizations, video recordings to observe social interactions, and even tracking devices to monitor movement and spatial relationships within a group or population.

Once the data is collected, researchers can analyze it to gain insights into various aspects of songbird social behavior. They may study the structure and dynamics of vocalizations, examine patterns of interactions within groups, investigate mating behaviors, or explore the effects of environmental factors on social behavior.

By studying songbird social data, researchers can deepen our understanding of how these birds communicate, form social bonds, establish territories, and navigate their environment. This knowledge can contribute to broader studies of animal behavior, communication, and evolution.

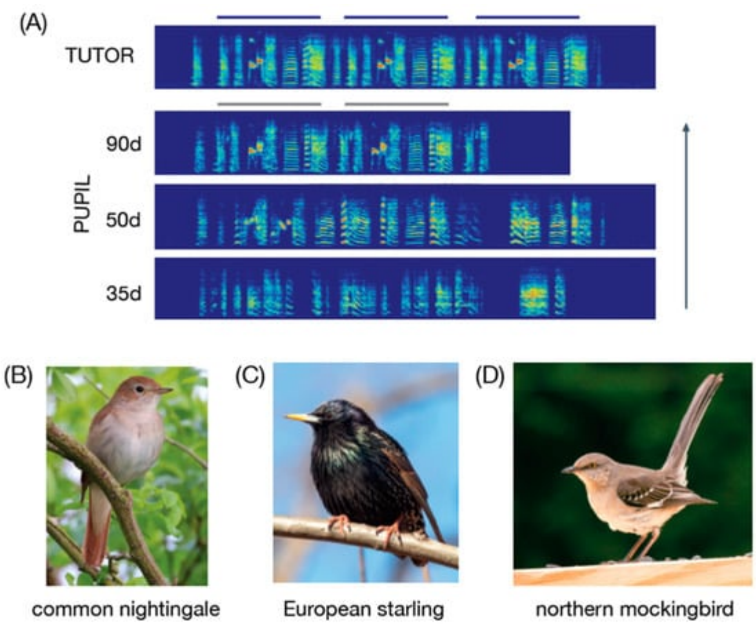

Figure 2.1: Song development and images of songbirds commonly studied with regard to late song (S2) learning.

Chapter 3

DATA PREPARATION

Data preparation and visualization are crucial steps in analyzing and understanding data. In this particular case, data was initially stored in EDGES format, which includes information about the, second node, and the weight of each edge in separate rows. To facilitate further analysis, the data was converted into CSV (comma-separated values) format using the pandas library.

Once the data was in CSV format, a networkx object was created from the dataset. Networkx is a powerful Python library for studying the structure and dynamics of complex networks. It provides a wide range of tools and algorithms for network analysis.

With the networkx object in hand, a pyvis network was obtained. Pyvis is a Python library that allows for the creation of interactive network visualizations. It utilizes the visJS library, which is a JavaScript library for visualizing network data. The result of this process is a visually appealing and interactive network visualization.

The output of the pyvis network visualization is typically in HTML format, which can be easily shared and viewed in web browsers. This format allows for easy exploration and interaction with the network, enabling users to gain insights and make discoveries from the data.

By leveraging the power of data preparation and visualization techniques, analysts and researchers can effectively explore and communicate complex network structures and relationships.

Chapter 4

GRAPH

In this chapter, we will draw different graphs of this data.

4.1 Network Graph

Let's see what out network looks like!

