Visualization and Implications of Different Networks

Jeffrey Bandurski
DATA-53000
Data Science Program
Lewis University
jeffreybbandurski@lewisu.edu

Shirley Ramirez
DATA-53000
Data Science Program
Lewis University
shirleymramirez@lewisu.edu

Deepkumar Patel
DATA-53000
Data Science Program
Lewis University
deepkumarpatel@lewisu.edu

Abstract - The goal of this project is to visualize three different network datasets, where each network is analyzed and visualized by applying different algorithms and layouts using Gephi.

I. Introduction

This report focuses on highly connected networks of social networks and a network of companies that compete in the internet industry. Even though these networks are unrelated, the principles followed to visualize each network in this report can still be applied to most of the networks. Each dataset contains two separate files, specifically, the nodes and edges, where each node has its own label. Datasets are publicly available on the network repository website under the social networks and labeled network categories.

The purpose of this report is to analyze each network and apply different algorithms to detect communities, authoritative nodes, influential nodes, and broadcasting nodes that can quickly spread information throughout a network. Using all this information, we can find common interests between the entities, design recommendation systems, and even try to systematically remove most central nodes to study how the network might behave in the foreseeable future. Network analysis has many implications, the analysis and different methods shown in the future sections are no way limited to these datasets, therefore they will be generalized and the idea can still be applied to any type of network.

The future sections of this report describe the dataset, brief overview of algorithms and layouts, analysis of each network along with its discussion, and a conclusion. Annexed is a list of sections with a more detailed description included in this report. Section II describes the dataset. Section III describes how statistical algorithms work and their importance in a network, along with layouts used to visualize a network. Section IV describes how each algorithm is applied to each dataset. In addition, the results are visualized and analyzed with a brief discussion. Section V is the conclusion of this report.

II. Data Description

A. Dolphins dataset

The “Dolphins” dataset was acquired from The Network Data Repository website and represents a social network of bottlenose dolphins that lived off the coast of Doubtful Sound, New Zealand in 2003. The network is undirected and consists of 62 nodes, one for each of the dolphins in this community. Each node is labeled with the name of a dolphin and the edges represent frequent associations between them [1]. Figure 1 is an initial visualization of the Dolphin network. No algorithm has yet been applied to it and since this is a relatively small network, no filtering was necessary. The layouts used were Fruchterman Reingold, which gave it a circular shape, followed by the Expansion layout, which reduced crowding and increased the readability of the labels.

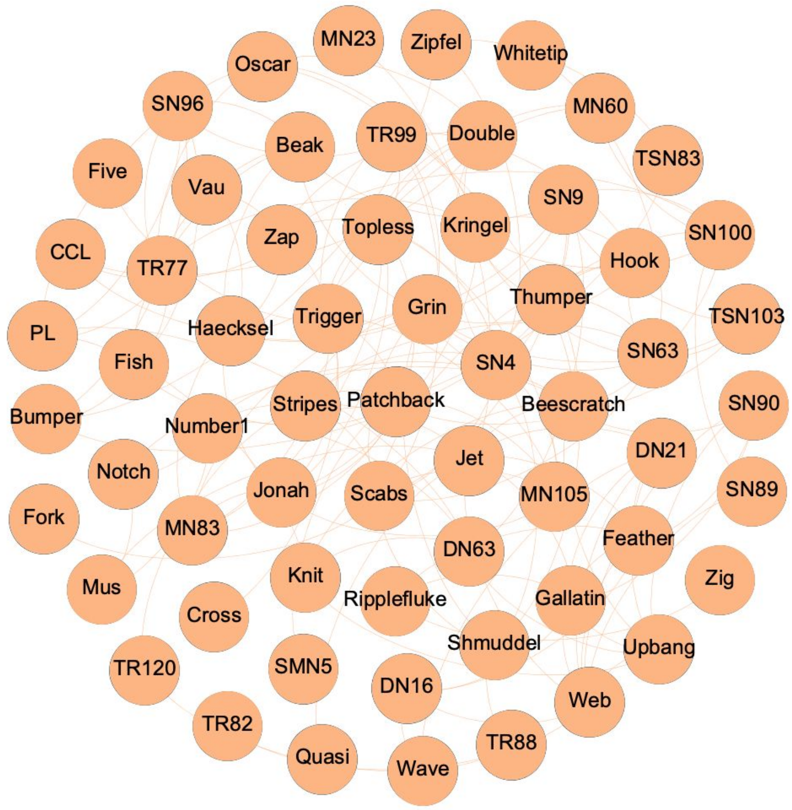

Figure 1. Dolphins dataset prior to the application of any algorithm or layout.

B. Sports dataset

This dataset [2] contains a network of mutually liked Facebook pages from the sports category, collected in November of 2017. Nodes in this dataset represent the pages, and edges represent the mutual likes by users. All individual nodes also include labels, which represent who the page belongs to, essentially it is a Facebook page name. It is also worth mentioning that the pages included in this dataset are verified by Facebook. It means that the pages are authentic and are often searched pages on the website.

Originally this dataset contained 13.9K nodes and 86.8K edges. However, for the purpose of this report a degree range filter was applied to visualize the large network in its simpler form, and at the same time, it helps reveal high-level edge patterns in a network. In an ideal situation, we should not get rid of any nodes, but for demonstrating purposes, the degree range filter helps reveal nodes with a high number of degrees and eliminates nodes with low degrees. In addition, it is easier to comprehend the values on a smaller number of nodes, and it makes it more manageable. Therefore, the final subset of the network contained 35 nodes and 110 edges. It is an unweighted and undirected graph with no self-loops.

In order to fully visualize the subset of the network, we found that Fruchterman Reingold worked the best for this dataset because it expands every node making the length of the edges uniform as shown in figure 2. In addition, the expansion layout was applied to remove any overlaps. We can see that figure 3 shows the Facebook pages (nodes) of athletes, sports companies and teams, along with the edges representing mutual likes between the nodes.

