Distribution Agreement

Brian Breeden Jr March 24th, 2014

Total Communicability of Temporal Matrix vs Aggregate Graph using a Food Web

by

Brian Breeden Jr

Michele Benzi

Adviser

Department of Mathematics and Computer Science

Michele Benzi

Adviser

Steve Batterson

Committee Member

Christopher Curran

Committee Member

2014

Total Communicability of Temporal Matrix vs Aggregate Graph using a Food Web

by

Brian Breeden Jr

Michele Benzi

Adviser

An abstract of

a thesis submitted to the Faculty of Emory College of Arts and Sciences

of Emory University in partial fulfillment

of the requirements of the degree of

Bachelor of Sciences with Honors

Department of Mathematics and Computer Science

2014

Abstract

Total Communicability of Temporal Matrix vs Aggregate Graph using a Food Web

By Brian Breeden Jr

Node centrality is an important metric in network analysis, allowing the most influential nodes in a system to be identified. This thesis analyzes a directed food web network in order to discover the most influential predator and prey species in the ecosystem, and to examine the different possible constructs for analyzing time-dependent networks. The communicability matrices for each individual month as well as the aggregate graph, and three variations of the temporal communicability matrix were computed. The individual months were compared, revealing a unique top predator species for almost every month, but a mostly fixed list of top preys. The aggregate graph appeared to favor the species expected from comparing the graphs. The temporal matrices, on the other hand, yielded seemingly conflicting results between predators and prey on what resulted in a top species. Predator rankings seemed to depend heavily on which temporal matrix was used, while top prey rankings were relatively stable.

Total Communicability of Temporal Matrix vs Aggregate Graph using a Food Web

by

Brian Breeden Jr

Michele Benzi

Adviser

A thesis submitted to the Faculty of Emory College of Arts and Sciences

of Emory University in partial fulfillment

of the requirements of the degree of

Bachelor of Sciences with Honors

Department of Mathematics and Computer Science

2014

Acknowledgments

I would like to thank my advisor, Michele Benzi, for guiding me through this entire project. I would also like to thank Ernesto Estrada for both suggesting the project for this thesis and providing the data for me to use. Finally, I would like to thank Steve Batterson and Christopher Curran for serving as my honors committee members.

Contents

1 Introduction **1**

2 Background **4**

3 Problem **10**

4 Methods **12**

5 Results and Discussion **14**

5.1 Communicabilities in Individual Months 14

5.2 Temporal Communicability vs Aggregate Graph 21

6 Conclusion **29**

7 Appendix **32**

List of Tables

1 Top Total Broadcast Communicabilities, February 14

2 Top Total Receive Communicabilities, February 15

3 Top Total Broadcast Communicabilities, April 16

4 Top Total Receive Communicabilities, April 16

5 Top Total Broadcast Communicabilities, June 17

6 Top Total Receive Communicabilities, June 17

7 Top Total Broadcast Communicabilities, August 18
8 Top Total Receive Communicabilities, August 18
9 Top Total Broadcast Communicabilities, December 19
10 Top Total Receive Communicabilities, December 20
11 Top Total Broadcast Communicabilities, Aggregate 21
12 Top Total Receive Communicabilities, Aggregate 21
13 Top Total Broadcast Communicabilities, Temporal 22
14 Top Total Receive Communicabilities, Temporal 23
15 Top Total Broadcast Communicabilities, Reversed Temporal 24
16 Top Total Receive Communicabilities, Reversed Temporal 25
17 Top Total Broadcast Communicabilities, Delayed Temporal 26
18 Top Total Receive Communicabilities, Delayed Temporal 26
19 Food Web Species 32
20 Food Web Species continued 33
21 Food Web Species continued 34
22 Food Web Species continued 35

1 Introduction

Network science is a rapidly growing interdisciplinary field, drawing scientists and academics from many disparate fields across science [1]. From describing the interactions of proteins in a cell to the connectedness of companies within an industry or the politics of medieval Europe, network theory can provide new tools for researchers in almost any discipline. The study of networks looks at individual nodes or players in a network, and the connections between them. From this, researchers can discover much about a system that was before unknown.

Centrality is an important concept in network theory with many applications. Using node centrality measures one can find the nodes that are the most essential or influential to a network. This allows researchers to better determine how information flows through a network. Many inferences can be made from finding central nodes, from influential businessmen in a corporate network to relevant pages on the internet. These ideas can give insight that otherwise would be unattainable without such measures.

However, there is no universal method for determining centrality. Several different measures have been derived, each with their own strengths and weaknesses, from degree centrality to subgraph centrality to the PageRank algorithm made famous by Google's search engine. Depending on the network, the different centrality measures can lead to the same results, or very different results. For example, one person in a social network could have the most connections in the network, making him the most central by degree. However, another person could be the single mutual friend between two distinct groups in the network, making him the central connection in communications between the two groups, which other centralities would highlight.

Another facet of networks that has received attention recently is that of the passage of time. Many applications of network theory treat a network as a static, unchanging object. However, especially with social networks, connections between members of the network may come and go over the course of time. Static observations of groups cannot capture the fluidity that defines such networks. In such instances, newer methods of analyzing networks have been developed to try and account for the changes in the network. Aggregate graphs, temporal graphs, and comparing graphs of each time period can all add insight to the evolving structure of the network. As with centrality, no single method is accepted as being the most effective or efficient means of studying time-dependent networks.

One example of a time-dependent network is a food web. In an ecosystem, predator-prey relationships are very important for understanding the make-up of the environment. However, not all relationships are constant throughout the year. Bears may be a major predator in parts of the year, but in winter, during hibernation, they are absent from the food web. This creates a different dynamic in the ecosystem as the seasons change, but does not negate the bears' effect on the ecosystem, even in winter. Despite bears not hunting during winter months, their prey and other similar organisms are kept in check by the predation during other seasons. A simple graph of the food web network taken during the winter does not show this influence. Likewise, a graph taken during the summer fails to account for the bears' absence during winter, giving a chance for prey to recover.

