

# Investigating Real and Simulated Complex Networks

Hrishikesh Joshi

October 15, 2021

## Contents

<b>1</b>	<b>Abstract</b>	<b>3</b>
<b>2</b>	<b>Introduction</b>	<b>3</b>
2.1	Food networks . . . . .	4
2.2	Graphs to model networks . . . . .	4
<b>3</b>	<b>Background</b>	<b>4</b>
3.1	Graph terminologies . . . . .	5
<b>4</b>	<b>Graph metrics</b>	<b>7</b>
<b>5</b>	<b>Application of graph theory on food web</b>	<b>10</b>
5.1	How these metrics can be calculated? . . . . .	10
5.2	How these metrics can be applied on a food web? . . . . .	10
5.3	Which measures can be applied to food web? . . . . .	11
<b>6</b>	<b>Use of NetworkX</b>	<b>13</b>
<b>7</b>	<b>Future Scope: Predicting the alterations to a food web using Deep Learning</b>	<b>13</b>
7.1	So how the algorithm will work? . . . . .	14
<b>8</b>	<b>The Research questions</b>	<b>15</b>
<b>9</b>	<b>Experiments and results</b>	<b>15</b>
9.1	Directed Graphs . . . . .	16
9.2	Undirected Graphs . . . . .	17
<b>10</b>	<b>Discussion</b>	<b>20</b>
10.1	Analysis for directed Graphs . . . . .	20
10.2	Analysis for undirected graphs . . . . .	22

<b>11 Contributions</b>	<b>23</b>
<b>12 References:</b>	<b>24</b>
<b>13 Conclusion</b>	<b>26</b>

## 1 Abstract

As the project suggests, Networks can be found anywhere and everywhere. In this research, the use of various graph metrics have been made to uncover some interesting facts about the graph. Following the research, some new information regarding food webs emerged, such as the fact that not all measurements can be applied to both directed and undirected graphs, the number of species in a food web and their connection, and so on. Measures like transitivity and node connectivity are used to determine the dependency among diverse species in a food web, and the flow hierarchy determines if there is any feedback loop or mutual perdition in the food web. The compared numerous food webs using graph metrics uncover similarities and differences in this study. Based on the results of applying numerous graph metrics, the following conclusions have been reached. the research tries to answer various questions, such as. What are the differences between the food webs for directed and undirected graphs? How do species interact with one another? The number of species that can be found in various graphs, The effect of increasing the number of nodes and edges in a data set. If there is a mutual perdition factor in the food web. The impact of graph size and order on the food web. These findings reveal surprising details about distinct food webs and how they interact. In the future, the metrics derived can be analysed using link prediction and deep neural networks to determine the impact of node introduction and elimination.

**Keywords:** Average Degree, Density, Average Indegree, Average Outdegree, Flow Hierarchy, Connectivity, Node connectivity, Average Degree Connectivity Eccentricity, Transitivity, Clustering, Eccentricity, Clustering, Eccentricity, Eccentricity, Eccentricity, Order, Size

## 2 Introduction

Networks are everywhere, from biological networks to transportation networks, from the food web networks to computer networks and from interstellar networks to neurological networks [8][2][5]. Graphs can be used to model entities and relationship between entities in networks . Individual objects, alone does not give enough information, but when the relationship between an object and other objects is considered and analyzed more knowledge can be extracted about the network as a whole. For example, a single cell does not give much information about the immune system. However, modelling the relationships between multiple entities of the immune system (eg red blood cells, white blood cells, antibodies, macrophages, spleen, bone marrow) as a network gives more information about the immune system as a whole. Graphs constructed from these networks are central to analyzing these complex networks. [8]

## 2.1 Food networks

In an ecosystem there exists a complex network with thousands of species which interact with each other in various ways. One such interaction can be described as a predator/prey relationship where the predator acquires energy by consuming prey. A network called as “Food Web” can be formed by considering organisms in food web as nodes and their interaction can be shown as edges. In food webs there exists a hierarchy based on who consumes who. A food web describes habits of the organisms that exist in the food chain and how they survive by consuming others [2] [5] [4].

Food chains can also change due to extinction of a species or introduction of a new species. Organisms play a key role in maintaining the food chain [3] [13] but due to natural and human interactions, some organisms have become extinct and some new species have been introduced due to human migration. Such changes can have a huge impact on the food chain as well as food web [6] [3] [12].

[8] [11] [3] Food chains can also change due to extinction of a species or introduction of a new species. Organisms play a key role in maintaining the food chain [6] [15].

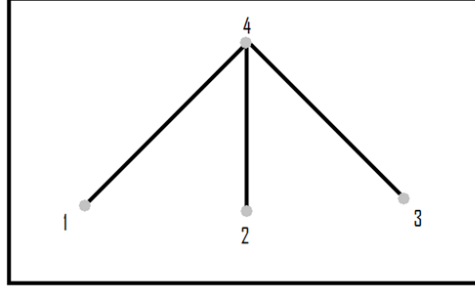
The “x” eats “y” relationship is more complicated in practice where multiple relationships may be encapsulated as a food web. One of the ways to show the complexity of a food web can be an example of a burger. A burger consists of different organisms from different levels of the food chain: patty is produced from beef (from cattle), lettuce (that is consumed by cattle), cheese (that is made from milk from cattle) and a roll (that is made from wheat which is consumed by cattle). The consumer, a human, eats all of these elements. Thus, an organism can eat a variety of organisms belonging to different levels.

## 2.2 Graphs to model networks

Graphs provide an abstraction of interactions between pairs of entities in networks and so are vital to studying relationships in networks [11] [12]. An object would be uninteresting if it were evaluated in isolation. It only gets interesting when you explore its relationships to other objects. Like on Biological level, The links between Cells, Molecules and Biological mechanisms can represent a network. Graphs can model real life networks or they can simulate networks [12] [3].

## 3 Background

The study of graphs, mathematical structures used to model pairwise relationships between entities, is known as graph theory in mathematics. In this context, a graph  $G = (V, E)$  is made up of a set of vertices  $V$  (also known as nodes or points) connected by edges (also called links or lines). Graphs are one of the most important topics in discrete mathematics. [4] Graph theory can be applied to various fields such as computer science, linguistics, physics, chemistry, social sciences,



$$\begin{bmatrix} 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 \\ 1 & 1 & 1 & 0 \end{bmatrix}$$

Figure 1: Example of a matrix of a undirected graph.

biology and mathematics [4] [6]. Here we are going to discuss the application of graph metrics to compare real and simulated networks in food webs.