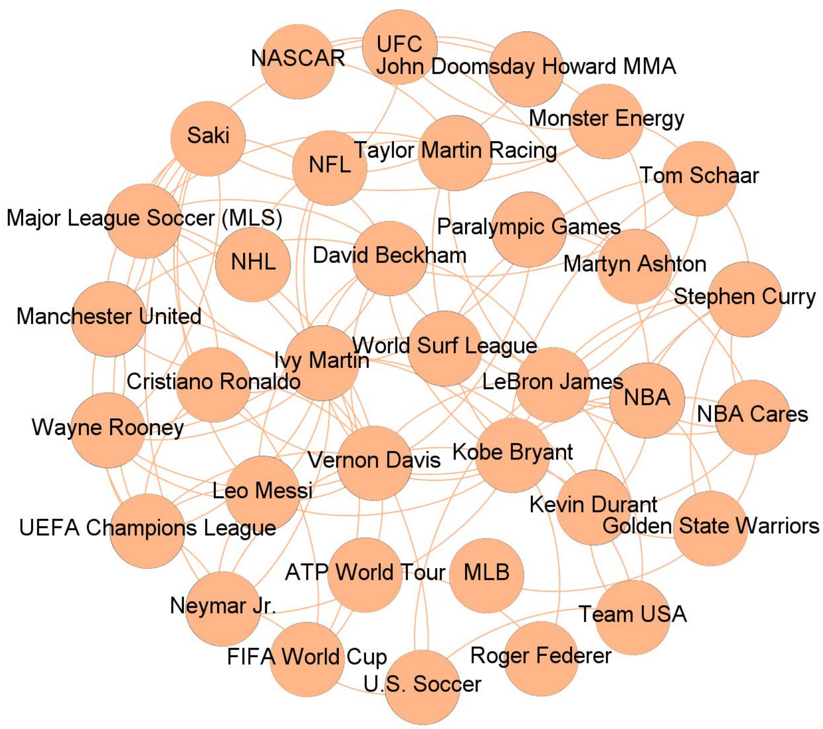

Figure 2. Filtered sports network

C. Internet Industry Partnerships dataset

This dataset [3] contains a network denoting the partnerships between different internet industries. These industries are split into three subtypes, content, commerce, and infrastructure. These subtypes determine the label of each node. After obtaining the folder for the network, a Python script [4] was used on its nodes and edges files to replace the original unlabelled numbers with their corresponding label. From there, the nodes and edges were input into Gephi.

This network originally consisted of 219 nodes and 631 edges. In order to more easily parse the data, a filter was used in Gephi, now only showing nodes and their corresponding edges that had a degree of 5 or higher. This reduced the number of nodes to 84 and the edges to 369. The network is unweighted and undirected. To further improve the readability of the visualization, we then applied a layout. For this case, a Fruchterman Reingold layout was used, giving a circular, uniform look that's easy to parse. Additionally, an expansion layout was used to remove any overlap between nodes and to make edges easier to parse.

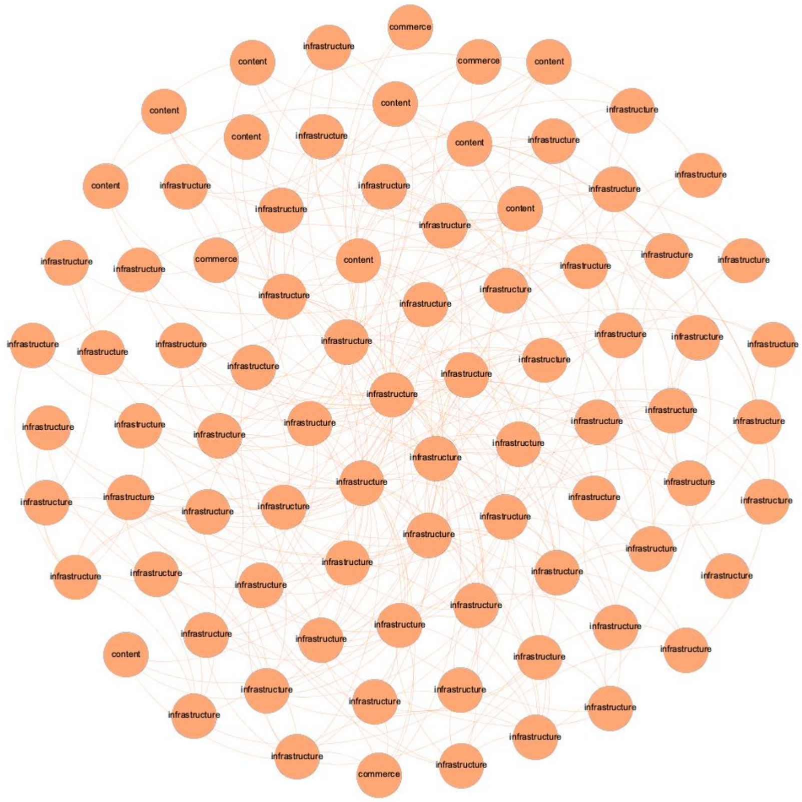

Figure 3. Filtered Industry Network (Included subtypes: Content, Commerce, and Infrastructure)

III. Algorithms and Layouts

Algorithms

A. PageRank

PageRank is an algorithm that was designed in mind with social networks and web pages. Its purpose is to determine the relative importance of the nodes in a network, giving it a corresponding value [5]. This gives you a way to sort networks by some overall measure of relevancy. This importance is based on not just the amount of connections a node has, but also how many connections the nodes' connected to it have. One way to think of it is that the significance of a node is deemed by the amount of nodes that flow into it and the amount of connections those connecting nodes have.

