The Progress of Prosthetics Technology- Final Report

1. Introduction

The earliest known prosthetic body parts can be dated back to over 3,000 years ago in Egypt as simple extensions using materials such as leather and wood. It is with recent technology that the development of prosthetics has become an increasingly more refined process, using more comfortable, efficient, and affordable materials while also increasing functionality. The network our team has chosen to analyze is the citation network of research articles related to the progress and development of prosthetic technology. More specifically, we are focused on the outer extremities such as feet, arms, legs, etc. A few members of our group that have friends and family that are amputees, and we would like to not only further educate ourselves, but we would like to educate others as well regarding prosthetics. Our goal is to also utilize this network to track the progress of development over a short period of time and potentially predict where the future of prosthetics technology lies. To do so, we chose to build our network using articles around the time of World War Two (1940s). We gathered several research papers from each decade (from 1940s-2020s) and then we put those papers into the researchrabbit.ai software. The software then connects our nodes (research papers) and our edges (an article that cited that node). Initially, our project was going to have around 1,000 nodes, but after we refined our search to only outer extremities it severely decreased the size of our network to around 80 nodes and 170 edges.

To refine our analysis further, we divided our research questions into the following three areas of focus: authors, date of publication and keywords. Regarding the authors of the articles, our focus was to observe any trends in unique authors with the network. This should cause nodes to fall in line with community behavior, since the hubs of a network would have a significant number of unique nodes connecting to it. Analyzing the recurring keywords within articles, the network should display a low closeness centrality between nodes sharing an attribute like keywords. We will do so using researchrabbit.ai, keyword filtration system in tandem with the network in Gephi. The goal of last analysis, dates of publication, is to determine whether the year the article was published correlates with the degree of a certain node. We predict that nodes with publication dates post World War 2 (1945) will display a higher degree. We developed our network by searching for articles in PubMed by decade and then by relevance. We then added them into Research Rabbit and used the 'Later Works' function to generate a network based off papers that cited the ones added. We used this to manually create files to export into Gephi for further analysis.

2. Initial Network

The network we have generated has a total of 79 nodes and 164 edges (**Figure 1**). The network is a directed one and the weights of the edges are one since a research paper would not cite the same paper twice in the bibliography of the article. Upon initial inspection of the graph shown in **Figure 3** below, we can vaguely discern one or two potential hubs. We can also make out a few scattered nodes with a degree of zero that are their own separate components. Running the Network Overview (shown in **Figure 2**) statistics on Gephi, we were able to ascertain that our graph density was not very dense with a result of G=0.027. The nodes with a degree of zero will cause a modularity towards the left of our degree distribution, while the hubs will become outliers with only a few of them having a significantly high degree.

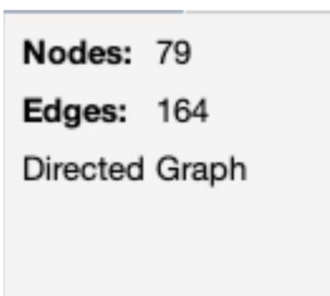


Figure 1: Number of nodes and edges in the network

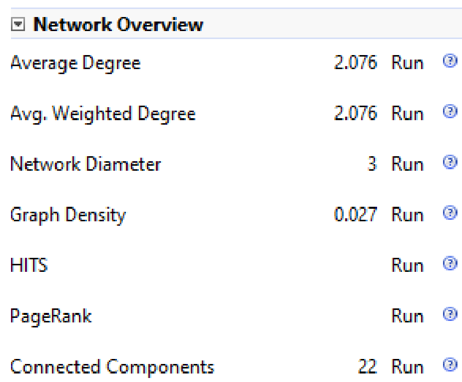


Figure 2: Network Overview statistics of the network in Gephi

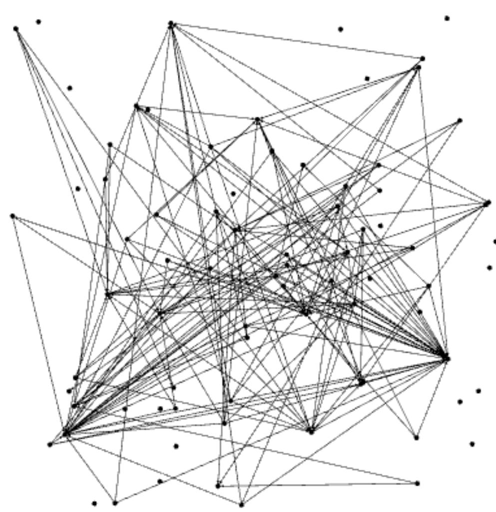


Figure 2: Initial graph of the network on Gephi

3. Layout Algorithms

The “Yifan Hu” algorithm runs extremely fast and stops automatically. The algorithm reduces complexity by utilizing a force-directed model, in which the forces are powered by a Barnes-Hut calculation, ultimately making the speed of the calculation n^2 rather than *n log n*. This algorithm clusters the nodes by hierarchy of degree. The nodes with a higher degree are towards the center of the graph, while those with a lower degree are pushed out. Therefore, the nodes with no links are left in the distance. From the network shown in the **Figure 4.1** below, we can see that most of the node’s cluster in one large community, while all of the nodes with a degree of zero have been scattered away from the connected component.