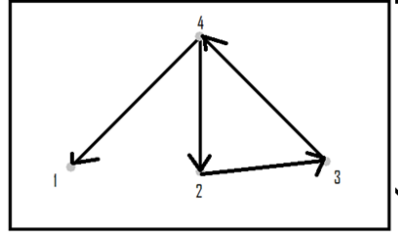
An undirected graph is a graph in which edges have no directions. These are distinguished from directed graphs where the edges have directions (some edges may be multi-directional). [6] The order of a graph is number of vertices and the size of a graph is number of edges in the graph respectively. [6] [14] A graph  $G = (V, E)$  can be represented as an  $|V| \times |V|$  Adjacency matrix where  $A_{i,j} = 1$  if and only if there is an edge from vertex  $i$  to vertex  $j$  (as seen in Figure 1 and Figure 2). A matrix is symmetric if the upper triangular and the lower triangular matrices are the same. The adjacency matrix of an undirected graph is symmetric, but the adjacency matrix of a directed graph may not be symmetric. [6] [1]

For example, the graphs in the figure 1 and 2 can be represented by an adjacency matrix. As shown the food web can be modelled as a directed graph where species are represented by nodes and their relationship is represented by edges [3]. This graph can be represented as binary adjacency matrix. [6] [11] [13].

### 3.1 Graph terminologies

Graphs come with a variety of terminologies, but most of them are self-explanatory as we get to know more about the concept [10]

**Neighbors** : If there is an edge  $u, v \in E$ , a vertex  $u$  in a graph  $G = (V, E)$  is a neighbor of (or equivalently next to) a vertex  $v$ . If  $u, v \in E$ , a vertex  $u$  is an in-neighbor of a vertex  $v$  in a directed graph, and if  $v, u \in E$ , it is an out-neighbor [10].



$$\begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 1 & 1 & 0 & 0 \end{bmatrix}$$

Figure 2: Example of a matrix of a directed graph.

**Neighborhood:** For an undirected graph  $G = (V, E)$ , the neighborhood  $N_G(v)$  of a vertex  $V$  is its set of all neighbors of  $V$ . We're discussing the out neighbors in a directed graph. The union of the neighborhoods of a set of vertices  $u, v$  is their neighbor [11] [6] [10].

**Connectivity and reachability:** If a from  $u$  to  $v$  exists in  $G$ , a vertex  $v$  can be reached from a vertex  $u$ . If all vertices of an undirected graph can be reached from all other vertices, it is said to be linked. If all vertices of a directed graph can be reached from all other vertices, it is said to be strongly connected [6] [10].

**Cycle** :a cycle is a path that starts and finishes at the same vertex in a directed graph. A cycle can be of any duration (i.e. a self loop). A simple cycle has no repeated vertices other than the start and finish vertices, which are always the same [6] [11] [10]. A (simple) cycle in an undirected graph is a path that starts and ends at the same vertex, has no repeated vertices other than the first and last, and is at least three lengths long. Here only basic cycles will be discussed.as the routes,the frequency cycles will be dropped.[6] [11] [10].

**Forests and trees:** forest is an undirected graph with no cycles, whereas agree is one that is connected. When all edges of a directed graph are changed to undirected edges, the result is an undirected forest (or tree). A rooted tree has one vertex that is identified as the root. The edges of a directed graph are usually all directed towards or away from the root node. [10].

**Acyclic graphs that are directed** :A directed acyclic graph is one that has no cycles (DAG) [10]

**Distance:** :The shortest path in a graph  $G$  is the distance( $u, v$ ) between two vertex  $u$  and  $v$  from  $u$  to  $v$ (minimum number of edges). It's also known as the shortest path length between  $u$  and  $v$  [10].

**Diameter** : The maximal shortest path length over all pairs of vertices is the graph's diameter [10]

**Multigraphs** Multi-edges are when a graph allows several edges between the same pair of vertices. Multi-graphs are graphs having several edge [10].

## 4 Graph metrics

A graph metric is a measure of some property of the graph, the vertices or the edges [6] .The graph measures or metrics can be divided into two categories:

**Global measurements** They refer to a graph's global qualities and, as a result, each graph is represented by a single integer.

**Nodal measurements** They refer to the attributes of nodes in a graph and, as a result, are represented by a vector of numbers, one for each node

**Some of the graph metrics are mentioned as below** Most definitions are standard. We refer the reader to wolfram website.[16]

Metric	Definition
Vertex Count	They represent the number of vertices in a graph which is an order of a graph
Edge Count	They represent number of edges in the graph which is size of a graph

Table 1: Basic Metrics

Metric	Definition
Vertex Degree	In an Undirected Graph, The Vertex Degree represents number of edges connected to a Vertex
Vertex In Degree	it represents the number of incoming edges to a selected vertex
Vertex Out Degree	It represents the number of outgoing edges for a vertex

Table 2: Degree Metrics

Metric	Definition
Graph Distance	It returns shortest path between two vertices as length.
Graph Distance Matrix	It returns the distance matrix for a graph
Vertex Eccentricity	It is used to get longest shortest path from a given vertex
Graph Radius	This is used to get the lowest eccentricity between vertices for a graph
Graph Diameter	It is used to get the longest distance between two vertices in a graph

Table 3: Distance Metrics



Metric	Definition
Vertex Connectivity	It shows number of pathways that can be traversed through two vertices that are vertex-independent which is subset of vertices where no two vertices represent the same G edge.
Edge Connectivity	It shows the number of vertices without any edges for two connecting vertices
Graph Density	It shows the number of possible edges for a graph.
Graph Link Efficiency	This shows how tightly the edges in a graph are connected