In Gephi, PageRank is largely determined by the number of nodes in the network and the connections (edges) said node has. Specifically, the PageRank of one node will also depend on the PageRank values of all other nodes connected to it and the number of connections those. For a directed network, it depends on the PageRank of the incoming node connections and their number of outgoing connections. In the end, this will give you a list of decimal values that, together, add up to 1 [5].

B. Modularity

One common goal of analyzing networks is to determine if there exist within them smaller subgroups, or clusters. Modularity describes “the density of connection within clusters compared to the density of connections between clusters”[6]. Networks with high modularity have dense connections within the cluster, but sparse connections between clusters [7]. Modularity is also the name of the community detection algorithm implemented in Gephi, based on the Louvain Method.

The Louvain Method is an algorithm used to detect clusters, also called communities, within a larger network [8], which is broken down into two phases [6]. In the first phase, each node is considered a cluster and compared to a neighboring node. If adding them together would result in a new, more optimized cluster, then they are added together, if not then the node moves on and is checked against the remaining nodes. This phase ends when each node has gone through this process and the maximum local modularity has been reached [6]. The second phase involves a similar process as the first phase, except that instead of individual nodes, the new clusters created in the first phase are compared to one another [6].

Once the different communities have been detected by this algorithm in Gephi, they can be visualized following the Gestalt principles of Similarity and Proximity. Utilizing the Partition method, all the nodes belonging to the same community will be attributed the same color [9]. The Layout methods can then be employed to efficiently place the nodes close to the nodes with which they share a link, which often results in nodes of the same community being set near another.

C. Betweenness Centrality Distribution

This algorithm measures the number of times a node appears between every other node, assuming the most amount of information flows through the shortest path between each node [10]. By traversing the graph, it finds the shortest path between each node, and then simply counts the number of times a node falls between every other node. Higher the summation, greater the betweenness centrality. This algorithm is useful for finding the amount of influence a node has over the flow of information in a network. Sometimes nodes with high betweenness centrality are considered the most important nodes because these nodes ensure the connectivity of all the other nodes. In the future section, we will use this algorithm to find important nodes in all the datasets.

D. Closeness Centrality

Using the previous algorithm, we can find nodes with the highest flow of information, but this algorithm helps us find nodes that can efficiently spread the information in a network. It works by calculating the shortest paths between all the other nodes, and then sum of all the paths by the total number of nodes. Finally, the result is inverted to determine the closeness centrality value [11]. This algorithm is very useful when we want to find nodes or individuals who can quickly influence the network, and in the next section we will find such nodes for all three networks.

Layouts

A. Fruchterman Reingold

The Fruchterman-Reingold layout algorithm is considered to be among the most useful of Gephi’s Layout options. It works by simulating the network as a system of mass particles, where the nodes are the particles and the edges are energy-holding springs between them [8]. The algorithm works by moving the nodes around in order to get it to a configuration that minimizes the energy between the nodes [12]. Fruchterman-Reingold is often favored because it allows users to specify the area of the graph [8] as well as the gravity of the layout, to avoid the dispersion of the nodes [12]. Despite its usefulness, however, it is also known for being rather slow, so it is advisable to limit its use to networks of 1,000 nodes or less [12].

B. Expansion

The Expansion Layout is simple and straightforward, but incredibly helpful for readability. As the name implies, if the nodes of a graph are too close together, this layout can be utilized to change the scale of the network graph, in a way that increases the distances between the nodes [13]. This layout is often used when the nodes are too close for the labels to be legible, but it is often advised to limit its use to only once or twice [14].

C. ForceAtlas 2

ForceAtlas 2 is the second version of the ForceAtlas Layout. ForceAtlas was “made to spatialize Small-World / Scale-free networks” [8]and is particularly useful for data exploration because it does not introduce bias when it creates the network graph [12]. ForceAtlas 2 uses the same equations as its predecessor but is optimized to handle larger networks [15]. Also, ForceAtlas 2 has a linear-linear model and allows for the user to set the gravity, which means the attraction and repulsion are proportional to the distance between the nodes [15], and the resulting clusters are in a tighter layout [8]. Additionally, ForceAtlas 2 allows the user to control the area which the graph can cover and it has a “Prevent Overlap” setting, which prevents overlap between the nodes [16].

IV. Network Analysis

A. Analysis of all the datasets using *PageRank*

Here, we applied the PageRank algorithm to each of the three networks shown above. For each of the three, probability was set at 0.85 and epsilon was set at 0.001. Additionally, networks were set as undirected and unweighted.

We wanted to then display the values obtained after running the algorithm. First, the shape of each network is maintained from above. We then set the size of each node to correspond to its PageRank value. The same idea was used to color the nodes, where a range of colors corresponds to size of the PageRank values. The results of this for the networks are shown below, as Figures 4, 5, and 6.

Applying PageRank to the dolphin network shows us the ‘importance’ of each dolphin to their larger social network. Dolphins like Grin (value: 0.0321), Trigger (value: 0.0318), and Jet (value: 0.0313) have the largest amount of relevancy, whereas dolphins like Cross (value: 0.00508), MN23 (value: 0.00542), and Quasi (value: 0.00542) are least relevant. There’s a large number of nodes that have similar rankings as well, that can be broken up into three categories: small (around 0.008), medium (around 0.020) and large (around 0.028). In other words, there are some highly social dolphins, some that are moderately social, and some that seem to be almost ostracized. Another note of interest is that the large value nodes tend not to have primary or secondary connections to other similarly valued nodes. This can be observed in the below section, denoting modularity classes. The dolphins with larger PageRanks all seem to be in different classes from one another.