Using centrality measures that take time into account can yield valuable insight into the working of food webs and other networks. Comparing different time periods in a food web can show the ebb and flow of certain species influence, and possibly highlight a central species that would otherwise be hidden behind spikes in the importance of other species. Also, employing temporal and aggregate graphs of the networks can lend further insight by combining the effects of each time period. This thesis will employ all three methods of accounting for time in analyzing the central nodes of a specific food web, in order to show the additional information gained by each, as well as the differences found between the different methods.

2 Background

Network science is the study of networks in the world around us and how the pieces, or nodes, interact. A network is an arrangement of objects that are connected to each other in some way. They can be any object, from people to tree roots to web pages. In network theory, the objects being described are referred to as nodes or vertices, and the connections between them are called edges or links, which connect two nodes at a time. A network is represented by a graph G, defined as $G = (V, E)$ where V is the set of nodes with $|V| = n$ and E is the set of edges $E = \{(i, j)|i, j \in V\}$. In a simple network, all edges are unweighted, meaning that every edge is treated as equal.

The main tool for studying graphs of simple networks is the adjacency matrix A. A is an $n \times n$ matrix. Each element of the matrix a_{ij} is defined by

$$a_{ij} = \begin{cases} 1, & \text{if } (i, j) \text{ is an edge in } G, \\ 0, & \text{else.} \end{cases}$$

Adjacency matrices are very useful in network theory, as most algorithms for studying networks use the adjacency matrix as a representation of the network itself.

There are two main types of networks, based on the reciprocity of edges in the network, namely undirected and directed networks. In undirected networks, all connections between nodes are mutual. An example would be friends on Facebook. If Alice is known to be friends with Bob, then Bob must be friends with Alice. These mutual connections mean that information can travel along each edge in either direction. Thus, in an adjacency matrix for an undirected network, $a_{ij} = 1 \Rightarrow a_{ji} = 1$.

Adjacency matrices of undirected networks are therefore symmetric, and there are twice as many nonzero entries in the matrix as there are edges in the network.

In directed networks, each edge has a direction attached to it. One example would be Twitter followers, as following someone on Twitter does not mean that they are following back, although they have the option to. Another example would be the food web. If a bear eats salmon, it does not necessarily mean that salmon eats bears, and most relationships in a food web are very one-sided. These networks have a flow to them, as information (or nutrition in the case of food webs) mostly flows in one direction along each edge. In the adjacency matrix for a directed network, $a_{ij} = 1 \nRightarrow a_{ji} = 1$. Directed networks have asymmetric adjacency matrices with a number of nonzero entries equal to the number of edges.

One of the main subjects of analysis in networks is that of *centrality*. Centrality measures the degree to which a node influences the rest of the network. Using centrality, one can create a ranking of the most important nodes in a network, or the least. This has been a focus of researchers in network theory since the emergence of the field. As such, there are several methods of measuring centrality.

The most basic centrality measure is degree centrality. In degree centrality, a node's influence is simply measured by $C_d(i) = d_i$, the degree of the node. The degree of a node is the number of other nodes directly attached to the original node by edges. The node with the highest degree would be considered the most important by degree centrality. In directed networks, due to the two possible orientations of an edge, there are two measures of degree. The out-degree, d_i^{out}, counts the number of nodes connected by edges pointing out from the original node. Conversely, the in-degree d_i^{in} counts the number of nodes connected by edges pointing towards the original node. Both of these can be used for degree centrality of directed networks, depending on whether the most important sender or receiver would give the information the researcher is seeking. This split in centrality translates to other centrality measures as well, with out-degree as a type of broadcast or "hub" centrality and in-degree as a type of receive or "authority" centrality.

Another important measure of centrality is Katz centrality. Katz centrality, like many similar centrality measures, is found by manipulating the adjacency matrix to create a vector of centrality values. These manipulations allow centrality measures to take into account indirect relationships between nodes. To do this, the centrality measures look at the entries of A^k which equal the number of walks of length k between pairs of nodes. A walk of length k is an ordered set of $k+1$ nodes i_l where each node is connected by directed edges to the node before and after it. A closed walk is one where $i_1 = i_{k+1}$. Counting all the closed walks in G requires summing the series

$$\sum_{k=0}^{\infty} A^k.$$

While this series may appear useful, without further manipulation the series diverges, which would make the data useless. Thus it is required to somehow cause the series to converge, which is what methods like Katz centrality employ. Katz centrality is measured using the formula

$$K_i(\alpha) = [(I_n - \alpha A)^{-1}\mathbf{1}]_i = \mathbf{e}_i^T (I - \alpha A)^{-1}\mathbf{1} = \mathbf{e}_i^T \sum_{k=0}^{\infty} \alpha^k A^k \mathbf{1}$$

where I_n is the identity matrix of size n, $\mathbf{e}_i$ is the ith standard basis vector, $\mathbf{1}$ is a vector of ones, and α is the factor that makes the series converge. For Katz centrality to be defined, α must be in the range $0 < \alpha < \frac{1}{\lambda_1}$, where λ_1 is the principal eigenvalue of the adjacency matrix. By the Perron-Frobenius Theorem, λ_1 is real [8]. As α tends towards 0, then Katz centrality will give the same rankings as degree centrality. If α tends towards $\frac{1}{\lambda_1}$, then the rankings will reduce to those given by eigenvector centrality, $C_{ev}(i) = \mathbf{q}_1(i)$, where $\mathbf{q}_1$ is the eigenvector associated with λ_1 [2]. This range of possible α values allows researchers to fine-tune Katz centrality to find the level in which they can get the most appropriate measures for the specific network in question. In a directed network, one uses $[(I_n - \alpha A)^{-1}\mathbf{1}]_i$ as a broadcast centrality measure, and $[(I_n - \alpha A^T)^{-1}\mathbf{1}]_i = [\mathbf{1}^T(I_n - \alpha A)^{-1}]_i$ as a receive centrality measure