Table 4: Connectivity Metrics

Metric	Definition
Closeness Centrality	It shows the inverse average distance for every vertex.
Between Centrality	It shows the percentage of shortest routes passing through vertices in a graph.
Degree Centrality	It shows the list of vertex degree for vertices in a graph G which is number of edges incident to vertex
Eigenvector Centrality	It shows the Eigenvector centrality for graph's vertices which is used to find influence of a graph based on given score for high scoring node, If the node is connected with multiple nodes then the score is higher.
Katz centrality measures	they show how nodes can be traversed from a vertex as a hub which can be used for showing influence of a node in a graph
Radiality Centrality	It shows the countable number of paths for a graph's vertices .
Status Centrality	This can be used to determine if relationship relation between graphs are positive or negative

Table 5: Connectivity Metrics

## 5 Application of graph theory on food web

Application of Graph Theory on food web food web is a summary of interactions between multiple species. A mathematics graph is made up of two sets, often denoted  $G(V,E)$ , with the components of  $E$  being relationships between the elements of  $V$ . Both concepts can be graphically described as diagrams of relationships among species or elements, which helps to clarify the connection between the real-world biological system and the abstract mathematical system [8] [14] [12]

A diagrammatic representation of a graph is extremely useful for visual study and understanding of complex interactions and relations. Because of this, it can be difficult to obtain nontrivial information by studying the graph's properties when it becomes large.

### 5.1 How these metrics can be calculated?

From simple measurements such as the order of the graph and edges in a graph, which tells how big and dense it is, to vertex degree measures, which tell how well-connected each node is locally, the Graph measures can be used to characterize graphs.

### 5.2 How these metrics can be applied on a food web?

As can be seen, the matrix representation of a graph is a very convenient way to process data on a computer. In many graph theory applications, matrices are a natural way to express the problem.

We refer to an ecosystem as a specific geographic area that is defined in terms of geographical features such as landforms, plants and animals that live there, as well as the prevailing climatic conditions such as temperature and rainfall, etc. There are many interconnected food chains in those environment which makes a foodweb [4].

**consumer/overlap graph** Comparing a food web's species as vertices of an undirected graph, a food web's competition graph is defined by joining the vertices corresponding to predators that share at least one common prey. The competition graph is also referred to as the consumer graph or the overlap graph because of the way it overlaps with each other [3]. In a similar way, the resource graph in a food web is defined by joining the vertices corresponding to the species (preys) that have at least one predator in common [7].

**The H Clique**  $G = (V,E)$ , where  $V = [v.1], [v.2], \dots, [v.n]$  and  $E = [e.1], [e.2], [m]$  are the vertex and edge sets, respectively. Let's pretend that the edges are positioned at random. A directed graph or digraph is one in which each edge of the graph or digraph has a distinct direction. If every pair of different vertices is nearby, a subgraph  $H$  of  $G$  is said to be complete. A clique, or maximally-completed graph  $G$ , is a graph  $H$  with no neighbours [7].

**The common enemy graph.** So, let's say  $k$  preys on at least two species  $I$  and  $j$ ), and the members of an in-neighbourhood group can help identify the species. There must be at least two edges  $(i,k)$  and  $(j,k)$  in the food chain.  $I$  and  $j$  are connected by an edge in this case, and the food web's Resource Graph, or Common Enemy Graph, has been created [7] [4].

### 5.3 Which measures can be applied to food web?

The directionality of the links is one of the most visible structural aspects of genuine food webs that distinguishes them from other graphs. The transmission of energy from prey species to predators is described by links, which are predator-prey relationships in a complex way. A food web shows how the species are interconnected and how the energy moves from photosynthesis to decomposition among various species forming smaller food chains within a food web. They can also be used to describe the impact of predation on a prey species, taking into account that the interaction between nodes is asymmetrical [14] [2]. To form a food web, it is preferred as a directed graph, often known as a digraph. The adjacency matrix, which is no longer symmetric about the diagonal, reflects the asymmetry. [2][3]. When a network comprises species interactions, the direction of those interactions, and some measure of the connections, such as interaction strength, it displays a food web.

**Here are some of the metrics that can be applied on a food web**

**Basic Measures** Basic Measures such as Vertex count can be used to show number of species in a food web where as Measures such as Edge count can be used to show number of connections in a food web, By analyzing Vertex Count we can find out how many species exists in a food web where as with Edge we can analyses the complexity by knowing the number of connections [16]

**Degree Measures** Basic Degree measures such as Sum of vertex degree shows the number of edges in a graph from each vertex, This will represent the depth of a network, Here by comparing different food webs from various areas such as bay ,deserts or grasslands we can find out how deep food web of a particular region is By analyzing Vertex Indegree and Vertex Outdegree we can analyze the hierarchy in different food webs [16] [2] [9].

**Connectivity Measures** They can be used to find out how the elements in different food webs are connected, By using density measures we get the idea of interactions among the species where as vertex connectivity can show the number of pathways from a node, They can be used to analyse how the relation is formed in a particular food web. As food web is not as simple as food chain where the relation is simple like Predator-Prey, Here the relations can be complex. Like humans do consume sharks, But man-eater sharks are known to hunt humans as well. Another example would be lions and hyenas. They both

are known to kill each other, Also there are some species in nature like tiger Salamander who consume members of their own species, vertex connectivity can be used to analyse the relation between those vertices. As for Graph Link Efficiency, It can be used to analysed how tightly the species are connected [16] [15][1].

**Centrality Measures** [5] Centrality measures are useful to compare food webs as they show the influence of a species, Measures such as status centrality can be used to analyse positive and negative connections in various food webs, Degree Centrality can be analysed to get an idea of a number of species consumed by an animal, like a lion may consume gazelle, zebra and giraffe but in the same environment even though cheetah is a predator, it may not be able to consume larger species like zebra. The Katz Centrality which is Which is utilised to determine an actor's relative level of influence inside a network can be used to analyze the influence of a species from multiple food webs, Like It can be used to analyse the influence of an Apex Predator from different geographical locations, It can be used to analyse and compare the Influence of lion in the Savanah compared with influence of a jaguar in American region and influence of a tiger in the Central and South East Asian Region[16] [5][3].

**Reciprocity and transitivity measures** The Measures such as mean clustering coefficient which is The degree to which nodes in a graph cluster together is measured. can be used to number of nodes are clustered together with a single node, It can be help to analyses how find out which species has the most influence and impact in different food webs where as local clustering coieffienct which is how close the neighbours are to forming a clique in the neighbourhood can be used to subdivide the food web into food chains or smaller food webs and can be used to analyse the species with most influence in each divided part [16] [15] [1].