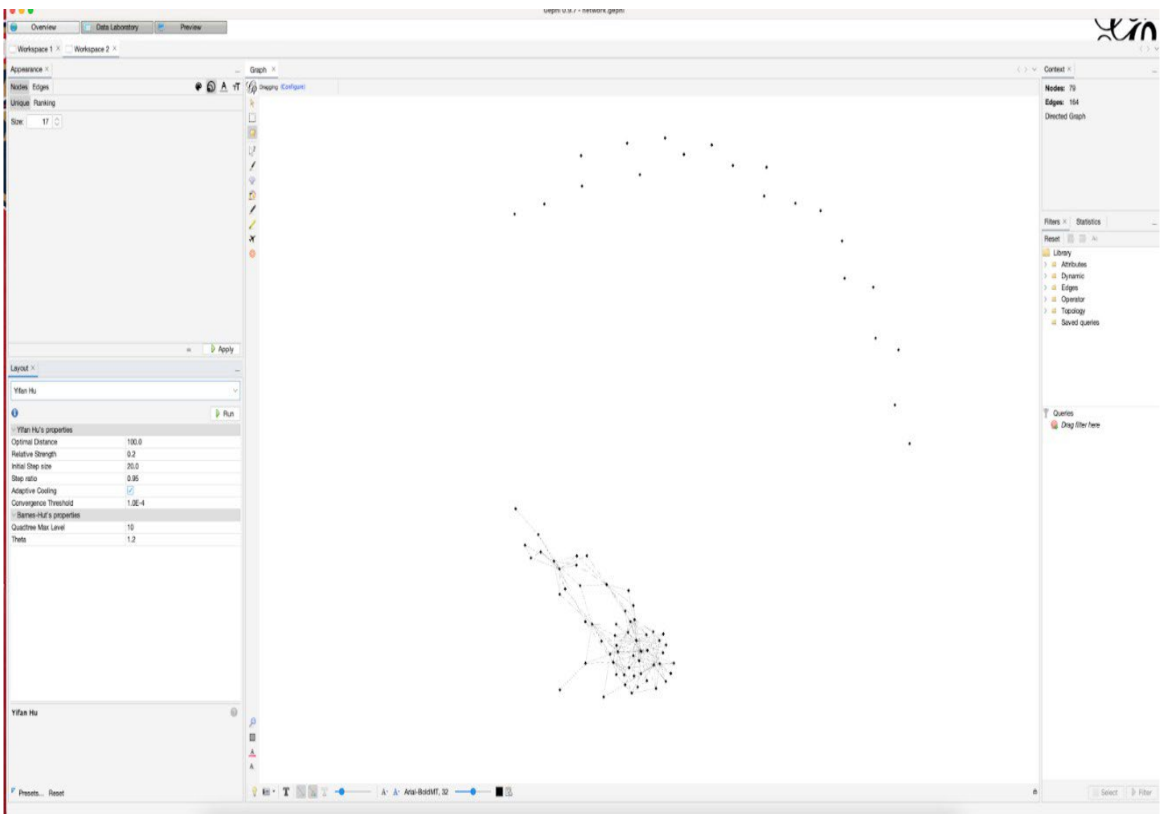


Figure 4.1: Yifan Hu Layout of network.