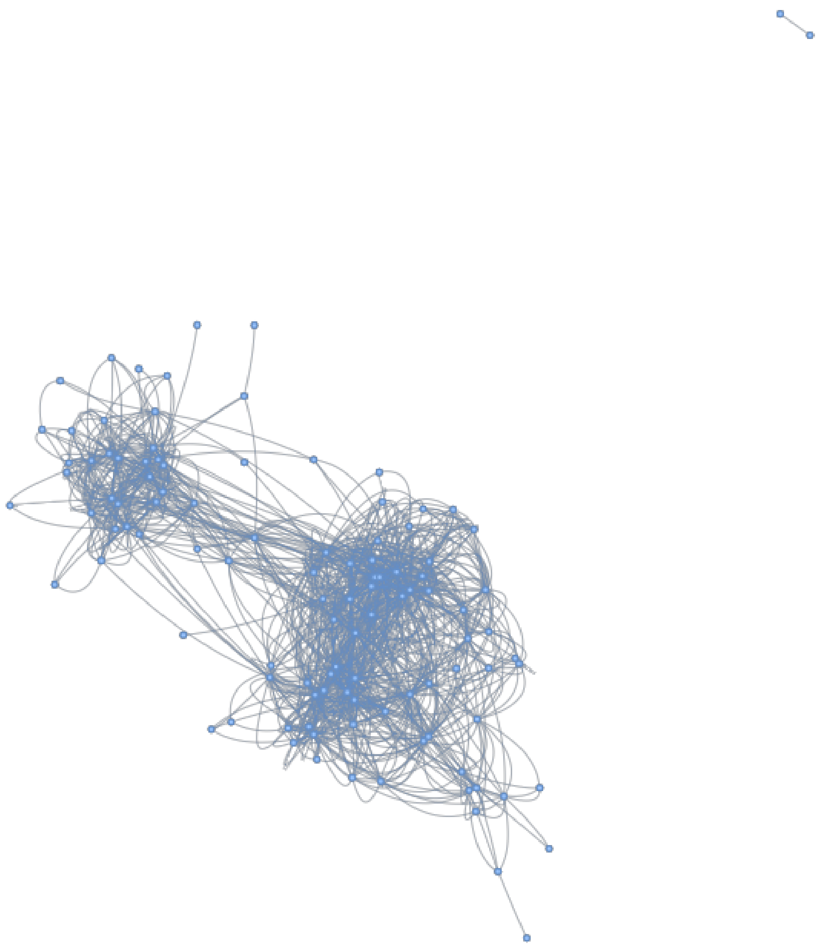

Figure 4.1: Songbird-social network graph.

4.2 Degree Graph

Degree counts how many edges are connected to a node.

Let's scale the node sizes of our toy network based on their total degree numbers.

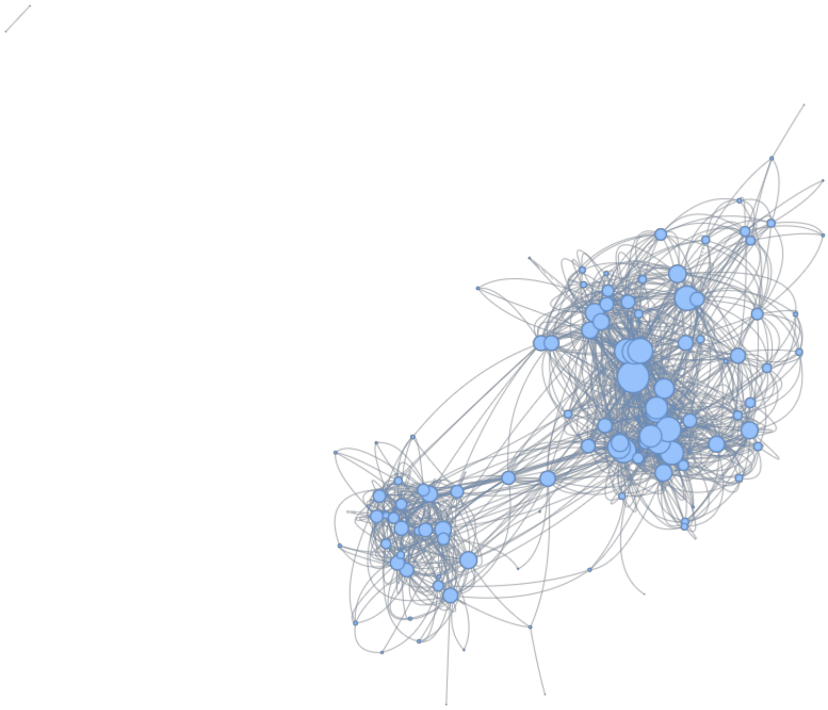

Figure 4.2: degree Graph.