**Homophily, Assortive Mixing and Similarity measures** Measures such as graph assortivity can be used to divide the food web into smaller food webs, It can be used to analyse and compare the details of smaller sub-groups of a food web. Measures such as vertex corelation Similarity can be analysed for similarity between species eg. A tiger can eat a blackbuck, Along with tiger other species such as Leopard, Crocodile or Wolf can also consume the blackbuck, When compared with different food webs such as Americas, We can see that alligators and jaguars both have a mutual prey where as mean degree connectivity can be used to analyze the complexity of a food web[16] [3] [1] [5]. Measures such as jaccard similarity can be used to determine similarity as the set of animals in a food web is finite, Jaccard similarity can be used to identify the number of species that can be used to identify the similarity in consumption by species Eg. a rat can be eaten by an eagle, snake or other multiple species, This will return the the similarity of connected vertices between two nodes, Also compared with multiple

food web, It can show which species are mutual in different environments like a Leopard can be found in Different food webs across the world [16] [15] [1] [5].

## 6 Use of NetworkX

Network X 2.6.2 will be used in the research. By creating the nodes hash able objects, this tool can be used. This could be an XML file, another graph, a text string, an object, or an image. This is the node object that can be customized None object in Python, on the other hand, is not allowed to be used as a node. To design the food web networks, the Python Network X tool will be utilized. All of these food web graphs and networks will be studied using the Python module Network X. It is simple to construct and uses free software[7].

**The Package can be used for**

- Creation of food web networks
- Determine the number of k-cores and subgraphs.
- Make the random graphs.
- Conversion of graphs into various formats.
- Examine the centre, diameter, and proximity.

There has been some progress in decoding full food web interconnections, which show the feeding interactions and who consumes whom in the biological community. Then this network is being compared with other real life and simulated networks to better grasp how different networks and subgroup networks<sup>11</sup> can be used to understand graph complexity and evaluate them for resemblance.[8] [11][12].

## 7 Future Scope: Predicting the alterations to a food web using Deep Learning

**Introduction of a new species** The introduced species, also known as non-native species, is a species that has spread outside of its native distribution range due to natural or human causes, either intentionally or accidentally. Different impacts on the local ecology. Naturalized species are introduced species that become established and spread beyond their original introduction site.

**Extinction of a species** Extinction is nothing but death of all members of a species or organism. Due to Natural or Human based causes, many species have gone extinct in the past and many are critically endangered which can have a great impact on a food web

## 7.1 So how the algorithm will work?

As we know each species in a food web is represented by a node and their connectivity is shown by vertices and these connections between different nodes form dependencies and influence over each other. A food web is altered by addition of a new species or extinction of one. We consider species as nodes and their interactions are shown by edges. Each Node in a food web (Each Species) in a food web has its role to play as they are interconnected and related and elimination or introduction of a node can alter the entire food web. For example Asiatic Cheetahs used to be main rivals of black bucks, But as they've gone extinct black buck population in the region has been exploded it altered the entire food web as well, Similarly Introduction of new invasive species such as Foxes, Rabbits or Common Myna due to Human migration have altered the food chain of various geographical locations across the world. The algorithm will try to analyse the effect of elimination or introduction of a node(species) from a food web and generate new simulated food webs and matrix can be constructed from these simulated networks which will give the insights to predict how the alteration in a food web will be once a species is introduced or eliminated. To implement such algorithm on a food web.

**Implementation of the algorithm using link prediction** Link prediction is a major difficulty with network-structured data. Link prediction approaches use some scoring functions, such as common neighbours and Katz index, to ascertain the likelihood of a link. As a result, they can be used in a variety of situations. Its readability, scalability, and, in some situations, their ease of use. However, each and every one of them Heuristic makes a strong assumption when two nodes are likely to link. Their use is limited in networks where these assumptions fail.

Instead of employing established approach, a more sensible approach would be to learn an appropriate one from a specific network. We hope to develop a function mapping subgraph patterns to link existence by extracting a local subgraph around each target link, thus automatically learning an approach that suits the present network. The theory integrates a variety of approaches into a single framework and shows that they can all be well approximated using local subgraphs. Our findings reveal that local subgraphs store a wealth of information about link presence. Second, we offer a new method for learning heuristics from local subgraphs using a graph neural network that is based on the  $\alpha$ -decaying theory (GNN). Its experimental results indicate unrivalled performance across a wide spectrum of challenges [17].

Here we are trying to use this link prediction algorithm to generate the newer food webs using such graph neural network algorithms.

Deep Neural Network algorithms can be used, Some of these algorithms are as below

- Monolithic Deep Neural Networks
- Local Deep Neural Network

- Community Based Neural Network

With the use of these algorithms, The effect can be analysed and new simulated networks can be generated to predict how food networks will be if a certain species is introduced or eliminated. For example, If we observe the example of Chesapeake, We can see species such as foxes, mule swans and nutria are introduced to the region due to human migration. Now this has had a great effect on the local food web as their population grew.

## 8 The Research questions

- What is the Plan?

This research will aim to compare and analyse diverse datasets, both real and simulated, using various graph metrics on datasets from various food web networks. The NetworkX package was used to work on the graphs and produce the metrics, and it is a powerful tool for graph analysis.

- Which Graph matrices can be used on food webs? Various measures such as basic metrics like order and size of the graph, measures that show connectivity, In degree and out degree, assortativity, complexity, influence, hierarchy can be applied on the graph
- How different graphs can be compared and analysed to find similarity and difference between various real and simulated food web networks with the application of those matrices?
- How useful these metrics are in this context? The metrics utilised provide some important information that may be used to compare different food web datasets from different environments. This reveals some interesting information regarding how different environments are similar and distinct, as indicated in the results.
- How Graph metrics can be calculated? The graph measurements are the results of analysing the graph; they can range from fundamental metrics like the number of nodes and edges to metrics like distance, probability, coefficient, and influence of a node, among others. To apply these measures, we used the networkX software.