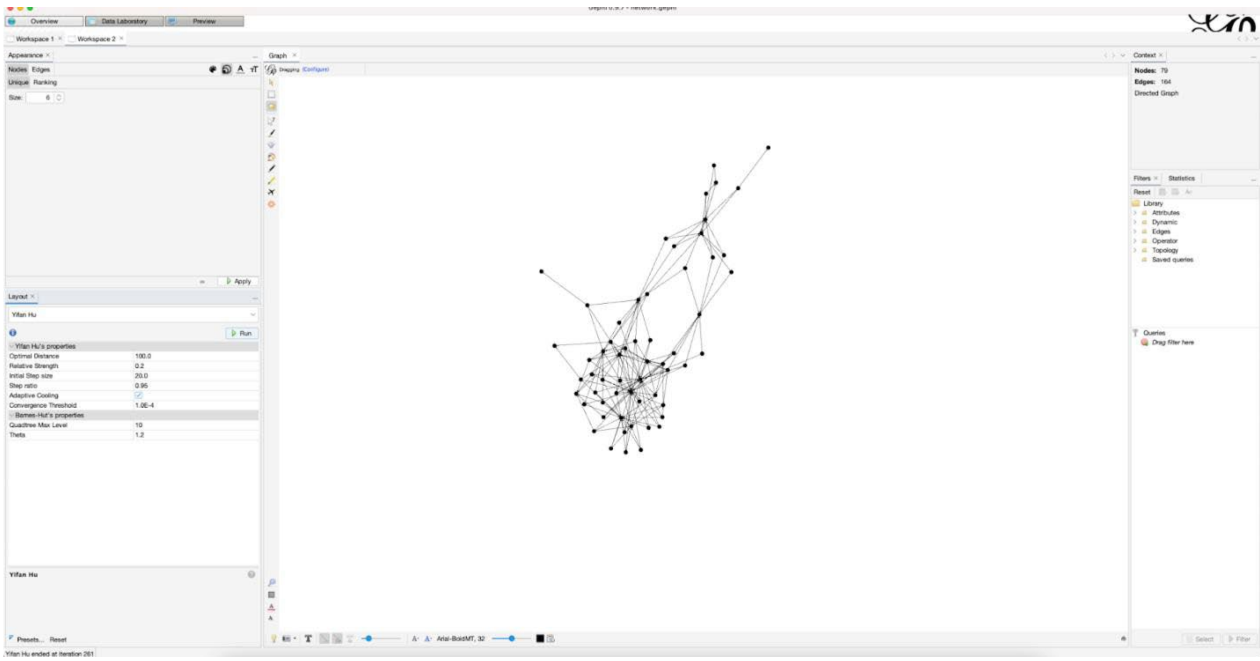


Figure 4.2: Yifan Hu layout of network excluding nodes with a degree of zero.

The “Fruchterman Reingold" layout simulates nodes as mass particles and edges as strings. The edges are treated as undirected regardless of the graph layout and displays the network in a circular format. This allows us to understand the topology of the network. This is due to the algorithm placing topologically near nodes in the same cluster while those that are placed far from each other are topologically far. The nodes that have no edges are pushed to the outside. This layout works well as shown in **Figure 5** below as it allows the hubs of the network to be emphasized in the center of the graph.

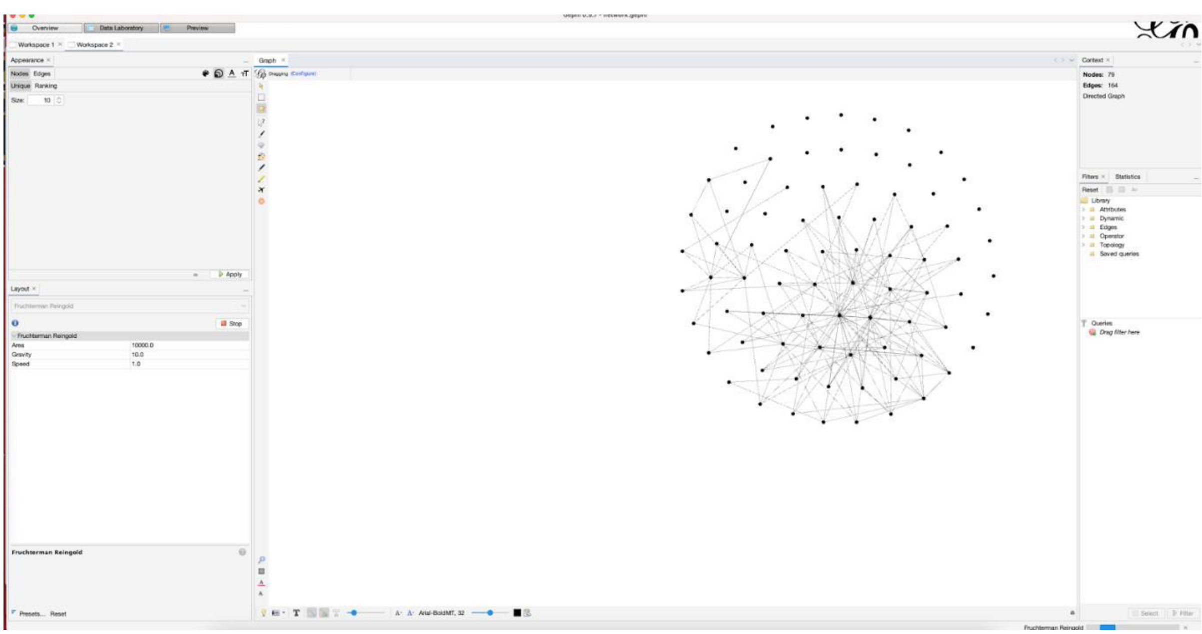


Figure 3: Fruchterman Reingold layout of network.

Force Atlas: The “Force Atlas” layout works well for smaller networks such as this one. It is also useful for exploring the network as it doesn’t have any biases when plotting. With a complexity of $O(N^2)$, the algorithm pulls strongly connected nodes together and pushes the weakly connected nodes apart from each other. **Figure 6** clearly displays the nodes with no edges pushed to the outside of the network. In addition, the hubs and well-connected nodes are centered together on the right side of the graph while the nodes with lower degree values are on the left side. This layout displays a network very similar to the Yifan Hu layout shown previously.