4.3 Betweenness centrality Graph

Betweenness centrality is a measure used in graph centrality of node within a network. It measures the extent to which a node lies on the shortest paths between other nodes in the network. In other words, it calculates how often a node acts as a bridge or intermediary between other nodes.It's important to note that betweenness centrality is a relative measure, meaning it is calculated in relation to the other nodes in the network.

Nodes with high betweenness centrality are considered to have significant influence or control over the flow of information or resources in a network.

The figures 4.3 shows the size of nodes as their degree.

4.4 Eigenvector Centrality

Eigenvector centrality is another measure theory to determine the importance or centrality of a node within a network. It assigns a score to each node based on the concept of eigenvectors. In eigenvector centrality, the importance of a node is not only determined by the number of connections it has but also by the importance of its neighboring nodes. A node is considered important if it is connected to other important nodes.
Eigenvector centrality is useful for identifying nodes that have influence or control over the entire network, rather than just being intermedi like in betweenness centrality. Nodes with eigenvector centrality are considered to be influential or powerful within the network.

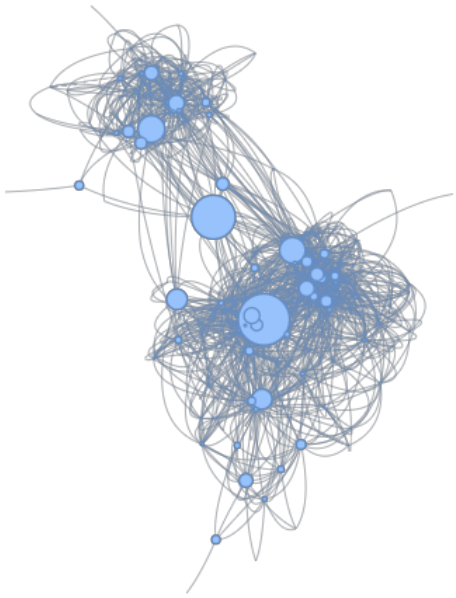

Figure 4.3: betweenness Centerality Graph

The figure 4.4 shows the size of nodes as their eigenvectors.

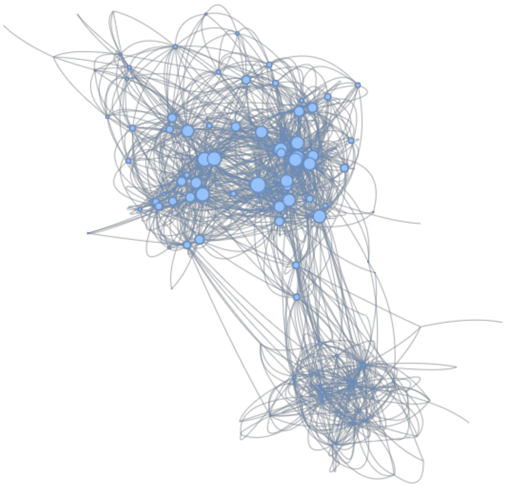

Figure 4.4: eigenvector centrality Graph.

4.5 triangle graph

In graph theory, a triangle graph refers to nodes are connected to each other, forming a triangle shape. In a triangle graph, each node is directly connected to the other two nodes, creating a closed loop.

Understanding the presence and distribution of triangles in a graph can help analyze social networks, biological networks, communication networks, and various other types of networks.