## 9 Experiments and results

The experiments were conducted as follows:

- Firstly, the data which is in the form of adjacency matrix was read from the files.
- Most of these files were in .net format.

- As not all metrics can be applied on both directed and undirected graphs, two different program codes were created for directed graphs and undirected graphs respectively.
- Then the suitable metrics were applied on the data.
- The results for the metrics are shown in the tables below.
- Then with the help of these tables the results were examined.

## 9.1 Directed Graphs

Dataset	Nodes	Edges	AVG:Degree	SD:Degree	Density
Chesapeakebay	36	531	29.5	10.64	0.422
Coachella Valley	30	276	18.4	7.45	0.312
Grasslands	67	124	3.70	2.79	0.028
Heliaster	13	23	3.53	1.98	0.147
Little Rock	182	2427	26.67	21.85	0.073
Scotch broom	154	369	4.79	7	0.015
Silwood	154	369	4.79	7	0.015
Skipwith	35	379	21.65	8.57	0.318

Dataset	AVG:INDEGREE	STD:INDEGREE	AVG:OUTDEGREE	STD:OUTDEGREE
Chesapeakebay	0.42	0.24	0.42.5	0.34
Coachella Valley	0.31	0.21	0.31	0.21
Grasslands	0.02	0.03	0.02	0.03
Heliaster	0.14	0.20	0.14	0.08
Little Rock	0.07	0.11	0.07	0.05
Scotch broom	0.01	0.04	0.01	0.01
Silwood	0.01	0.04	0.01	0.01
Skipwith	0.31	0.33	0.31	0.15

Database	Flow Hierarchy	Strong Connection	Weak Connection
Chesapeakebay	0.78	False	True
Coachella Valley	0.65	False	True
Grasslands	1	False	True
Heliaster	1	False	True
Little Rock	0.90	False	True
Scotch broom	0.98	False	True
Silwood	0.98	False	True
Skipwith	0.89	False	True



The Database	Node Connectivity	AVG:Degree Connectivity	SD:Degree Connectivity	Clustering
Chesapeakebay	4.57	13.31	10.63	0.53
Coachella Valley	2.61	10.56	4.19	0.39
Grasslands	0.07	2.84	1.75	0.13
Heliaster	0.28	3.50	2.42	0.15
Little Rock	1.11	31.11	20.26	0.18
Scotch broom	0.05	10.92	8.15	0.06
Silwood	0.05	10.92	8.15	0.06
Skipwith	2.30	13.22	9.67	0.35

## 9.2 Undirected Graphs

Database	Nodes	Edges	AVG Degree	SD Degree	Density	Diameter
Chesapeakebay	36	482	26.77	8.28	0.76	3
St Martin	23	46	4	2.48	0.18	4
St Martin	44	200	9.09	5.37	0.21	4
Ythan Estury	98	388	7.91	6.67	0.08	5
Ythan Estury Full	134	594	8.86	7.65	0.06	4

Database	AVG Eccentricity	SD Eccentricity	Node Connectivity	Transitivity
Chesapeakebay	2.13	0.34	0.83	0.87
St Martin 1	3.43	0.49	0.18	0.20
St Martin 2	3.04	0.42	0.32	0.33
Ythan Estury	3.68	0.54	0.15	0.19
Ythan Estury Full	3.62	0.48	0.11	0.18