Figure 4: Force Atlas layout of network.

The “Yifan Hu” algorithm seems most useful due to its ability to cluster the nodes by hierarchy of degree. The nodes with a higher degree are towards the center of the graph, while those with a lower degree are pushed out. Therefore, the nodes with no links are left in the distance. This allows the user to visually understand which nodes in the network have the highest degree value. It also allows the user to interpret the clusters of nodes that are most connected with each other.

4. <u>Emphasizing the Network</u>

In order to add emphasis to the diagram, we have sized the nodes by a degree ranking, with the minimum size being four and the maximum size being twenty-four. The nodes have also been colored by degree ranking, showing that the darker nodes represent hubs in the network. The nodes have also been labeled. The **Figure 5** shows the nodes labeled with a fixed size. **Figure 6** displays the node label size with node size as an attribute.

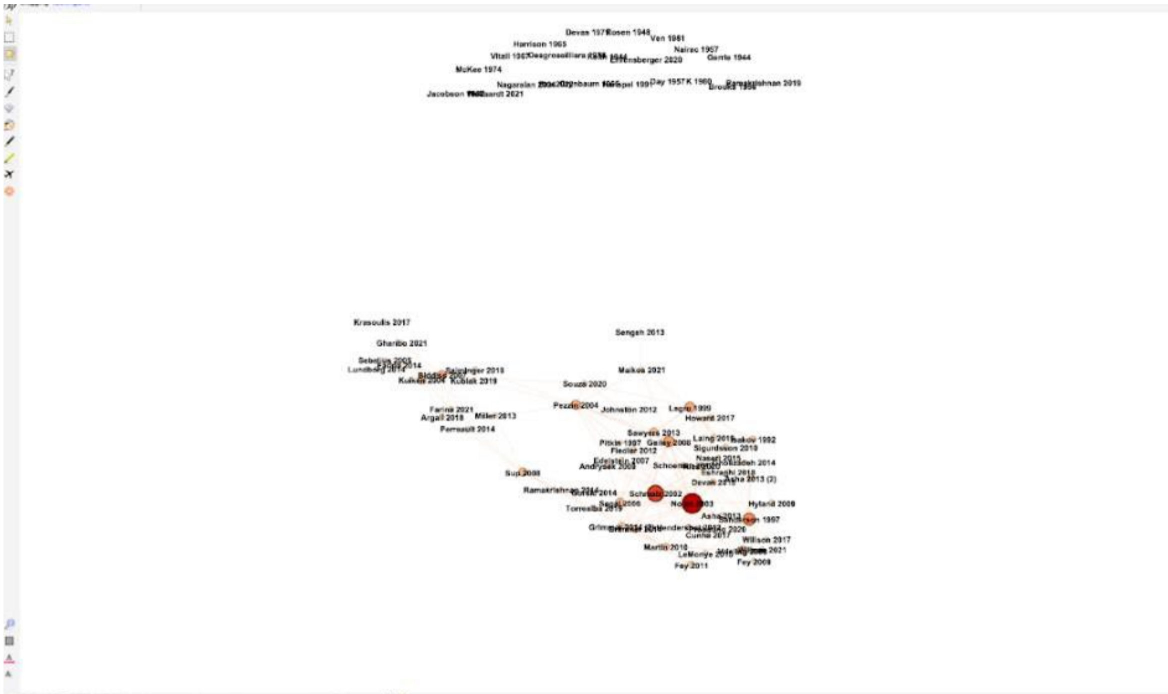


Figure 5: Network emphasized thought color coding of nodes by degree ranking and fixed size.