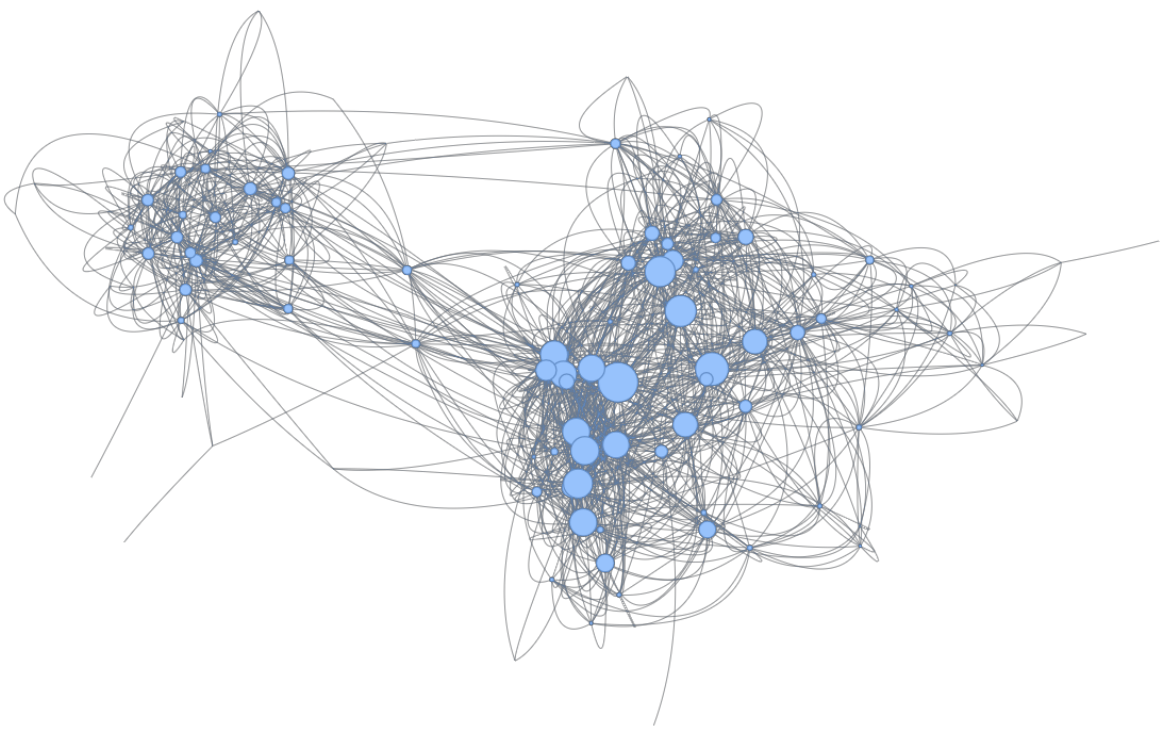

Figure 4.5: triangles Graph.

4.6 cliques Graph

clique refers to a subset of nodes in a graph where every node directly connected to every other node in the subset. In other words, a clique is a fully connected subgraph within a larger graph. A graph can have multiple cliques of different sizes and configurations.

Cliques are often used to study the concept of social cohesion or tightly knit communities within a network. represent groups of individuals who have strong connections and interact closely with each other.

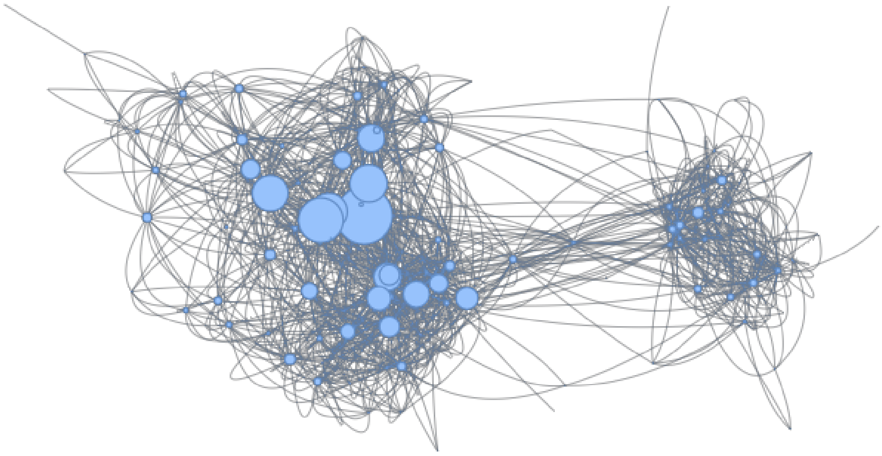

Figure 4.6: cliques Graph

4.7 average degree connectivity Graph

Average degree connectivity, also known as average degree a measure used to quantify the average of edges or connections that each node has in a graph. The average degree connectivity provides insights into the overall connectivity or density of a graph.
It important to note that the average degree connectivity is a global measure that provides an overview of the graph's connectivity but may not capture the local variations or specific patterns of connections within the network.