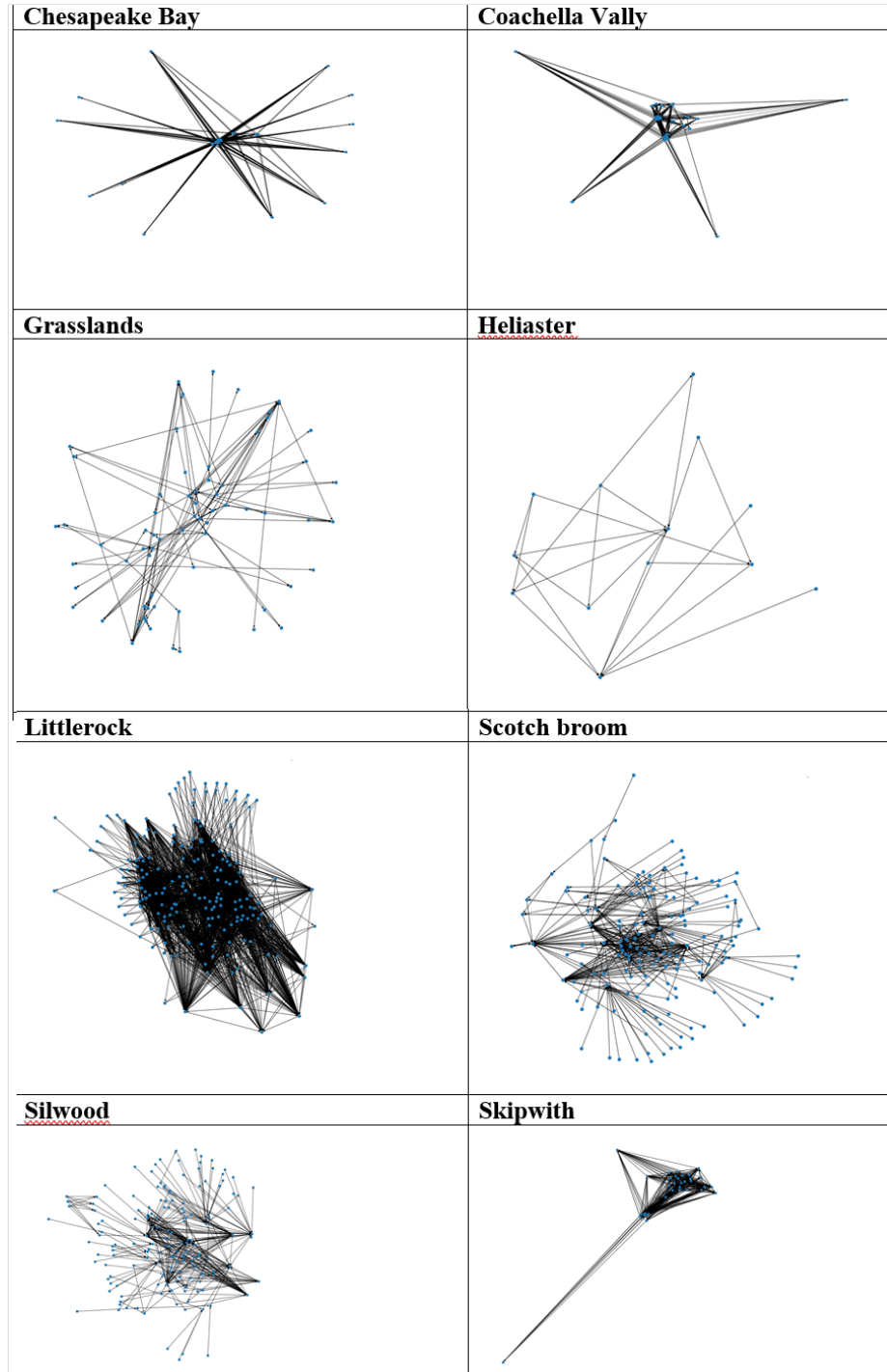


Figure 3: Directed Graphs

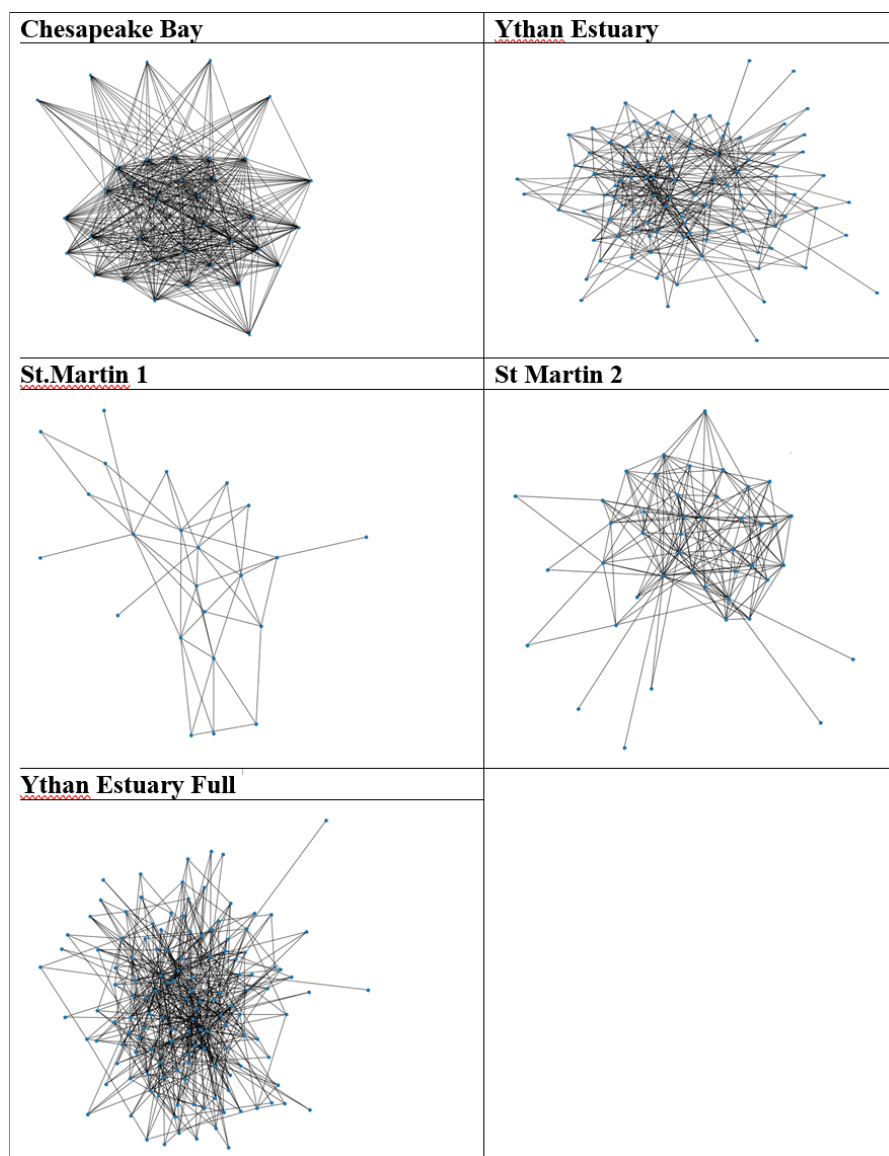


Figure 4: Undirected Graphs

## 10 Discussion

As the analysis on multiple directed and undirected graphs has been done, It can be seen that not all the metrics can be applied on both directed and undirected graphs. The results are generated after suitable metrics have been applied on both directed and undirected graphs For many of the metrics, as the graphs varied in size, there was a need for standard deviation which is also calculated in the results

### 10.1 Analysis for directed Graphs

**Basic Metrics** Here, we can see that the data obtained for the food web for little rock has the maximum number of species followed by silwood and scotch broom (182, 154 and 154 respectively), and the Heliaster has the lowest number of species (13) i.e. participating in a food web Now if we look at order of the graph,we find that the order of the graph is significantly higher for Littlerock and also the number of edges for Littlerock is higher as well (2427) and for cheasepeakbay it's 531. It is lowest for the heliaster

Now, based on the order of the graph and edges, we can calculate the average degree. It has been discovered that Cheaspeakbay has the highest average degree (29.5), and that, despite having a higher number of nodes and edges, Littlerock has a lower average degree (26.67), indicating that it is less complex than Cheaspeakbay. Now, Heliester, the graph with the smallest number of nodes and edges, has the lowest average degree (3.53).

Now as the data size differs for all dataset,Use of average will not give accurate results for every metrics,Hence we use standard deviation as the standard deviation is a measure of the variation in ap erticular graph. The standard deviation of degree for Cheasepeakbay is the highest(10.64) even though nodes and edges for LittleRock are higher,But still the deviation of degree for Cheaspeakbay,Littlerock and Skipwith is higer compared to other graphs.

Even though Littlerock has relatively higher average degree,Yet it's standard deviation is lower (0.07),Which is an example why standard deviation can give accurate analysis. The deviation of cheaspeakbay skipwith and cochelalla vally is higher compared to others,The deviation for Littlerock(The Graph with Most Nodes and Edges) is the lowest

Now, if we look at the density,Cheaspeakbay has the highest density (0.42), followed by Skipwith and Cochellavally, with densities of 0.318 and 0.312, respectively. When we look at grasslands, we can observe that all except one graph is sparse since grassland has an extremely low density of 0.028.