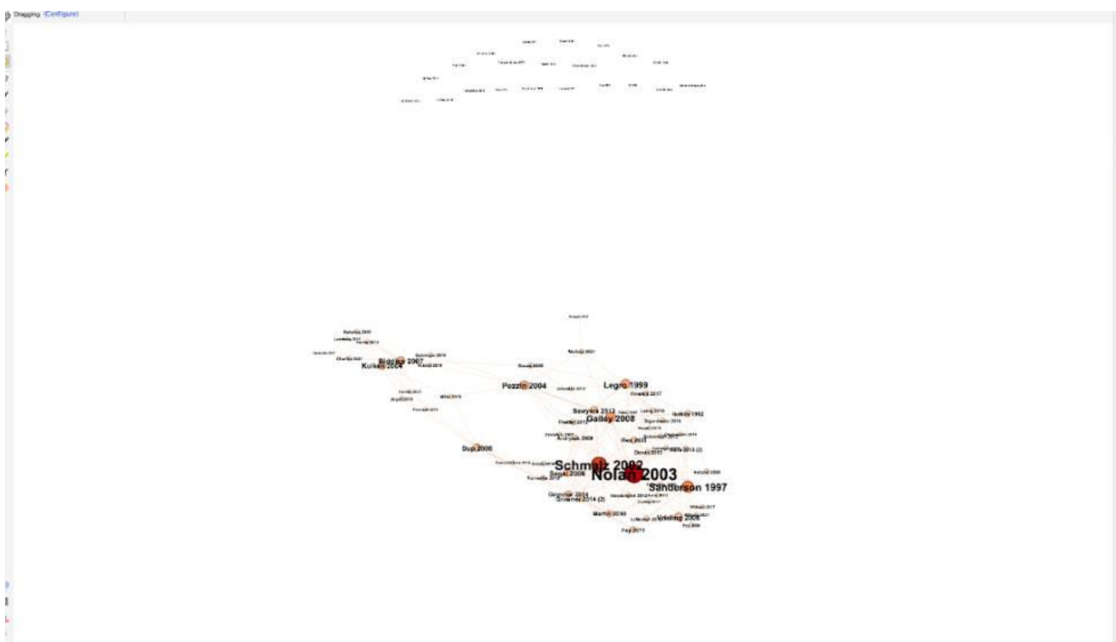


Figure 6: Network with nodes sized according to degree rank.

This new diagram has helped visualize which nodes are most important in terms of degree, closeness centrality, and betweenness centrality. The sizing and color-coding of the nodes being filtered by degree ranking makes it extremely noticeable which nodes in the middle of the graph have higher degree. For example, it is clearly visible that Nolan 2003 has the highest degree of all (k=29) the nodes, with Schmalz 2002 being a close second (k=23).

5. Statistics

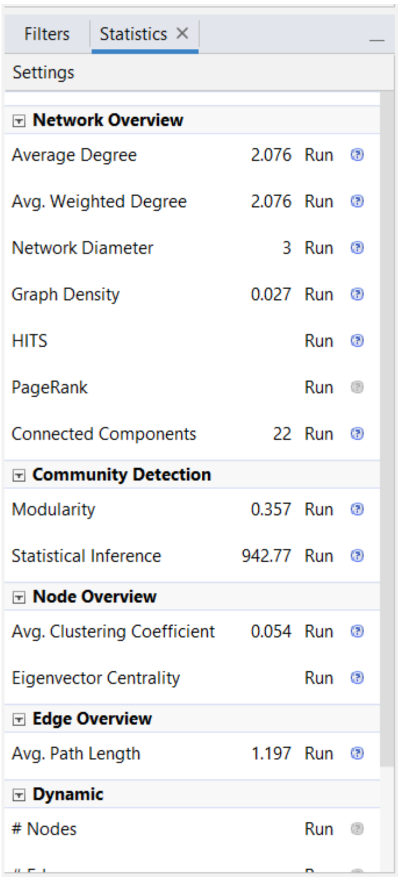


Figure 7: All statistics run on Gephi

Degree distribution (shown in **Figure 8** below) measures how many edges a given node has. Random networks fit a Poisson distribution while small world networks will tend to follow a power law distribution. This means that nodes with lower amounts of edges appear in a network with higher frequency. Looking at our dataset, we notice that nodes with lower edges tend to underperform what would be expected if the dataset followed a power law distribution. Due to our dataset not including all citations, this phenomenon is likely caused by the construction of the network by ResearchRabbit.ai than a fundamental fact about the data.

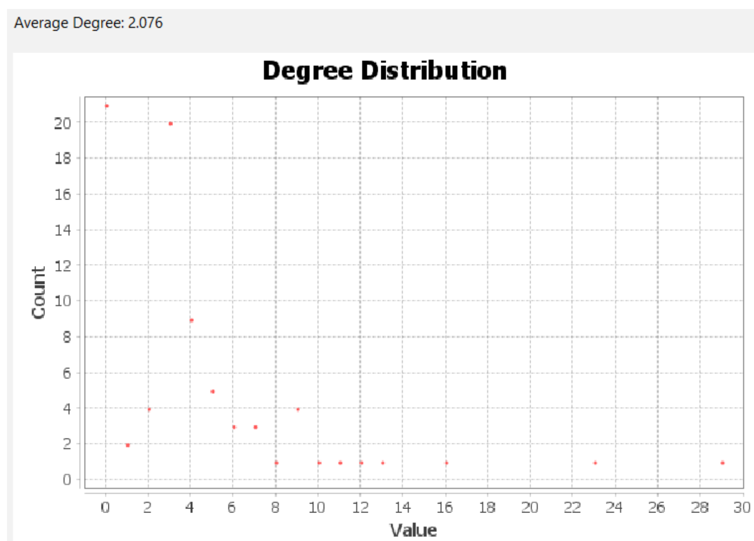
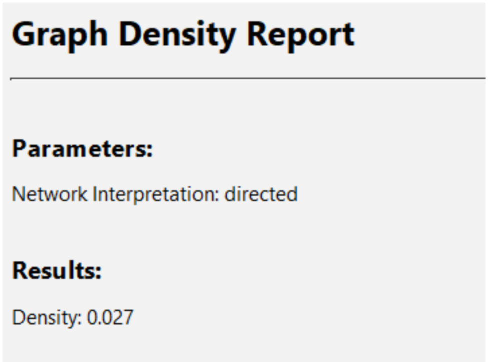


Figure 8: Degree distribution of network.