Another walk-related centrality measure is subgraph centrality. Subgraph centrality uses the diagonal entries of the exponential of the adjacency matrix,

$$SC_i(\beta) = [\mathrm{e}^{\beta A}]_{ii}$$

where the 'inverse temperature' $\beta > 0$ measures outside disturbances on the graph. In most cases, $\beta = 1$. However, like α in Katz centrality, β can be adjusted to work for the graph being used. The exponential is related to the infinite series displayed earlier through the power-series expansion of $e^{\beta A}$:

$$\mathrm{e}^{\beta A} = I + \beta A + \frac{(\beta A)^2}{2!} + \cdots + \frac{(\beta A)^k}{k!} + \cdots = \sum_{k=0}^{\infty} \frac{(\beta A)^k}{k!}.$$

The weight placed on a walk of length k in this case would be $\frac{\beta^k}{k!}$.

Another measure similar to subgraph centrality is *total communicability*. Instead of simply taking into account closed walks, found by the diagonal entries of the exponential, total communicability uses all walks ending at a given node. Total receive communicability is found by

$$TC_i(\beta) = [\mathrm{e}^{\beta A^T}\mathbf{1}]_i = \mathbf{e}_i^T \mathrm{e}^{\beta A^T}\mathbf{1}.$$

Similarly, $e^{\beta A}$ gives a measure of broadcast communicability. Total communicability is especially useful for directed networks, where closed loops may not exist. In those cases, $SC_i = 1$ for all i. In addition, with larger networks, it is faster to compute the row or column sums of $e^{\beta A}$ than to compute the diagonal entries of $e^{\beta A}$ [3].

One main problem of adjacency graphs is how to show the passage of time. There is no place to put time-related information into the adjacency matrix, as all entries a_{ij} have a value of either 0 or 1 depending on the existence of a related edge. To address this, several methods can be used. The first method is simply to make a separate adjacency matrix for each point in time where data is available. The distinct times can then be compared to each other to find trends in the data over time.

Another solution to the inclusion of time is the aggregate graph of the network. The aggregate graph takes the set of adjacency matrices from each point in time $\{A_1, A_2, \ldots A_n\}$, and creates a single adjacency matrix where

$$a_{ij} = \begin{cases} 1, & \text{if } \exists k \text{ such that } [A_k]_{ij} = 1, \\ 0, & \text{else.} \end{cases}$$

This new matrix combines the individual matrices so that if an edge exists at any point in time between two edges, it exists in the aggregate graph. This graph is useful as it gives a single adjacency matrix for the entire time period. However, the aggregate matrix does not take into account what length of time the edge exists; it only measures if the edge exists at all.

When looking specifically at communicability and subgraph centrality of time-dependent networks, one solution is the temporal communicability matrix of a network [5]. This graph takes the exponential of the adjacency matrix for each point in time, and multiplies them together in order, so that

$$C_T = e^{A_1} e^{A_2} \dots e^{A_n}. \tag{1}$$

The temporal matrix thus takes into account the length of each edge's existence, as the longer the edge exists, the more times it is included in the final calculation. Matrix multiplication does not commute, so the order in which the matrices are multiplied matters. Computing the temporal communicability matrix with a different ordering of matrices $A_1, ..., A_n$ may yield a different result than the original ordering.

3 Problem

The study of networks as they change over time is still a relatively new field. Older methods treated the network as a static, unchanging structure. This is useful for some networks, but many are much more fluid. Being able to capture the evolution of the network over time and analyze them at the same level as with static systems is essential to further understand changing networks.

The food web is a typical example of a changing network. As seasons change, certain animal species' effects on the ecosystem wax and wane. A brown bear is not going to show up on a food web during months when it is hibernating. Likewise, animals that rely on deciduous tree leaves will not show up as eating such foods during the winter when there are no leaves. Static network analysis methods cannot take such measures into account. To more accurately understand such networks, the passing of time must be taken into account.

Using data from a food web taken at five different times during the year, this thesis will use temporal analysis to show how the network evolves over time. This will be done with all three mentioned methods of looking at networks over time: static graph comparison, the aggregate graph, and temporal communicability. The food web will also be used as a computational testbed of the three methods, highlighting any similarities and differences between each method, as well as any potential problems or inconsistencies found in the methods.

Analyzing the centrality of a food web can be very important for ecologists as well as mathematicians. On the mathematical side, it provides a common network for testing different algorithms and methods. For ecologists, the data can be used to help preserve or monitor the ecosystem. By determining the top species by centrality, ecologists know which species are the most influential in the group. Then, if such species are dwindling in numbers, ecologists know to ramp up efforts to preserve them, as their disappearance would have a large effect on the ecosystem as a whole.

The food web data that is used in this thesis was graciously provided by Professor Ernesto Estrada, University of Strathclyde, Glasgow, Scotland. Professor Estrada obtained the data from a paper titled "Temporal Variability in Predator-Prey Relationships of a Forest Floor Food Web" [4]. The food web data itself was collected from a forest in Southwestern Cork County, Ireland. The full list of species can be found in the appendix of this thesis.

4 Methods

The data used in the thesis contains 113 species or groups of species. It was originally in the form of five tables, one for each month. In each table, a species was named, followed by a list of the species it preyed on during that month. Species for which there was no predator or prey relationship for a specific month were not listed in the relevant table. In total, 98 species were identified as predators during at least one month, and 50 were identified as prey.

Upon receiving the raw data, an edge list was created for each month, listing each predator-prey relationship found in the data. We then used the `edgeL2adj` code found in the Matlab Tools for Network Analysis toolbox [7] in order to create adjacency graphs of each month.