4.8 kCore Graph

In graph theory, a k-core is a a larger graph where every node has a degree of at least k within that subgraph. In other words, a k-core is a maximal subgraph in which all nodes have a degree of at least k. The concept of k-core is useful in various applications, such as community detection, network analysis, and understanding the robustness or vulnerability of networks. Higher values of k result in smaller and more tightly connected k-cores, while lower values of k lead to larger and less dense k-cores.

In this link, the dynamics of all graphs are placed.

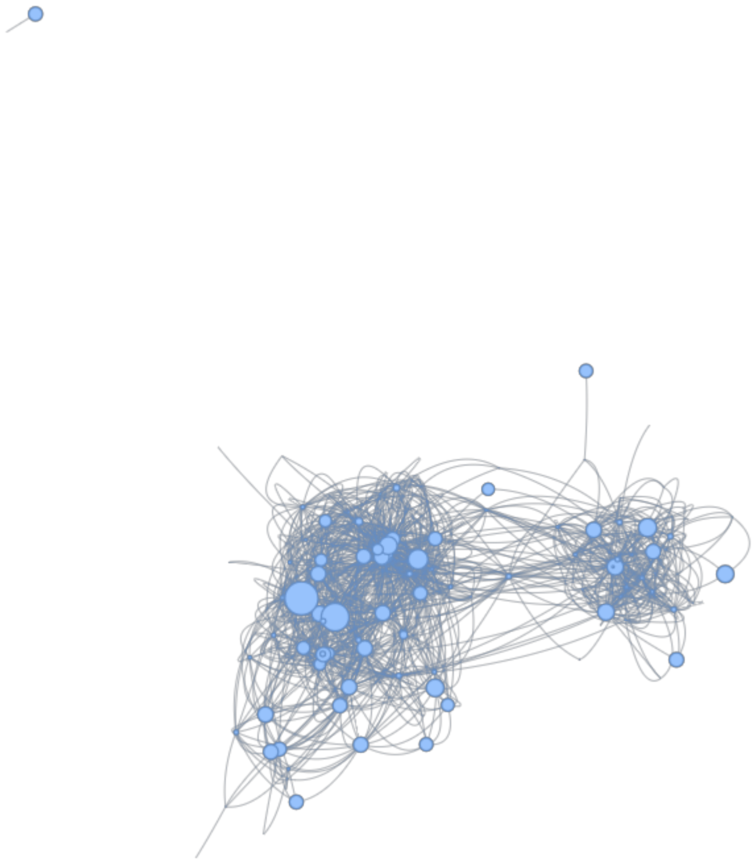

Figure 4.7: average degree connectivity Graph.