Graph Density Report

Parameters:

Network Interpretation: directed

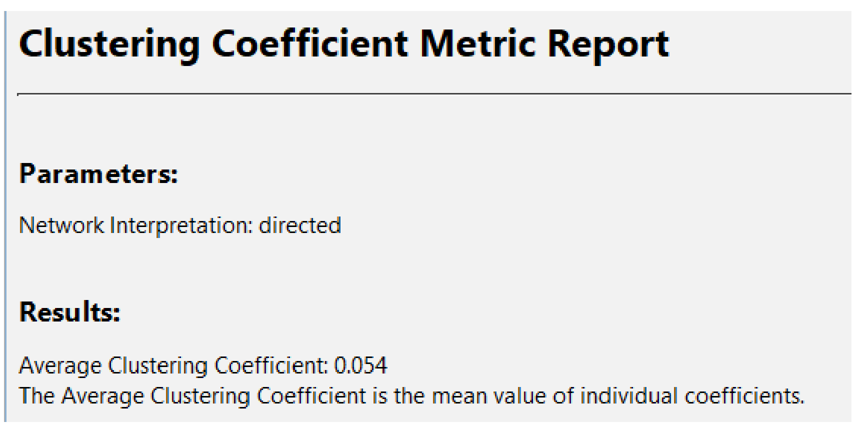
Results:

Density: 0.027

Figure 9: Graph density report

Density is a statistic about the network that ranges from 0 to 1. Gephi calculated the graph density to be 0.027 (**Figure 9** left). A density value of 1 means the network is a complete network with all nodes sharing edges. Density is defined as edges divided by total edges possible. Since this network is directed, the formula for maximum edges is n*(n-1) where n is the number of nodes a given network has. While Density is normally a small value, our network has an especially low density. This is caused by the number of unconnected nodes in the network bringing the total average down.

The average clustering coefficient is a network science statistic that measures the tendency of nodes in a cluster to form communities and cluster together. This value ranges from 0 to 1. The average clustering coefficient is calculated by taking an average of each node's clustering coefficient value. This value is calculated on each node by finding the percent of this node's neighbors that form edges with the node's other neighbors. A value closer to 0 indicates nodes do not form communities while a value closer to one indicates this network tends to form strong communities. Our network has a low average clustering coefficient value of 0.05. This indicates the network does not have a tendency to form communities. Since a good portion of our network's nodes do not have edges, these nodes will bring this average down. We can see from the network pictured in **Figure 10** that there are two distinct areas within the network that display community-like behavior.


Clustering Coefficient Metric Report

Parameters:

Network Interpretation: directed

Results:

Average Clustering Coefficient: 0.054
The Average Clustering Coefficient is the mean value of individual coefficients.

Figure 10: Clustering coefficient metric report

6. <u>Filters</u>

The first filter our team decided to run was the "Partition Hub" filter (results shown in **Figure 11**). This was done to show how many nodes withing the network displayed hub-like behavior. Two distinct hubs can be made out from our emphasized graph, but after running the filter, we learned that 43% of the nodes in our network had a degree high enough to be considered hubs.