To compute the communicability matrix for each month, the `expm` function in Matlab was used. Using the `exp` function on a matrix simply takes the exponential of each entry in the matrix; `expm` computes the matrix exponential. Three temporal communicability graphs were made. The original temporal communicability matrix was made with February as the starting point going chronologically forward from there. Then the order was reversed to make the second temporal matrix. The third temporal matrix was created by shifting the starting point to June, and going chronologically forward, wrapping around from December back to February and April. Since the data is month by month, and therefore should be cyclical in nature, this ordering should be accurate. To make each temporal communicability graph, each month's communicability matrix was multiplied in the order mentioned.

The aggregate graph was more complex to create. First an array X that contained each month's adjacency graph was made. Then the aggregate graph *AggGraph* is made and for each i and j from 1 to 113, $AggGraph_{ij} = \max[X]$. This took the entry a_{ij} from each month and set $AggGraph_{ij}$ to the maximum entry, which would be 1 if any month has a link there and 0 otherwise. The `expm` function was then used to find the aggregate communicability.

Once the communicability matrices were formed, each communicability graph was multiplied by a column of ones in order to get the broadcast communicabilities, which translates to their status as a predator. A higher total broadcast communicability signifies a more influential predator in the network. A row of ones was then multiplied by the communicability graphs to get the receive communicabilities, which shows the each species' status as a prey. Similar to the broadcast communicability, a higher receive communicability shows a higher influence as a prey species. Next, a log log plot of each month's broadcast communicability was taken against its receive communicability, as well as the same for the temporal and aggregate graphs, using the `maloglog` function native to Matlab to check for any correlation there. Finally the temporal communicability was plotted against the aggregate communicability for both broadcast and receive, to see how similar the two measures were.

5 Results and Discussion

In this section, the top ten predator and prey species for each month are displayed, along with the top species of each for the temporal matrix and aggregate graph.

5.1 Communicabilities in Individual Months

For each individual month that has data available, the top predator and prey species were determined. These will be compared in this section to look for any possible patterns in the data. Such trends could be used to learn about the ecosystem's structure, or to provide a basis for studying the temporal and aggregate results.

Table 1: Top Total Broadcast Communicabilities, February

Species Number	Species	Total Communicability
14	carabus granulatus	34.99250926
53	family staphylinidae	26.99250926
77	order coleoptera larvae	21.96453646
42	family linyphiidae	12.5279728
88	order nematoda	10.0279728
75	order acari	8.527972799
93	oribatidae	8.527972799
104	pterostichus strenuus	8.027972799
16	deroceras reticulatum	5
51	family sminthuridae	5
78	order collembola	5

February is a sparse month in the data. There are only 31 predator species that have data for this month as opposed to the full set of 98 predator species, so the set of potential top predators is very limited. Due to this, many of the species on the list for February do not end up as high on the lists for other months. The top predator for the month of February is the *carabus granulatus*, a type of beetle common to North

Table 2: Top Total Receive Communicabilities, February

Species Number	Species	Total Communicability
17	detritus	33.34948495
66	microfungi	29.22261227
109	vegetation	27.99250926
12	bacteria	17.69463947
88	order nematoda	14.0279728
108	urtica dioica	8.154845485
75	order acari	8.154845485
78	order collembola	8.154845485
89	order oligochaeta	7.873127314
77	order coleoptera larvae	6.154845485

America and Europe. Most species in the data are arthropods, a classification that includes insects and spiders, among other species.

For the prey, the data is even more limited. Only 19 different preys have data for February. However, the top species are much more common year-round than for predators. In fact, the top four prey ”species” are not animal species at all, but rather types of food for herbivorous animals. Detritus (dead plant and animal matter), fungi, plants, and bacteria act as the main sources of food at the bottom of the food chain. These four food sources act as the indirect energy source for all species in the ecosystem, so some ordering of them is expected at the top of every month's data. As for animal prey species, the top one in February is *order nematoda*, or roundworms.

The lower number of species in February is most likely due to winter conditions. Arthropods are cold-blooded, and so may cope with the cold temperatures of winter by going dormant [6]. Since animals do not eat while dormant, any dormant predator species would be missing from the February data. Likewise, if all relevant predators are dormant, their prey would not show up in the data for February as well.

Table 3: Top Total Broadcast Communicabilities, April

Species Number	Species	Total Communicability
67	nebria brevicollis	73.25770079
101	pteroshicus melanarius	72.90256277
1	abax parallelepipedus	47.20252777
13	bembidion lampros	45.13336834
65	loricera pilicornis	39.95950346
98	platynus dorsale	33.90355786
90	order opiliones	32.54225173
71	olatynus obscurus	30.839864
102	pterostichus diligens	30.839864
62	lacinius epphipiatus	29.04225173

Table 4: Top Total Receive Communicabilities, April

Species Number	Species	Total Communicability
17	detritus	133.3515677
66	microfungi	108.4508651
109	vegetation	85.7338009
12	bacteria	67.28431595
78	order collembola	58.72073958
88	order nematoda	54.95821584
89	order oligochaeta	49.7418067
80	order diptera	41.08953646
75	order acari	40.56776678
81	order diptera larvae	22.96453646

April is a much more active month. There are 61 predators and 38 prey species, although some of those overlap. In addition, none of the top ten predators for February appear as the top predators for April. The top predator for April is *nebria brevicollis*, the European Gazelle Beetle.

The prey ranking is very similar to February. The main four food sources are again at the top, and four of the six other top preys were also present in February.

The top animal prey in April is *order collembola*, an arthropod commonly known as a springtail.