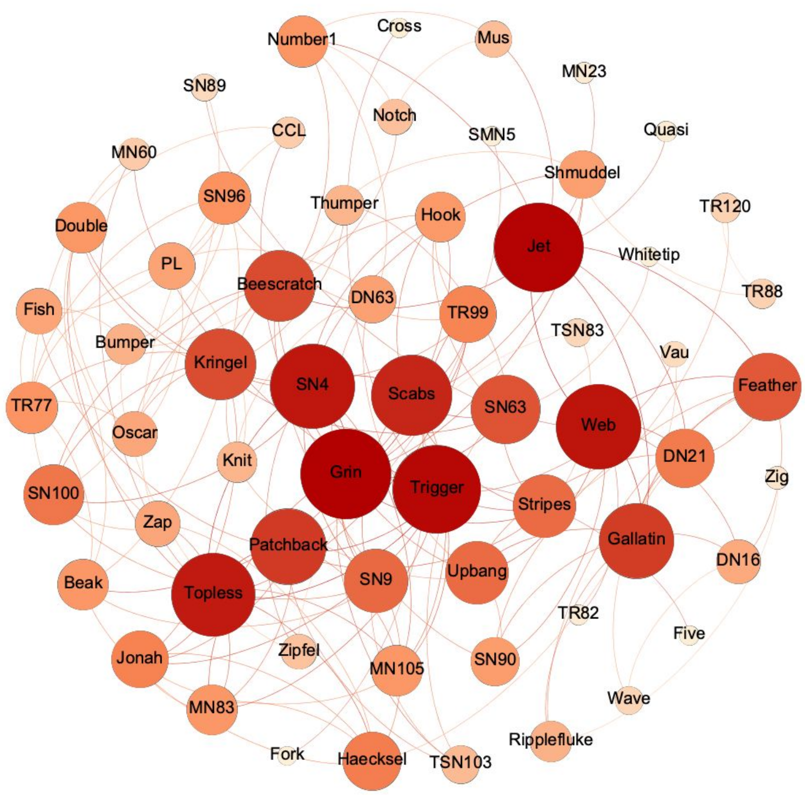

Figure 4. Dolphins Network with PageRank (Color and size indicate PageRank value. Lighter reds indicate smaller values. Darker reds indicate larger values)

Looking at the network below, the most ‘important’ terms from the sports network seem to Kobe Bryant (value: 0.0564), Vernon Davis (value: 0.0551), and Ivy Martin (value: 0.0551). These are all prominent, well-known athletes. On the other end, the less relevant terms tend to be athletic organizations (MLB, NHL, Nascar). Of course, despite calling them less relevant, this is after the network has filtered out the smaller degree nodes, so even the smaller values are still going to be fairly relevant. Considering this was captured in November 2017, this can give a purview into hot button topics in that time period. For example, a lot of terms relevant to basketball appear in this listing (Kobe Bryant, LeBron James, NBA). Additionally, terms relevant to soccer are prominent as well (FIFA World Cup, U.S. Soccer, MLS).

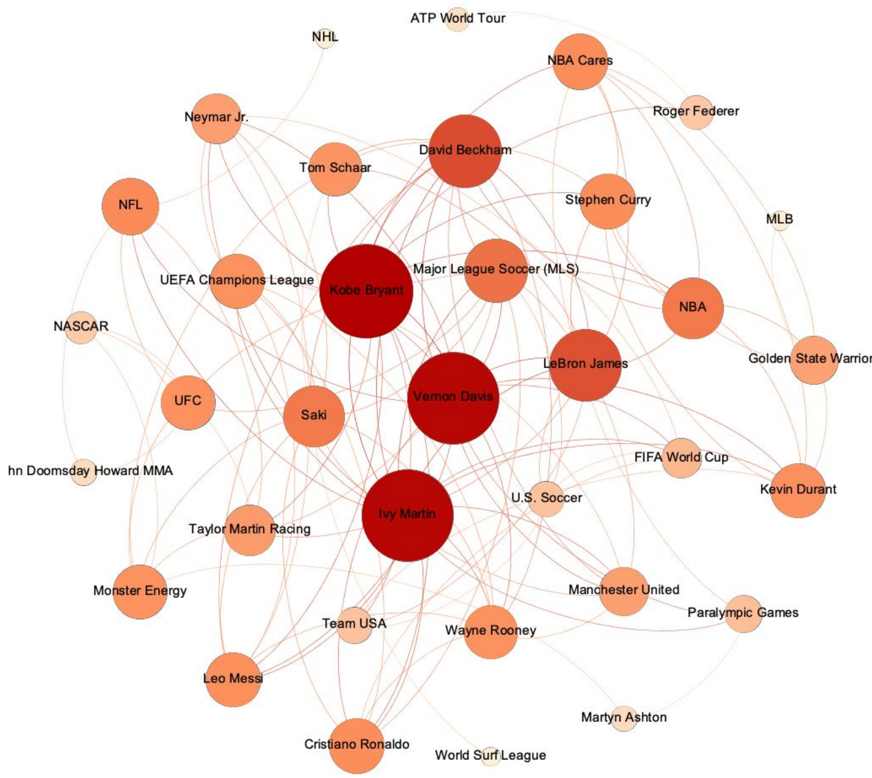

Figure 5. Sports Network with PageRank (Color and size indicate PageRank value. Lighter reds indicate smaller values. Darker reds indicate larger values)

In the filtered network below, we can see that the most of the nodes are already infrastructure based, with a handful of content and commerce industries. By the same line, infrastructure industries are generally the nodes that have the largest amount of relevancy as well. Commerce nodes, what few there are, tend to be among the least relevant. Some of the content nodes are strikingly relevant however. What this largely shows is that an infrastructure-labelled group is much more likely to belong to a larger network of partnerships with others on the internet, whereas commerce will likely not. And that there will be a handful of content-oriented groups that will be somewhere in between the other two in terms of partnerships.