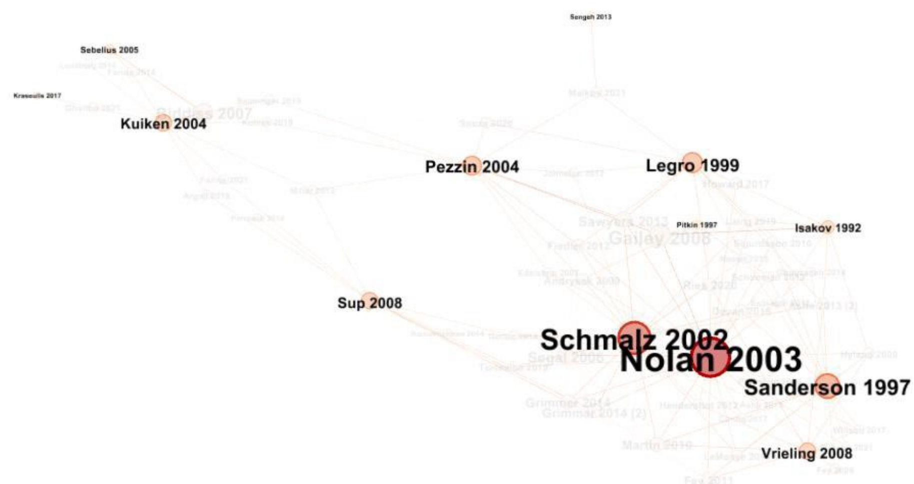


Figure 11: Partition hub filter

The next filter we decided to include was Modularity class. In this specific filter, there were 4 classes that made up roughly 73% of the graph. The percentage rates were 25.32, 18.99, 15.19, and 13.92 respectively. In **Figure 12**, we can see there are one or two decent sized communities, however, most of the nodes have a smaller degree. Then in **Figure 13**, we see the opposite happening. The communities have a high number of edges, but they have less nodes. In **Figure 14**, there are not as many hubs and there is a small number of nodes. **Figure 15** has a similar issue. However, they are extremely close in modularity class distribution. Overall, this network has a weak Modularity community because there are four sizeable communities.

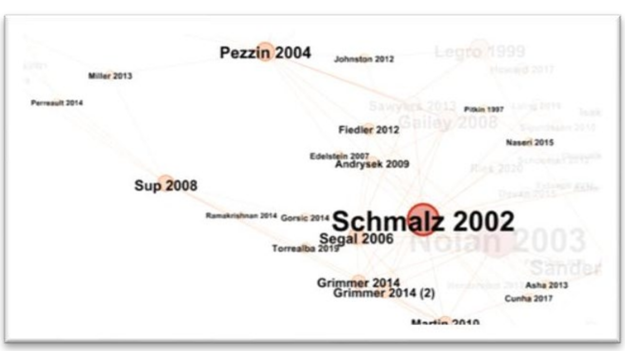


Figure 12: Modularity filter 25.32% group

Figure 13: Modularity filter 18.99% group

Figure 14: Modularity filter 15.19% group

Figure 15: Modularity filter 13.92% group

7. Notable Features

Homophily measures a network's tendency for nodes with similar attributes to cluster together. We were unable to determine from our network alone if homophily exists or does not exist due to the network being small. Additionally, the network having a small clustering coefficient tends to suggest homophily does not exist as no tendency to cluster regardless of attributes exist. While searching for foundational papers for our network, we noticed some nodes contained certain keywords within papers around a certain time period exhibited high clustering tendencies. For example, in the 1940s research was focused on dental prosthetics and we found it difficult to find research on limb prosthetics. Research in the 1960s to 1980swe tended to focus on understanding muscle movement patterns. We saw an increase of degree per node starting in 1967 indicated by the node labeled Jacobson 1962.

The network displays two distinct communities. However, one community is slightly larger than the other. The communities make up the large component, so they still interact with each other as opposed to the single node components that do not interact with each other at all. The larger community could be considered a giant cluster, but our network is too small to say definitively. Using research rabbit, we were able to discern a few trend keywords of interest within the network such as amputee, neural, implant, and gait. The results are shown in **Figure 16**, **Figure 17**, **Figure 18**, and **Figure 19** respectively. Most of the nodes in the network contained the keyword 'amputee', including the nodes with degree 0, and 'gait' appeared most in the larger community. The keyword 'neural' was mostly present within the smaller community, while the keyword 'implant' appeared most in the smaller components of the network. These trending keywords tie in with the community behavior displayed in the network.