When we compare the basic metrics like graph order, degree, and density, we can observe that Littlerock has the higher graph order (number of nodes) and size (number of edges) while Heliester has is a very small graph. However, But even though the order is highest for LittleRock,as it has lower density and degree compared to cheaspeakbay.

We can see that. The density is not highest. This means that even smaller graphs can have higher density and degree and vice versa. Based on the analysis on the degree metrics, we can see that nodes in cheasepeakbay are very strongly connected.

**Connectivity Metrics** Now the standard deviation of Indegree shows that it is highest for cheasepeakbay at 0.33, indicating that the number of incoming edges per node is highest for the bay, whereas it is lowest for silwood and scotchbroom, at 0.04, indicating that because of the higher number of incoming connections per node, the influence of upper level species in predator-prey relationships is higher for cheasepeak bay. And if we look at Vetex outdegree, we can find that the number of outgoing nodes for the standard deviation for cheasepeakbay is also the highest, at 0.34, indicating that the bay has a stronger influence of species in the lower ranks of the pyramid.

This suggests that cheasepeakbay and skipwith having a high level of complexity, whereas grasslands have the lower level of complexity as the standard deviation for both indegree and outdegree is the lowest, implying that the level of influence among species in grasslands is quite low.

Now if we analyse the connection, If a node can traverse to all other nodes, the connection is considered strong. If it cannot, It is considered weak. As they are directed graphs, every other node cannot be reached from any selected node, indicating that the connection is weak.

If we look at undirected graphs, we can see that there are no restrictions because there is no direction. This means that each node may be reached from any other node, and so the link is the strongest.

**Flow hierarchy** Now as the flow hierarchy means edges that are not participating in cycles

We can see that there are lower degrees of flow hierarchy in the graphs Cheasepeakbay and Cochella valley, with 78 percent and 65 percent, respectively. This suggests that there is a higher amount of mutual predation and feedback loop in the energy flow of these food webs. Mutual predation occurs when two species prey on each other, resulting in the development of cycles. Mutual predation occurs in a food chain. Now, if we look at data from grasslands and Heliaster, we can see that there is no cycle because the flow hierarchy is 100 percent. And while the flow hierarchy for graphs like Scotchbroom, Silwood, Skipworth, and Littlerock is close to 100 percent, there are still a few edges partaking in cycles.

**Assortivity metrics** Now, The Average node connectivity shows that the number of average nodes that must be removed to disconnect two selected nodes is very high for cheasepeakbay, which means that to disconnect two species from the food web, we must eliminate on average four species, which can also be justified by looking at average node connectivity.

So, on average, removing two related animals from a food chain necessitates the removal of four additional species, implying that two species are connected via four different species on average. However, in networks such as grasslands, where node connectedness is on average 0.07, two species can easily be separated if a single species is removed. This will have a negative impact on the food web, making ecosystems like Grasslands, Silwood, and Scotchbroom more vulnerable.

## 10.2 Analysis for undirected graphs

**Basic Metrics** When we compare the directed and undirected graphs of chesapeakebay, we can see that the order of the graph differs depending on the directions, even though the number of species is the same. Directed graph is 482 and undirected graph is 531. The degree of the graph increases from 26.77 to 29.5 as more edges are added, yet the density decreases from 0.76 to 0.42. The graph becomes increasingly thick as there are no directions. For the undirected graphs, the standard deviation of degree is highest for chesapeakebay and it's relatively higher for Ythan's full dataset as well. It's lowest for stmartin's first dataset.

Now if we compare St. Martin 1,2 and both datasets for Ythan Estuary. We can see that as the number of nodes and graphs are increased and as there are more nodes and edges, the graphs become more complex as their density and degree increases.

The Graph's density and degree are directly proportional to each other. When we compare the two datasets, the order of the graph is greater for both, but it's below 45 for the rest of the food webs. Even if the Order of the graph is 36, the number of edges is 482, which raises the degree of the graph to 26.77 and makes it the most dense graph with a density of 0.76.

When we compare the density, we find that chesapeakebay is the most dense, with the exception of the Ythan datasets, which have a density of less than 0.10, making them sparse. Because the Chesapeake Bay has the highest average degree of 26.78 and the highest standard deviation for degree of 10.64, it indicates that the nodes are highly interconnected, making it the graph with the most computational complexity.

**Complex Metrics** When we look at Vertex Eccentricity, we can see that the eccentricity for chesapeakebay is the lowest, with a value of 2.13 and a standard deviation of 0.34. Since eccentricity defines the longest path between two vertex points, this means that the chesapeake bay graph has the fewest hierarchical levels involved in the habitat. The average eccentricity for both datasets is 3.62 and 3.68 respectively, and 0.54 and 0.48 respectively, indicating that there are higher hierarchical layers in predator-prey relationships for Ythan databases.

That is, assuming each food web's food pyramids are taken into account. Lower levels can be found in datasets like chesapeakebay, whereas higher levels can be found in datasets like Ythan. Because transitivity depicts triangles in a graph, it can be used to determine a species' level of dependency in a food web. For example, if species "X" consumes species "Y," and species "Y" consumes

species "Z," what is the likelihood that species "X" consumes species "Z"? This will aid in a better understanding of the predator-prey connection.

Now, if we look at the diameter, it can help us determine how deep a species can go in a food web and what level of influence and dependence a particular food web has. For example, the diameter of the Chesapeake Bay is 3, which means that on average, a species in the Chesapeake Bay has an impact and dependence on other species up to two levels in hierarchy, whereas the impact and dependence of a species in the Chesapeake Bay is only one level in hierarchy.

## 11 Contributions

This research focuses on analysing and comparing various food web graphs, which aids in understanding the similarities and differences between various environments. This analysis explains how various food webs work and how they function.

The graph metrics, which are used to calculate and gather information from a graph, are examined in order to analyse these food webs.