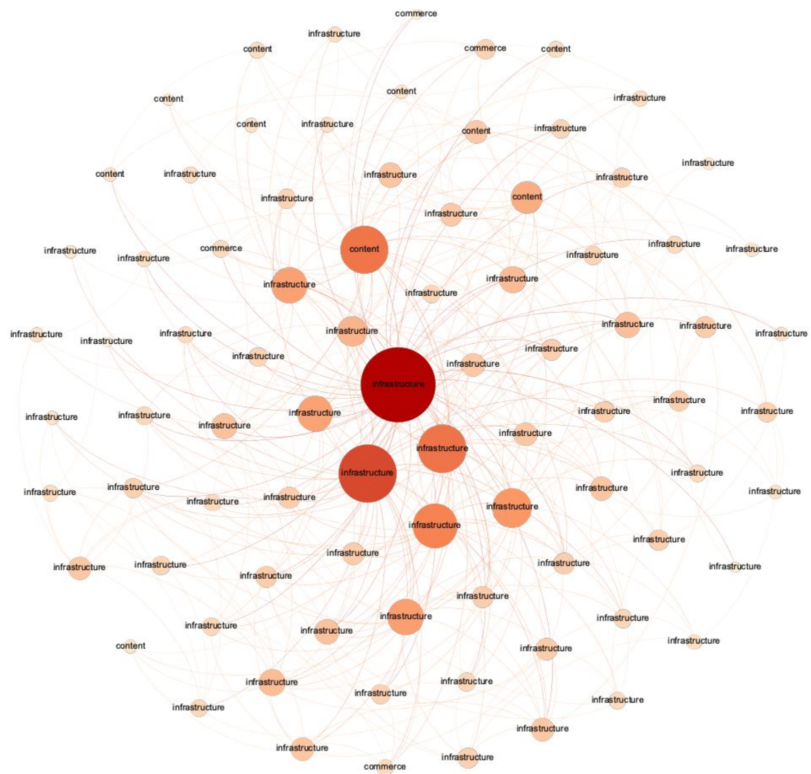

Figure 6. Internet Industry Network with PageRank (Color and size indicate PageRank value. Lighter reds indicate smaller values. Darker reds indicate larger values)

B. Analysis of all the datasets using *Modularity*

As previously described, Modularity is the property of a network system that describes the degree to which a network can be broken down into smaller groups, often referred to as communities. A higher modularity indicates that the nodes have a stronger relationship with the other nodes in their community than they do with nodes belonging to other communities. Conversely, low modularity suggests that while there may still be subgroups among the network, these clusters are not as isolated from each other as those of more highly modular networks. Modular analysis and visualization of the Dolphins, Sports, and Internet Industry Partnerships networks using the visualization tool Gephi, will help to illustrate this property. For each network, the Modular algorithm was applied to identify the number of communities existing in each network and to calculate the modularity score for each network. Each node was then assigned a new color corresponding to the cluster they belonged to, and then the ForceAtlas 2 and Expansion layouts were applied.

The Dolphin network's modularity score was calculated to be 0.517 and 5 distinct communities were identified (Figure 7). There are some dolphins that tend to have frequent interactions with members outside of their main group. Oscar, for example, has interactions with members of 4 of the 5 groups and TSN83 also seems to interact as often with members of its own (orange) cluster as with members of another (light green) cluster. In fact, TSN83 is almost completely surrounded by members of the green cluster in this visualization. For the most part however, intercommunity interactions are limited to a few dolphins in each group, while the rest limit their interactions to members of their own community. In the purple group, for instance, out of the 22 members, only 4 (PL, Beescratch, SN89, and DN63) show interactions with other clusters.

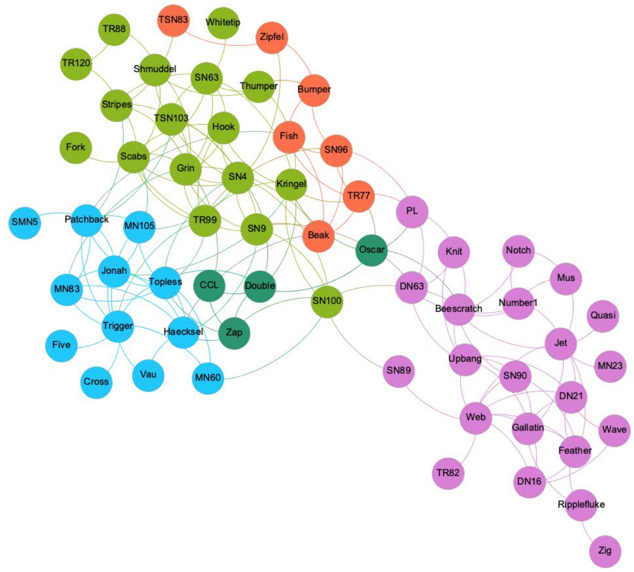

Figure 7. Modularity of Dolphins Network dataset (Groups = 5, Modularity = 0.517)

When the modularity algorithm was applied to the Sports network, 3 distinct communities were identified, and the modularity score was calculated to be 0.367 (Figure 8). The groups are not very well defined, with nodes like U.S. Soccer, for example, unexpectedly falling into the green group, which is primarily basketball themed. In this network, there seems to be many connections between the nodes of different clusters, with many of the green nodes connecting to blue nodes, and most of the red nodes connecting to at least one other cluster.