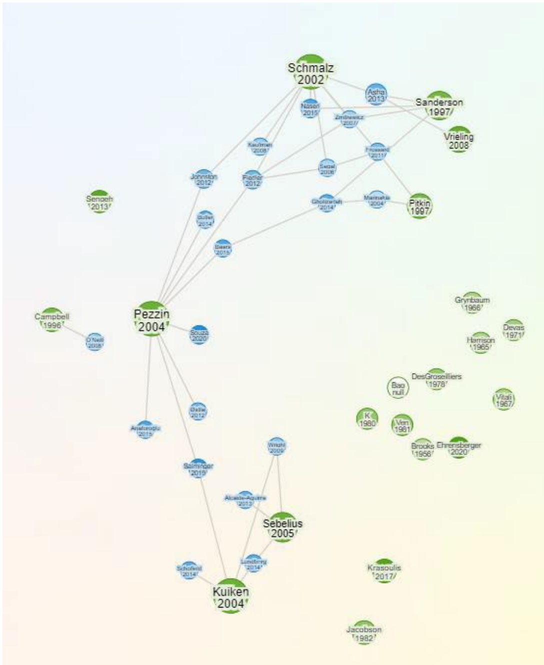


Figure 16: Research Rabbit amputee filter

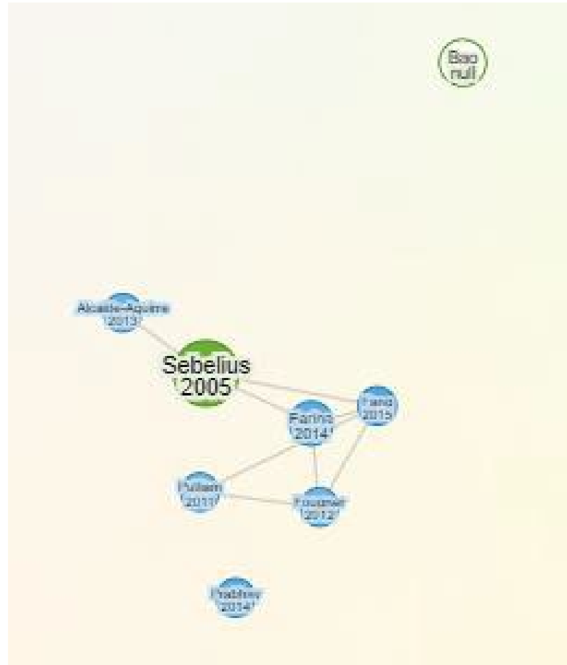


Figure 17: Research Rabbit neural filter

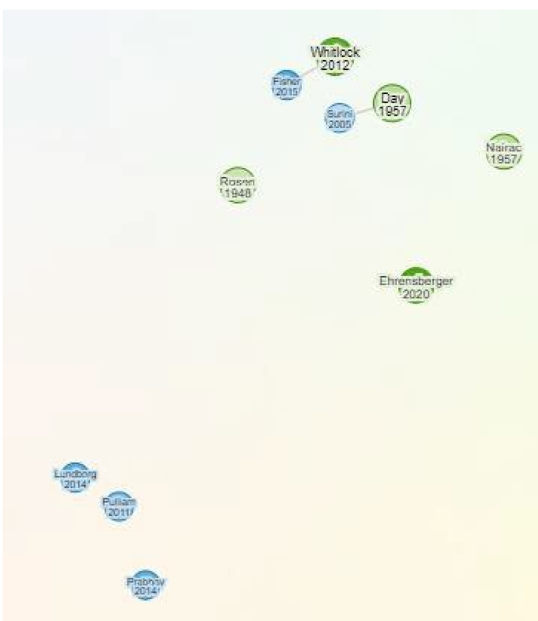


Figure 18: Research Rabbit implant filter

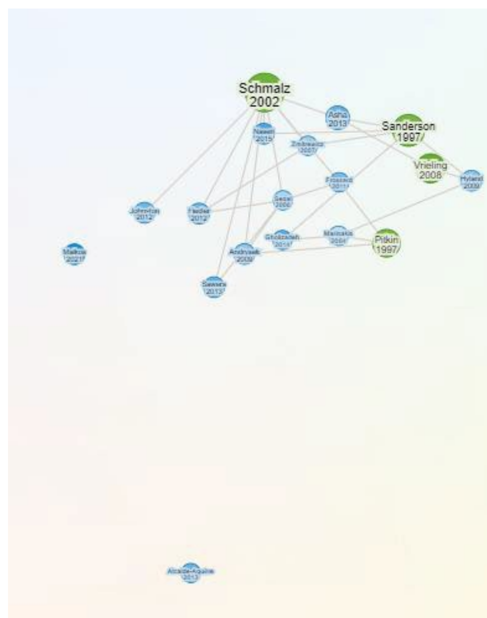


Figure 19: Research Rabbit gait filter

Since this network is directed, there is not much information on determining the robustness of the network. Robustness of a network determines its ability to remain functional if one or multiple nodes are removed. In the context of this scenario, the event of a 'failure' would most likely occur if an article is found to have falsified information. If that were the case, a spreading phenomenon would occur in which any article that cited the falsified one would contain false information as well. This would follow a linear threshold model where a particular influence will affect a node’s neighbors based off the strength of the connection between the node and its neighbor.

8. Results

From this analysis we were able to observe community behavior within the large component of our network. We were able to discern that there was a correlation between date of publication and links between the nodes. Most of the nodes with dates from the late 1990s onward displayed community behavior and made up the large component in the network. Furthermore, we also observed a significant number of nodes with unique author names linked to the hubs of the network, as well as a few nodes with the same author’s name. This is a clear display of homophily within the network, which indicates that there was an increase in collaboration within the network after the 90s. The nodes with dates of publication before this period all had a degree of zero. This indicates that in previous years there was very little interaction between researchers surrounding technological developments surrounding limb prosthetics. The keywords with the most notable appearance in Research Rabbit fell in line with the clustering of the network. This could indicate that current research has been focused on improving the functionality and accessibility of prosthetics for amputees.

Further analysis that would have made our study more interesting would be to create a scatterplot to determine correlation between the year of publication and degree of a node. Another interesting point to touch on in future research would be a frequency table of authors and keywords. However, a much larger dataset would be needed to properly analyze the scope of research. There was some was also some difficulty exporting the dataset from Research Rabbit into Gephi, but by scaling down the network we were able to manually create a node interaction CSV. From that, one member of the team was able to generate the necessary edges CSV utilizing Python. However, as a result, the degrees of the nodes did not show in the Data Laboratory tab in Gephi. Further data preparation would be needed to accomplish this.