Table 5: Top Total Broadcast Communicabilities, June

Species Number	Species	Total Communicability
101	pteroshichus melanarius	73.64577459
90	order opiliones	59.32268228
1	abax parallelepipedus	48.02216723
65	loricera pilicornis	45.02914293
13	bembidion lampros	40.92257726
62	lacinius epphipiatus	32.35381539
68	nemastoma bimaculatum	30.85381539
14	carabus granulatus	30.42345485
77	order coleoptera larvae	29.5839184
63	leiobunum blackwalli	29.40543055

Table 6: Top Total Receive Communicabilities, June

Species Number	Species	Total Communicability
17	detritus	116.2956221
109	vegetation	107.0483374
66	microfungi	84.60472799
12	bacteria	52.02268228
78	order collembola	46.00866087
89	order oligochaeta	39.81214872
88	order nematoda	39.66350636
80	order diptera	35.15917592
75	order acari	29.22261227
20	family aphididae	28.0559456

June is another busy month, with 69 predator species and 37 preys, again with overlap. The data is much more similar to April than February was, with six of the same top ten predators, albeit in a different order. June also has two top predators in common with February. The top predator for June is *pteroshichus melanarius*, the common ground beetle.

For prey, once again the top four food sources remain at the top, followed by five of the same top preys from April. The only new prey in the top ten for June is the *family aphididae*, a family containing half of all known aphids.

Table 7: Top Total Broadcast Communicabilities, August

Species Number	Species	Total Communicability
90	order opiliones	62.19580959
95	phalangium opilio	47.25924594
62	lacinius epphipiatus	33.85381539
64	leiobunum rotundum	30.85381539
63	leiobunum blackwalli	30.27855786
97	platynus assimile	30.24871238
76	order aranae	21.0279728
14	carabus granulatus	20.53417592
83	order gastropoda	20.0559456
16	deroceras reticulatum	19.0559456

Table 8: Top Total Receive Communicabilities, August

Species Number	Species	Total Communicability
17	detritus	82.6068582
109	vegetation	62.38237325
66	microfungi	58.13237325
12	bacteria	35.36996701
89	order oligochaeta	31.91725173
88	order nematoda	29.5839184
78	order collembola	26.77422743
80	order diptera	25.61938194
108	urtica dioica	20.77422743
87	order lepidoptera larvae	19.96453646

In August, there are 62 predator species and 27 prey species. It is an interesting month as far as top predators go, with only 4 species in common with June and 2 each with April or February. The top predator is *order opiliones*, which are also known as daddy longlegs.

As for prey, the field is surprisingly small, but the top species remain mostly the same as before. The only changes are the tenth species *order lepidoptera larvae*, otherwise known as caterpillars, which are not in the top preys for any other month, and *urtica dioica*, a common nettle plant. It is unclear why a plant species is listed separately from the vegetation category.

There is no data for the month of October. In their paper, McLaughlin, Jonsson, and Emmerson state that heavy rainfall and flooding in the area where they were collecting the data hindered their ability to retrieve a data set for the month of October. This makes the set a bit uneven, but in terms of the analysis does not hurt the results in any measurable way.

Table 9: Top Total Broadcast Communicabilities, December

Species Number	Species	Total Communicability
67	nebria brevicollis	73.92811283
90	order opiliones	39.74871238
61	lacinius ephipiatus	36.24871238
68	nemastoma bimaculatum	34.74871238
97	platynus assimile	32.24683969
53	family staphylinidae	27.84199421
77	order coleoptera larvae	27.64735474
83	order gastropoda	25.5559456
54	family thomisidae	20.2779728
76	order aranae	17.86130613

Like February, December has a much smaller set of species. There are only 40 predator species with data for December. However, there are 32 prey species, which is on par with the warmer months. This could be because while February is completely during the winter, early December is still fall. Thus, depending on when in the month the data was collected, as well as how late the frosts began that year, some species which become dormant in winter may not have done so yet. Continuing the

Table 10: Top Total Receive Communicabilities, December

Species Number	Species	Total Communicability
17	detritus	70.82268228
66	microfungi	58.68373089
109	vegetation	48.88450346
12	bacteria	32.70678819
78	order collembola	24.68469097
88	order nematoda	24.06027604
89	order oligochaeta	22.48876388
80	order diptera	20.36317882
108	urtica dioica	19.60135764
75	order acari	14.98817882

differences from the rest of the data, it only has four top predator species in common with August, and even less with the other months. Like April, however, the top predator is *nebria brevicollis*, the European Gazelle Beetle.

As for prey, it is again very similar, with nine of the same top ten prey species as August. The only different prey in December from August is *order acari*, which includes mites and ticks.

5.2 Temporal Communicability vs Aggregate Graph

This section looks at the top predator and prey species for the entire year, according to the temporal and aggregate communicability matrices. These results are compared to each other as well as to the trends found in the previous section.

Table 11: Top Total Broadcast Communicabilities, Aggregate

Species Number	Species	Total Communicability
67	nebria brevicollis	105.0924618
101	pteroshicus melanarius	99.73435096
90	order opiliones	73.95821584
1	abax parallelepipedus	60.68754628
13	bembidion lampros	57.38711371
95	phalangium opilio	54.91842186
14	carabus granulatus	51.50737325
65	loricera pilicornis	50.99555207
71	olatynus obscurus	46.84878239
102	pterostichus diligens	46.84878239

Table 12: Top Total Receive Communicabilities, Aggregate

Species Number	Species	Total Communicability
17	detritus	190.0412442
109	vegetation	171.6824658
66	microfungi	146.1159166
12	bacteria	84.40208273
78	order collembola	79.74742476
89	order oligochaeta	69.3009126
88	order nematoda	69.06893691
80	order diptera	60.2549155
87	order lepidoptera larvae	52.50117013
108	urtica dioica	50.24555207

The results for the aggregate graph are generally as expected. The only top predator for two separate months, *nebria brevicollis*, is the top overall predator. Second comes the predator with the highest single-month broadcast communicability, *pteroshicus melanarius*. The rest are all species that were in the top predators for at least one month. As the aggregate graph shows the combination of all relationships that were active at any one time, it should favor species that either had a very large contribution in one month, or multiple months with very different sets of preys. *Pteroshicus melanarius* fits the first description well, taking the top predator for June, the busiest month surveyed in the data. The only predator that beat it in the aggregate graph, *nebria brevicollis*, was the top predator in April and December. In addition, it had a higher total broadcast communicability in those two months than the top species in the final two months. These fit the expectation for top predators using total communicability from the aggregate graph.