These metrics provide valuable information about a network and can be used to compare and analyse various food webs. The analysis demonstrates how species ranging in size from the tiniest to the largest are linked by food, and how food is similar and different in many environments and geographical regions around the world. It can assist in the identification and comparison of data ranging from the number of species to the influence of a certain species in a food web. And it can also help to identify the food webs which are vulnerable.

We can locate various environments if we look about us. A desert is a dry, chilly environment with minimal water. Forests are densely forested with trees and other vegetation. They are densely forested or have a diverse range of plants. A grassy savannah with plenty of grass for animals to graze. Or in the oceans, where marine life thrives. We can see that these are very diverse, yet the underlying structure of the food web and energy flow remain the same.

Comparing these different environments can aid by generating significant insights that can be used to address questions such as

- how these food chains evolved. and how do they differ from one another?,
- Which species are common in different habitats?
- How did species in the food chain evolve?
- Which food webs are the most vulnerable?
- Which species have the most influence in various food webs, and which have the least?

These concerns can be answered by comparing multiple food webs of different ecosystems using graph metrics and analysing the acquired data and insights

Along with that, ML algorithms can be applied on the derived metrics to get useful insights such as effect of introduction or extinction of a species from a habitat.

If we look into the past, many times introduced species acted as a pest, altering the local food web. Similarly, extinction of a species also had a significant impact as well. For example, in Australia, the thylacine used to be a primary predator until it became extinct, or introduced species such as the common minke created a rivalry with the local noisy miner. When we apply deep neural network algorithms to the obtained graph metrics, we can simulate a new graph if a node is deleted or introduced, and since nodes represent species and connections are represented by edges, it forecasts the impact of introduction and extinction.

The application of deep learning might be quite beneficial, as many species have gone extinct, such as the Asiatic cheetah, and other species have been introduced to non-native places, such as the Common Myna or a Fox, altering the food web dramatically.

The research intends to anticipate the effect of extinction and introduction of a particular species by analysing data collected by metrics using deep learning and based on link prediction. This can help identify the possible impact and significance of extinction and introduction of a species.

As a result, in this research, numerous real and simulated food chains will be analysed and evaluated for similarity using graph metrics. The use of python libraries like network can be used to apply such metrics. This project also has the potential to build fresh graphs in the future by predicting the outcome of extinction and the introduction of new species using Deep Neural Network Algorithms and measurements. The data analysed provide valuable insights into diverse food web networks. These findings shed light on the structure, organisation, and scale of food webs in various geographic areas. It has also been discovered that the findings for directed and undirected graphs might differ significantly.

## 12 References:

### References

- [1] Tavis K Anderson and Michael VK Sukhdeo. Host centrality in food web networks determines parasite diversity. *PLoS One*, 6(10):e26798, 2011.
- [2] Roger Arditi, Jerzy Michalski, and Alexandre H. Hirzel. Rheagogies: Modelling non-trophic effects in food webs. *Ecological Complexity*, 2(3):249–258, September 2005.
- [3] Armando di Nardo, Carlo Giudicianni, Roberto Greco, Manuel Herrera, and Giovanni Santonastaso. Applications of graph spectral techniques to water distribution network management. *Water*, 10(1):45, January 2018.



- [4] Jennifer A Dunne, Richard J Williams, and Neo D Martinez. Food-web structure and network theory: the role of connectance and size. *Proceedings of the National Academy of Sciences*, 99(20):12917–12922, 2002.
- [5] David Easley, Jon Kleinberg, et al. *Networks, crowds, and markets*, volume 8. Cambridge university press Cambridge, 2010.
- [6] FERENC JORDÁN, ISTVÁN SCHEURING, and GÁBOR VIDA. Species positions and extinction dynamics in simple food webs. *Journal of Theoretical Biology*, 215(4):441–448, April 2002.
- [7] Phalguni Mukherjee. Studies on food webs using matrices associated with graphs. *International Journal of Applied Environmental Sciences*, 5(4):603–609, 2010.
- [8] M.D. Plummer. On n-extendable graphs. *Discrete Mathematics*, 31(2):201–210, 1980.
- [9] Daniel L Preston, Abigail Z Jacobs, Sarah A Orlofske, and Pieter TJ Johnson. Complex life cycles in a pond food web: effects of life stage structure and parasites on network properties, trophic positions and the fit of a probabilistic niche model. *Oecologia*, 174(3):953–965, 2014.
- [10] Md Saidur Rahman. Basic graph terminologies. In *Basic Graph Theory*, pages 11–29. Springer, 2017.
- [11] Ferozuddin Riaz and Khidir M. Ali. Applications of graph theory in computer science. In *2011 Third International Conference on Computational Intelligence, Communication Systems and Networks*. IEEE, July 2011.
- [12] Jaka Šircelj, Romi Koželj, and Lovro Šubelj. Effects of species extinction on ecosystems stability. *Uporabna informatika*, 26(3), 2018.
- [13] Morgan Sonderegger. Applications of graph theory to an english rhyming corpus. *Computer Speech & Language*, 25(3):655–678, July 2011.
- [14] Russell K Standish. Complexity of networks (reprise). *Complexity*, 17(3):50–61, 2012.
- [15] Daniel B Stouffer, Juan Camacho, and Luís A Nunes Amaral. A robust measure of food web intervality. *Proceedings of the National Academy of Sciences*, 103(50):19015–19020, 2006.
- [16] Eric W Weisstein. Graph. <https://mathworld.wolfram.com/>, 1999.
- [17] Muhan Zhang and Yixin Chen. Link prediction based on graph neural networks. *Advances in Neural Information Processing Systems*, 31:5165–5175, 2018.

## 13 Conclusion

As a result, in this research, numerous real and simulated food chains have been analysed and evaluated for similarity using graph metrics. The use of python libraries like network can be used to apply such metrics.

The food webs differ for the various environments, Yet the basic structure of the food web remains same . The application of various metrics on food web has given the important insights about food webs in different regions which is explained in results The Metrics applied using NetworkX gives important insights that reveal information about food webs of different environments

This research also has the potential to build newer graphs in the future by employing deep Neural Network Algorithms to anticipate the outcome of extinction and the introduction of new species. Here, We can make use of link prediction using deep neural networks based on the data collected from the graph metrics to predict the outcome of introducing and eliminating a node