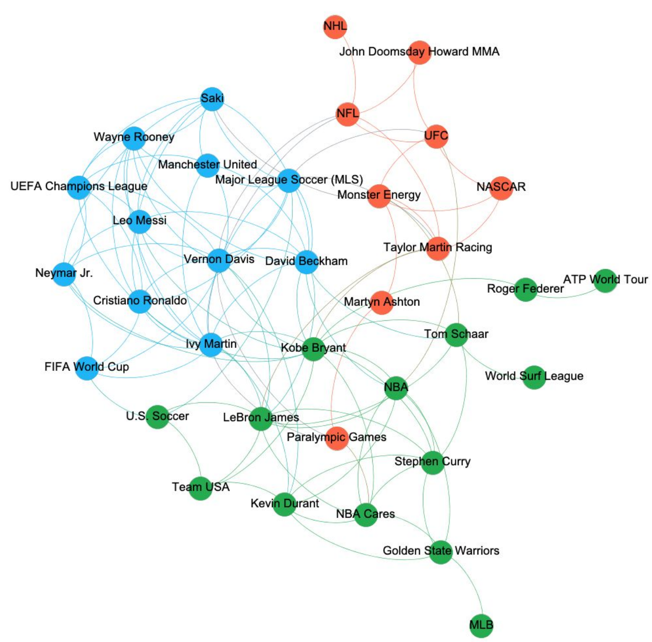

Figure 8. Modularity of Sports Network dataset (Groups = 3, Modularity = 0.367)

Finally, 7 communities were found within the Internet Industry Partnership network, and a modularity score of 0.271 was calculated (Figure 9). The low modularity score represents what is clear from Figure 7: nodes have dense connection to nodes in other communities. Nodes of the same color are in the general vicinity, but there are many which appear completely surrounded by nodes of other groups.

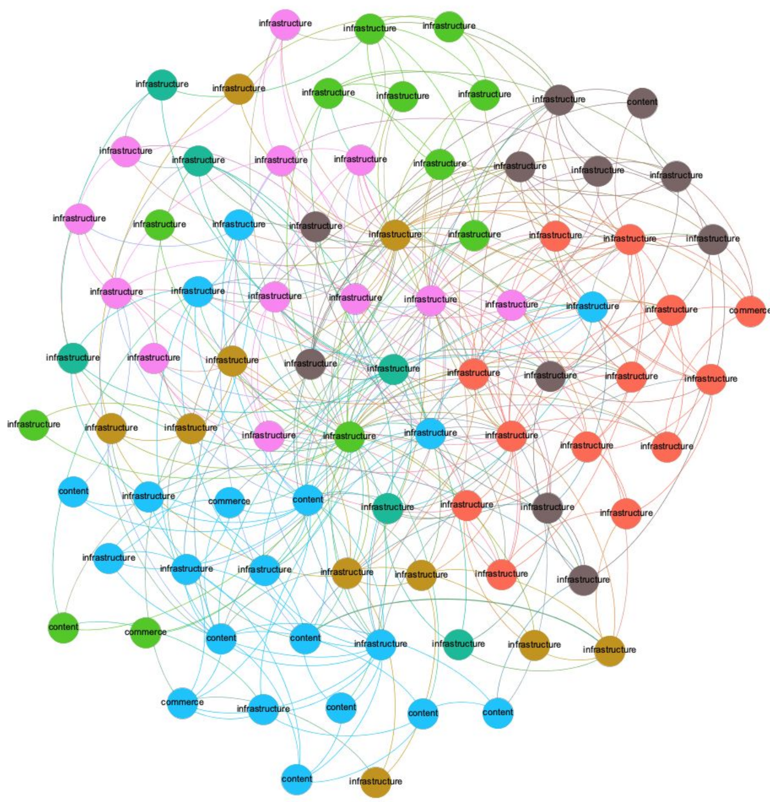

Figure 9. Modularity Internet Industry Partnerships Network Dataset (Groups = 7, Modularity = 0.271)

The modularity scores in these three networks are not unsurprising. Dolphins are well-known for being extremely social creatures, who live in fission-fusion societies [17]. They form close bonds with the members of their communities, so a low modularity score would be unlikely. However, coastal bottlenose dolphins, like the ones in this dataset are also known to have fluid associations, such as alliances that they may form to acquire food and protection, so some intercommunity interactions would be expected [18].

The Sports network shows three groups, and lower modularity than the Dolphin network, which is again, to be anticipated. Many people who would like a sports-related Facebook page are not just fans of a particular team, but of sports in general. So, if they enjoy watching one sport in one season, they might also enjoy another sport in a different season, and thus like Facebook pages that fall in different groups.

The Internet Industry Partnership's low modularity is not only expected but probably hoped for. The nodes represent the components of an industry, essentially, they exist to support one another, so in order for them to be successful they require strong links. The more nodes that link, the greater the coverage.

Modularity describes a system property, and it is not inherently good or bad. Sometimes, the degree of modularity is intentional, as with the internet, for example, where more connections are generally better, or it is just a reflection of other properties, like a dolphin's familial bonds, or a team's geographic location. There are some things to keep in mind however, that make the study of modularity important. High modularity can lead to some benefits, like when it helps to contain a disease outbreak [19] but it is difficult to obtain [20]. Low modularity, on the other hand, can increase the transfer of resources and ideas, but can make the whole network vulnerable. If one node goes down, every node it is connected to is at risk of also going down with it. When designing a system, these insights should be leveraged to create a functioning and robust network; data visualizations can help in this process.

C. Analysis of all the datasets using *Betweenness Centrality Distribution*

In the previous section we learned that "Betweenness Centrality" helps find the most important or influential nodes in a network. In this section, we will analyze each dataset and determine which nodes are important, and perhaps reveal meaningful explanations as to why they are important.

The first network to analyze is the "Dolphins" dataset, in which the nodes represent the name of bottlenose dolphins and edges represent frequent associations between them. After the betweenness centrality is applied, we are left with figure 10. Even in this small network of dolphins we can see there are few mammals that are important in holding the network together. In this case SN100, Beescratch, SN9 and SN4 are among the most influential dolphins in this community.