As far as preys go, it is also relatively unsurprising. The four non-animal groups take the first four spots, with detritus, which has led every month, as the top prey. The only surprising one is *order lepidoptera larvae* reaching the top ten preys, as it only made it for one individual month. However, there is not a large gap between the 10th top prey and the next few preys in each month, and caterpillars were the 11th most central prey in June as well, making the result more understandable.

Table 13: Top Total Broadcast Communicabilities, Temporal

Species Number	Species	Total Communicability
14	carabus granulatus	6055.278712
53	family staphylinidae	5889.148477
101	pteroshicus melanarius	4620.863632
67	nebria brevicollis	4474.815892
77	order coleoptera larvae	4262.7621
13	bembidion lampros	2894.081531
1	abax parallelepipedus	2600.833959
71	olatynus obscurus	2517.509681
102	pterostichus diligens	2517.509681
90	order opiliones	2490.711391

Table 14: Top Total Receive Communicabilities, Temporal

Species Number	Species	Total Communicability
66	microfungi	13062.10727
17	detritus	9947.430056
109	vegetation	9332.768638
12	bacteria	7048.732506
88	order nematoda	6325.212504
89	order oligochaeta	5706.629337
78	order collembola	1336.887082
108	urtica dioica	1203.928975
80	order diptera	884.2455249
100	porcelio scaber	838.2423443

The temporal communicability results are much more interesting. The top two predator species are the top two from February. The top temporal predator, *carabus granulatus*, is only a top five predator in February, and is eighth in two other months, but has data in four of the five months. Comparatively, *nebria brevicollis* is the sole top predator in two months, but only has data in those two months. The second most central temporal predator, *family staphylindae*, or rove beetles, is only in the top ten predators twice, but is one of only two predators to have double-digit total broadcast communicabilities for each of the five months. There are some odd rankings, however. *Olatynus obscurus* and *pterostichus diligens*, the eighth and ninth top predators, both only have data in the month of April. Two of the next three top predators, *order opiliones* and *pterostichus strenuus*, have data in four of the five months.

Looking at the prey, however, there are some unexpected results. Detritus, which was the top prey in each of the five months, is second to fungi in total receive temporal communicability. Looking a little farther down the list, *urtica dioica* is tied with *order acari* in one month and has 20% higher total receive communicability in two months, but does not have data for the other two while *order acari* has its highest values.

One would expect *order acari* to be higher after looking at how *family staphylindae* was the second top predator due to similar circumstances, yet it is not even in the top ten while *urtica dioica* is eighth. In addition, *order diptera*, the ninth top prey, also has data for all five months, and is only behind *urtica dioica* in February, where both species have their lowest total receive communicability. There are differences between matrix multiplication and simply multiplying the communicabilities or even adding the exponents in the formula using the scalar exponent rule $e^a e^b = e^{a+b}$, which may explain why the results were not as expected. That formula does not work with matrix multiplication because matrices do not commute; that is, $AB \neq BA$. Matrix multiplication acts differently than scalar multiplication, but when the original numbers are all higher, one still expects the result to be higher.

Table 15: Top Total Broadcast Communicabilities, Reversed Temporal

Species Number	Species	Total Communicability
67	nebria brevicollis	8883.151301
90	order opiliones	5231.543377
53	family staphylinidae	4075.26337
97	platynus assimile	3168.912788
68	nemastoma bimaculatum	2484.089337
77	order coleoptera larvae	2481.905297
95	phalangium opilio	2318.935459
61	lacinius ephipiatus	2155.764832
83	order gastropoda	1994.228282
63	leiobunum blackwalli	1947.094735

Looking at the reverse-order temporal communicability matrix, more of the predators seem out of place given what temporal communicability is attempting to depict. The top predator, *nebria brevicollis*, is the top predator from December. However, it only has data in two of the five months, and has double the total communicability as in the original temporal matrix. The second predator, *order opiliones*, is only missing

Table 16: Top Total Receive Communicabilities, Reversed Temporal

Species Number	Species	Total Communicability
66	microfungi	12747.58495
17	detritus	9372.396971
109	vegetation	8322.145932
12	bacteria	6955.929115
88	order nematoda	6140.504766
89	order oligochaeta	5490.713182
78	order collembola	936.0239767
75	order acari	607.6932015
100	porcelio scaber	500.53066
80	order diptera	488.8704106

data in February, but has more than double the total communicability as when February was the first month. Further on down the top predators list, *phalangium opilio* and *lacinius ephipiatus*, two species of harvestman spiders, are only active in one month each, August and December respectively. Yet, they are on the top predators list, right above a predator, *order gastropoda*, who has double-digit communicabilities in four out of the five months. In addition, eight of the ten species have data in December, whereas with the original temporal graph, only four of the top ten species have data for December.

The prey is very close to that of the regular temporal communicability. Microfungi again is the top prey despite detritus having a higher total communicability in every month. The top seven preys, in fact, are the same as in the regular temporal matrix. *Order acari*, which was unexpectedly not in the top ten preys before, is now the eighth top prey. However, *porcelio scaber* is now listed ahead of *order diptera*, which has higher total communicability in every month. *Order diptera larvae*, the eleventh top prey, also has higher total communicability in almost every month than *porcelio scaber*.