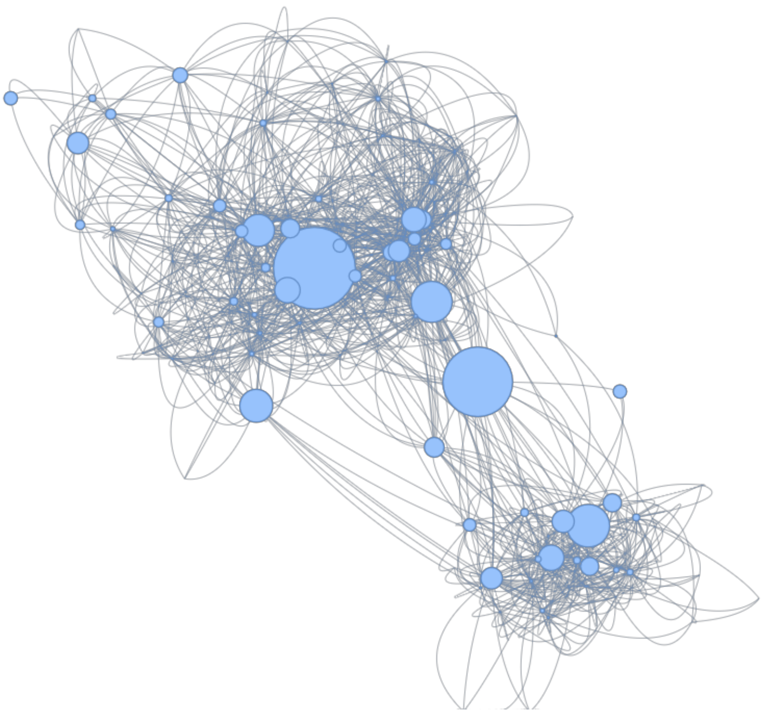

Figure 4.8: kCore Graph.

Chapter 5

CALCULATE SOME IMPORTANT PARAMETERS

In this chapter, we explain some important parameters related to networks and then calculate them for the our Songbird network.

5.1 Density

The density of a network refers to the number of connections or links between nodes in a network relative to the total number of possible connections. It is a measure of how interconnected the nodes are within a network.

the density of a network plays a crucial role functionality, efficiency, and resilience. However, it's important to note that the optimal density can vary depending on the specific goals and requirements of the network in question.

Our network density is equal to:

$$0.1713$$

5.2 Transitivity

The transitivity of a network refers to the tendency of a network to form triangles clusters. It measures the likelihood that if node A is connected to node B, and node B is connected to node C, then node A is also connected to node C. Transitivity is often used as a measure of the level of clustering or cohesion within a network.

Understanding the transitivity of a network can provide insights into the dynamics of social interactions information flow, and the overall structure of the network. It can help identify influential nodes, communities, or potential bottlenecks in the network.

Our network transitivity is equal to:

$$0.5583$$

5.3 Average Clustering

Average clustering, also known as average clustering coefficient, is a measure that quantifies the level of clustering or cohesion within a network. It provides an indication of how likely nodes in a network are to form clusters or triangles. The clustering coefficient of a node measures the proportion of connections between its neighbors that actually exist. It is calculated by dividing the number of connections between a node's neighbors by the total number of possible connections between them.

Average is important and organization of a network. It can help identify communities or clusters within the network, understand information diffusion patterns, and analyze the resilience and robustness of the network to failures or disruptions.

Our network average clustering is equal to:

$$0.6429$$

5.4 Degree Assortativity Coefficient

The degree assortativity coefficient is a measure that quantifies the tendency of nodes in a network to connect to other nodes with similar or dissimilar degrees. It provides insights into the assortativity or disassortativity of a network based on the degrees of its nodes. The degree of a node in a network refers to the number of connections or links it has. In a network, nodes can be classified as-degree nodes (having many connections) or low-degree nodes (having connections).

The degree assortativity coefficient is important because it provides insights into the structural organization of a network It helps understand how nodes with different degrees interact and form connections. It can also have implications for the spread of information, diseases, or influence within network.

Our network degree assortativity coefficient is equal to:

$$0.00027$$

5.5 Average Degree Connectivity

The average degree connectivity, also as the average degree or average node degree, is a measure that quantifies the average number of connections or links that nodes have in a network. It provides insights into the overall connectivity and degree distribution of a network. To calculate the average degree connectivity, you sum up the degrees of all nodes in the network divide it by the total number of nodes in the network.

The average degree connectivity provides a summary measure of the connectivity and degree distribution a network, helping to understand its structure and potential for interactions.

Our network average degree connectivity is:

43: 27.085,	41: 28.219,	28: 25.946,
24: 27.986,	14: 32.405,	29: 25.678,
56: 25.5,	40: 28.6,	30: 26.377,
13: 27.637,	21: 22.691,	33: 30.242,
19: 27.070,	35: 30.143,	22: 25.011,
38: 29.605,	31: 28.290,	23: 32.087,
44: 27.235,	27: 28.741,	18: 28.389,
11: 31.114,	17: 23.412,	39: 28.205,
12: 37.208,	26: 28.169,	25: 24.37,
15: 27.689,	20: 26.75,	3: 26.333,
7: 28.893,	10: 33.05,	4: 22.75,
2: 30.5,	5: 24.12,	6: 21.611,
	1: 9.834	

5.6 Adjacency Matrix

An adjacency matrix is a square matrix that represents the connections or relationships between nodes in a network. It is commonly used to represent the structure of a graph or network mathematically. In an adjacency matrix, each row and column corresponds to a node in the network. The value in each cell of the matrix indicates whether there is a connection or edge between the nodes.