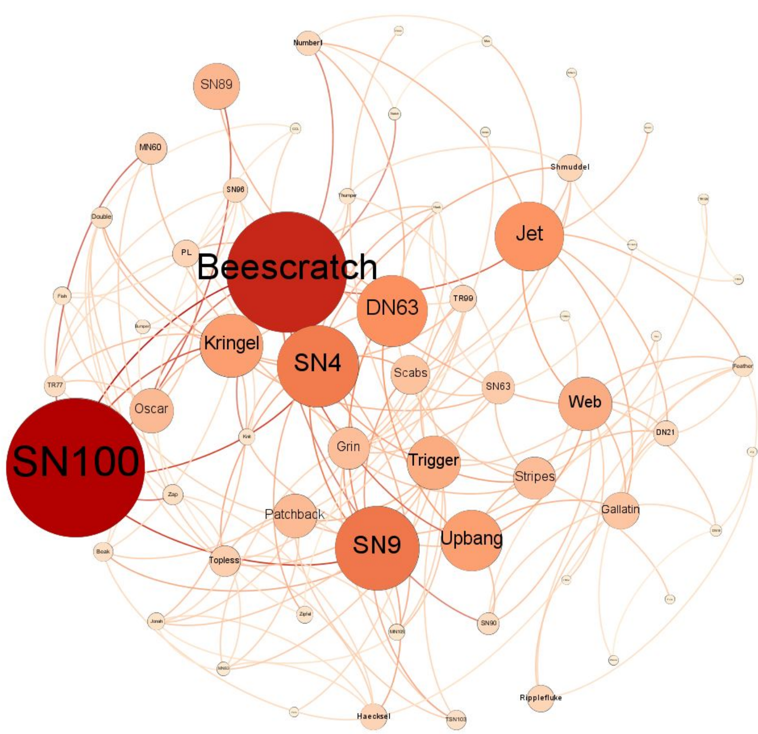

Figure 10. Dolphins network with betweenness centrality (Color and size indicate betweenness value. Lighter reds indicate smaller values. Darker reds indicate larger values)

To better understand why this statistical metric is important, let's assume there's a major outbreak of morbillivirus (virus that infects marine mammals [21]) in this community of dolphins. In order to slow the spread of the virus, dolphins with high betweenness scores would get vaccinated first because they can not only prevent the spread of such an outbreak but without them the community would not hold together.

The second network to analyze is the "Sports" dataset, and in this network the nodes represent the Facebook pages from the sports category, and edges represent the mutual likes by users. When we apply this algorithm, it shows Kobe Bryant, Vernon Davis and Ivy Martin are among the most influential athletes in this network as shown in figure 11. If we carefully observe the size for Vernon Davis and Ivy Martin, it is the same, because both of these nodes have the same betweenness value, and they also share the same number of degrees or connections. This network also reveals that basketball and football are among the most popular sports based on the nodes. It is also interesting to see how the best soccer players have low betweenness scores, even though it's the most popular sport in the works according to this article [22]. One reason we might see such a pattern is because the Facebook likes were collected based on the US population where basketball and football are widely famous.

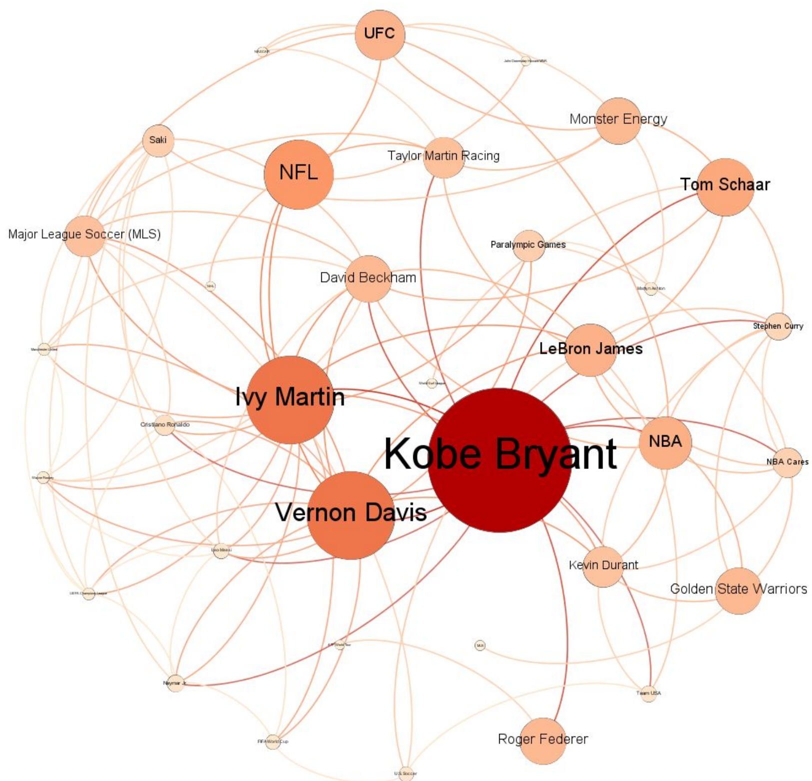

Figure 11. Sports network with betweenness centrality (Color and size indicate betweenness value. Lighter reds indicate smaller values. Darker reds indicate larger values)

Final network to analyze is the "Internet Industry Partnerships" dataset, and in this network the nodes represent a company that competes in the internet industry and edges represent partnerships among the companies. After applying this algorithm, figure 12 shows that infrastructure companies happen to be the most influential in the entire internet industry. Since most of the information passes through nodes with high betweenness, it is safe to say that investing, or partnering with infrastructure companies have a higher advantage in comparison to commerce businesses.