Table 17: Top Total Broadcast Communicabilities, Delayed Temporal

Species Number	Species	Total Communicability
90	order opiliones	8313.5487
101	pteroshicus melanarius	6877.271729
63	leiobunum blackwalli	5369.875849
1	abax parallelepipedus	4380.93008
65	loricera pilicornis	4307.620949
53	family staphylinidae	4046.251118
14	carabus granulatus	3336.182226
13	bembidion lampros	3161.735456
62	lacinius epphipiatus	2840.751257
77	order coleoptera larvae	2435.529808

Table 18: Top Total Receive Communicabilities, Delayed Temporal

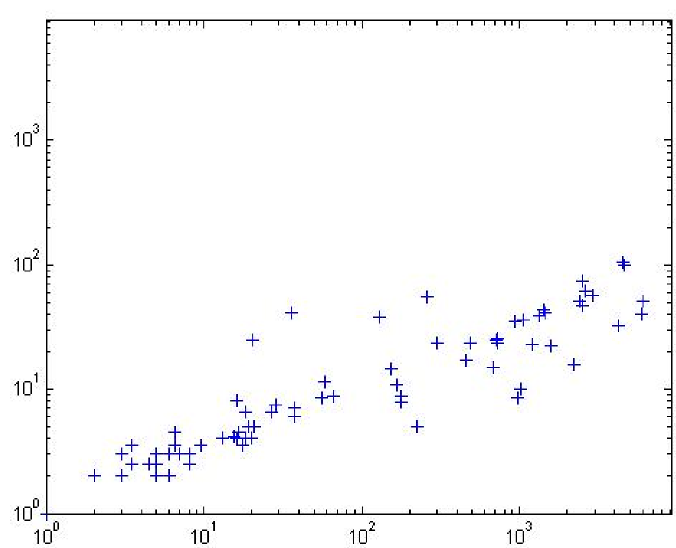
Species Number	Species	Total Communicability
66	microfungi	17229.36996
17	detritus	11900.89248
109	vegetation	10902.33953
12	bacteria	9093.491548
88	order nematoda	8440.631401
89	order oligochaeta	7668.282034
78	order collembola	1319.239965
75	order acari	802.9399933
100	porcelio scaber	758.9998716
80	order diptera	693.726721

Looking at the delayed temporal matrix starting at June, a pattern seems to be forming. The second top predator is the top predator from the month of June, and the top predator for the delayed temproal matrix is the second top predator for June and the top for August. Four of the top five predators using the delayed temporal communicability matrix only have data in two of the five months. However, they all have data during the month of June. All of these come before two species, *family staphylinidae* and *carabus granulatus*, who have data in 5 and 4 months respectively.

When comparing this to the irregularities found in other months, it appears that higher total communicability in months at the front of the multiplication order for each temporal matrix are given more influence over the final result, despite the β for all five months being set at $\beta = 1$. This is problematic, since changing the starting month or reversing the order of the months will change which species are seen as the top predators.

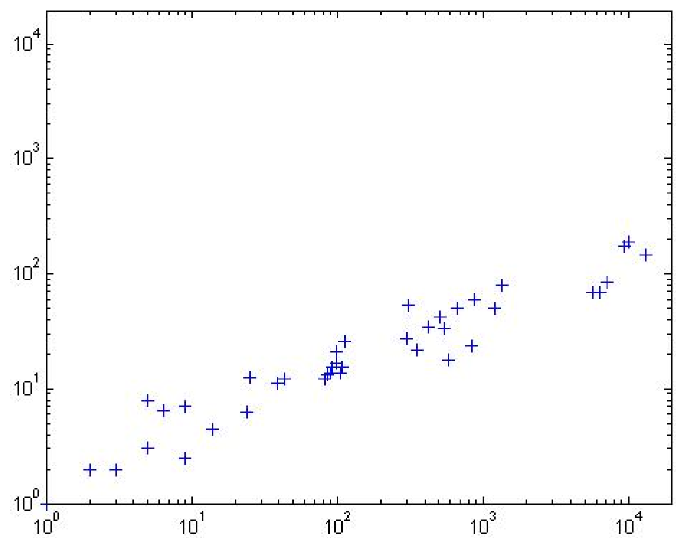
For preys, this conclusion does not appear to hold. The delayed temporal communicability graph has the same exact top ten prey species as the reversed temporal matrix. Between the three different temporal matrices, only the bottom three spots in the top preys change at all, and only between four different prey species. The preys seem much less affected by changing the order of months. This may be due to the stability of the top preys, as there is so little movement in the rankings from omnth to month.

Figure 1: Total Broadcast Communicability, Temporal vs Aggregate



$$\rho_{Temporal,Aggregate} = 0.7931$$

Figure 2: Total Receive Communicability, Temporal vs Aggregate



$$\rho_{Temporal,Aggregate} = 0.9049$$

As these results show, there is a strong but not complete correlation between the temporal communicability and the communicability of the aggregate graph. This means that while the two are related, there are some differences in results between the two measures. Since the objective of the temporal communicability is to take into account the passage of information over time and the duration of each edge's existence, which the aggregate graph does not consider, it can be seen as a positive sign for the temporal graph approach. However, some of the differences are found in the inconsistencies mentioned earlier, where the aggregate graph was closer to the expected ordering.

6 Conclusion

In this thesis, the concept of total communicability was used to measure and determine the top predators and preys in the ecosystem. First, each individual month was studied for trends and commonalities over time. Next the aggregate graph, a combination of each individual month's edges into one single graph, was constructed and used to analyze the data. Finally, the temporal communicability was constructed and analyzed to determine how that measure fit the data.

From month to month, there were varying levels of similarity. Preys mostly stayed the same year-round, especially the non-animal preys found at the top of each month's list. As for predators, there were a lot of different important predators, with no more than six out of the top ten in common between any two months. In addition, there were four unique top predators over the five months surveyed, as opposed to the single top prey being constant through each month.

The aggregate graph revealed results that fit in line with what was expected. By taking each predator-prey relationship that existed in any month and combining them into one simple graph, the aggregate graph gave results matching the highest single-month results for both predators and preys. It was possible for another species to feed on completely different meals at different times of the year, but this did not happen in this data set.