The adjacency matrix is a useful representation of a network as it allows for efficient computation of various network properties and metrics. It can be used to analyze the connectivity, clustering, centrality, and other structural characteristics of a network. Additionally, it can be used as input for various network analysis algorithms and computations.

Our network adjacency matrix is:

$$\begin{pmatrix} 0 & 1 & 1 & \dots & 0 & 0 & 0 \\ 1 & 0 & 1 & \dots & 0 & 0 & 0 \\ 1 & 1 & 0 & \dots & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \dots & 0 & 0 & 0 \\ 0 & 0 & 0 & \dots & 0 & 0 & 0 \\ 0 & 0 & 0 & \dots & 0 & 0 & 0 \end{pmatrix}$$

5.7 Incidence Matrix

The incidence matrix is another representation of a graph or network, similar to the adjacency matrix. However, while the adjacency matrix focuses on the connections between nodes, the incidence matrix focuses on the connections between nodes and edges. In an incidence matrix, the rows represent the nodes of the network, and the columns represent the edges. Each entry in the matrix indicates the relationship between a node and an edge.

The incidence matrix is particularly useful in analyzing directed graphs or networks with multiple edges between nodes. It can be used to compute various network properties, such as the degree nodes, the number of loops, or the connectivity of the network. It is also used in various network algorithms and computations such as finding spanning solving network flow problems.

Our network incidence matrix is:

$$\begin{pmatrix} 1 & 1 & 1 & \dots & 0 & 0 & 0 \\ 1 & 0 & 0 & \dots & 0 & 0 & 0 \\ 0 & 1 & 0 & \dots & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \dots & 0 & 0 & 0 \\ 0 & 0 & 0 & \dots & 0 & 0 & 0 \\ 0 & 0 & 0 & \dots & 0 & 0 & 0 \end{pmatrix}$$

5.8 k-core Graph

The length of a k-core in a network refers to the minimum number of edges that need to be removed in order to reduce the degree of all nodes in the k-core to k or less. In other words, it represents the minimum number of edges that need to be deleted to "peel off" the k-core from the network. A k-core is a subgraph of a network where all nodes have a degree of at least k. It represents most densely connected part of the network, where nodes are highly interconnected.

the length of k-core provides valuable information about the robustness, complexity, and hierarchical structure of a network, as well as the importance of nodes within the network. Our network length of a k-core is equal to:

$$18$$

5.9 Maximum Degree

Our network maximum degree is equal to:

$$117$$

5.10 Minimum Degree

Our network minimum degree is equal to:

$$1$$

5.11 Average number of triangles

Our network average number of triangles is equal to:

$$59.666$$

5.12 Maximum number of triangles

Our network maximum number of triangles is equal to:

$$117$$

You can see the entire calculation process in the code attached to the zip file.

BIBLIOGRAPHY

[1] C. Stahmer. 19 Network Analysis | Adventures in Data Science. https://ucdavisdatalab.github.io/adventures_in_data_science/network-analysis.html, may 8 2022.

[2] J. T. Sakata and D. Birdsong. Vocal Learning and Behaviors in Birds and Human Bilinguals: Parallels, Divergences and Directions for Research, dec 31 2021. [Online; accessed 2023-07-03].

[3] Anusha Senapati. A graph theory based proposal for the analysis of the rna processing pathway in covid-19 disease pathology. 2021.

[4] R. A. R. N. K. A. A. Others. aves-songbird-social | Animal Social Networks | Network Data Repository. https://networkrepository.com/aves-songbird-social.php.