Using temporal communicability was less predictable. Each of the three temporal matrices had different top predators, which seemed to favor the months that led off the ordering. The top February predators were at the top of the original temporal matrix, the December ones for the reversed temporal matrix, and the June ones for the delayed matrix. Also, each matrix had predator species in the top ten which only had one or two months of data, but had data in one of the first two months of the ordering. This apparent bias towards earlier months warrants further investigation to check whether it is a result of the formula, or a facet of the structure of the food web network. For prey, there were more consistent results, albeit ones still not expected from examining the individual months. The prey that had the top total receive communicability in every individual month was not the top prey, but instead second. Also, near the bottom of the top ten in the original temporal matrix, the formula appeared to favor the higher peak over the longer duration. Each of the three temporal matrices had these issues in the prey, and all had at least nine of the same top ten.

For future studies, in addition to testing the ordering of the temporal communicability graph, a major idea to look into is the use of β in the formula for temporal communicability. For this thesis, β was kept equal to 1, in order to ensure the months were weighted equally. However, since β can be viewed as an "inverse temperature," changing it to reflect the change in seasons may yield interesting results. Another possible idea, if one does not want to make the months uneven, is to simply set β as constant but not equal to 1. The parameter β has been found to be more useful for values near 1 [2], but setting $\beta = 1$ may not be the best weight, as agitations in the ecosystem from the method of collecting, or just from uncontrollable circumstances in the ecosystem at the time, could have altered the strength of the bonds in the predator-prey relationships. A value of β different from 1 could better represent such a situation.

References

[1] Estrada, E. *The Structure of Complex Networks: Theory and Applications*. New York: Oxford UP, 2012.

[2] Benzi, M., and C. Klymko. A Matrix Analysis of Different Centrality Measures. TR-2013-014 (2013), Emory University.

[3] Benzi, M., and C. Klymko. Total Communicability as a Centrality Measure. *Journal of Complex Networks*, 1.2 (2013): 124-49.

[4] McLaughlin, O., T. Jonsson, and M. Emmerson. Temporal Variability in Predator-Prey Relationships of a Forest Floor Food Web. *Advances in Ecological Research* 42 (2010): 172-258.

[5] Estrada, E. Communicability in Temporal Networks. *Physical Review E* Article 042811 (2013).

[6] Ennis, B. Winter Survival Strategies of Insects. Colorado State University, 5 Jan. 2010. Web. 21 Mar. 2014.

[7] De Weck, O. *Matlab Tools for Network Analysis*. MIT Strategic Engineering. Research Group: Olivier L. De Weck, Nov. 2011. Web. 21 Mar. 2014.

[8] C. D. Meyer, *Matrix Analysis and Applied Linear Algebra*, SIAM, Philadelphia, 2000.

7 Appendix

The appendix lists the species found in the ecosystem used in this thesis. The species are arranged in the order in which they were found in the data.

Table 19: Food Web Species

Species Number	Species
1	abax parallelepipedus
2	agonum muelleri
3	agonum viduum
4	alnus glutinosa
5	amara plebeja
6	anemone nemerosa
7	arion ater
8	arion distinctus
9	arion hortensis
10	arion subfuscus
11	asellus aquaticus
12	bacteria
13	bembidion lampros
14	carabus granulatus
15	carabus nemoralis
16	deroceras reticulatum
17	detritus
18	discus rotundatus
19	family anthocoridae
20	family aphididae
21	family bethylidae
22	family byhrridae
23	family ceraphronidae
24	family chrysomelidae
25	family clubionidae
26	family coreidae
27	family curculionidae
28	family cynipidae

Table 20: Food Web Species continued

Species Number	Species
29	family delphacidae
30	family diapriidae
31	family elateridae
32	family elmidae
33	family eulophidae
34	family formicidae
35	family gerridae
36	family gnaphosidae
37	family gyrinidae
38	family helidae
39	family hydraenidae
40	family hydrophilidae
41	family isotomidae
42	family linyphiidae
43	family neanuridae
44	family philochomidae
45	family platygastridae
46	family pompilidae
47	family ptiliidae
48	family scarabaieidae
49	family scydmaenidae
50	family silphidae
51	family sminthuridae
52	family sphecidae
53	family staphylinidae
54	family thomisidae
55	family tingidae
56	famly pselaphidae
57	fraxinus excelsior
58	genus microvelia
59	gerris lacustris
60	glomeris marginata
61	lacinius ephipiatus
62	lacinius epphipiatus
63	leiobunum blackwalli
64	leiobunum rotundum
65	loricera pilicornis
66	microfungi

Table 21: Food Web Species continued

Species Number	Species
67	nebria brevicollis
68	nemastoma bimaculatum
69	non-oribatidae
70	odellius spinosus
71	olatynus obscurus
72	oligolophus agrestis
73	oniscus ascellus
74	ophyiulus pilosus
75	order acari
76	order aranae
77	order coleoptera larvae
78	order collembola
79	order diplopoda
80	order diptera
81	order diptera larvae
82	order enchytraeidae
83	order gastropoda
84	order hemiptera
85	order hymenoptera
86	order isopoda
87	order lepidoptera larvae
88	order nematoda
89	order oligochaeta
90	order opiliones
91	order pseudoscorpionidae
92	order psocoptera
93	oribatidae
94	osmunda regalis
95	phalangium opilio
96	phylum bryophyta
97	platynus assimile
98	platynus dorsale
99	polydesmus angustus
100	porcelio scaber
101	pteroshicus melanarius
102	pterostichus diligens
103	pterostichus nigrita
104	pterostichus strenuus

Table 22: Food Web Species continued

Species Number	Species
105	ranunculus ficaria
106	rilaena triangularis
107	trechus obtusus
108	urtica dioica
109	vegetation
110	veronica montana
111	quercus robur
112	ranunculus repens
113	order lepidoptera