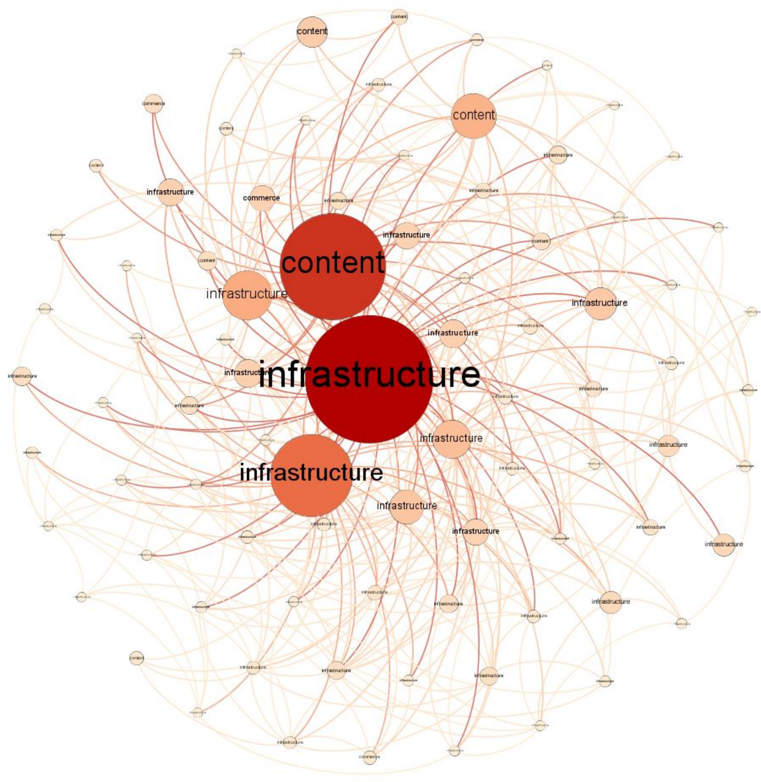

Figure 12. Internet Industry Partnership network with betweenness centrality (Color and size indicate betweenness value. Lighter reds indicate smaller values. Darker reds indicate larger values)

D. Analysis of all the datasets using *Closeness Centrality*

Similar to betweenness, closeness centrality makes use of the shortest paths between nodes to find nodes that can efficiently spread information throughout a network. Oftentimes, in an organization these nodes hold vital information and resources within a network.

The generated closeness centrality for "Dolphins" dataset shows many nodes that are close to each other as shown in figure 13. In a way, it is good, because it creates a tight sub-neighborhood or communities of dolphins. However, let's take the previous example and expand into it a bit further. Assuming there's an outbreak of morbillivirus in this network, then the chance of spreading the virus is so quick that it can infect the entire network because of high closeness centrality. If we compare the graph generated for modularity of dolphins and closeness centrality, we can conclude that purple communities of dolphins are least likely to spread the virus due to their low centrality. On the other hand, the light green community is most likely to spread the disease, due to high centrality value.

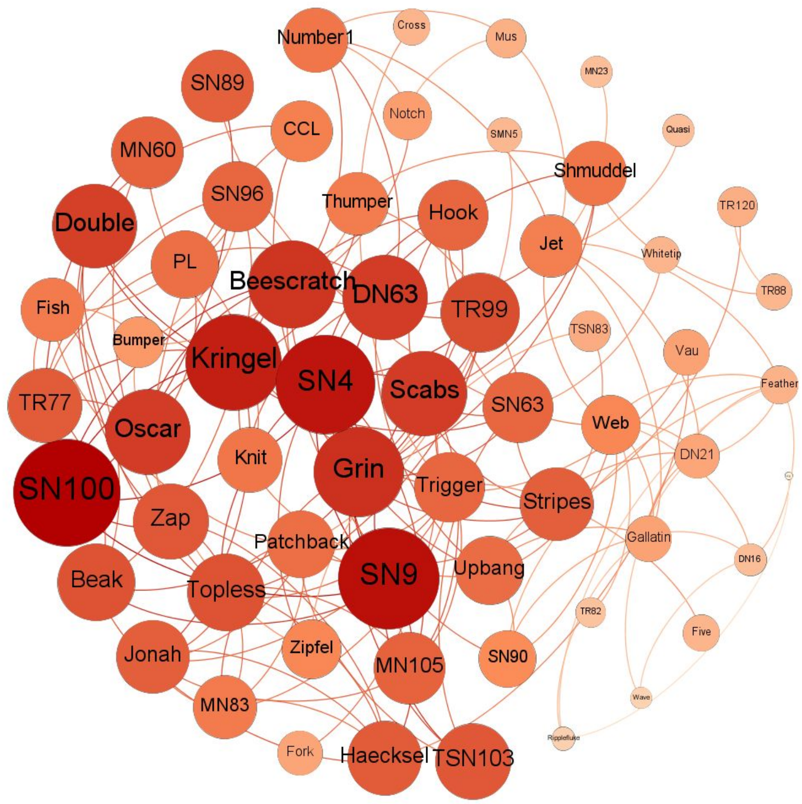

Figure 13. Dolphins network with closeness centrality (Color and size indicate closeness value. Lighter reds indicate smaller values. Darker reds indicate larger values)

From a brief overview in an earlier section, we learned that betweenness centrality and closeness centrality both use the shortest path to derive its values. That is why it is no surprise that high betweenness nodes still hold the position of high closeness centrality. However, with closeness centrality we see new emerging nodes that were once invincible. For instance, in figure 14 we can see that Kobe Bryant, Vernon Davis and Ivy Martin are among the top nodes with high closeness centrality values, and we also see professional soccer athletes along with a few teams and companies. There are many mutual likes/edges between the athletes in comparison to teams or companies, which is obvious because the majority of people follow the athletes. Having said that, there are major connections/links between the athletes and leagues/companies they are associated with almost similar closeness values.

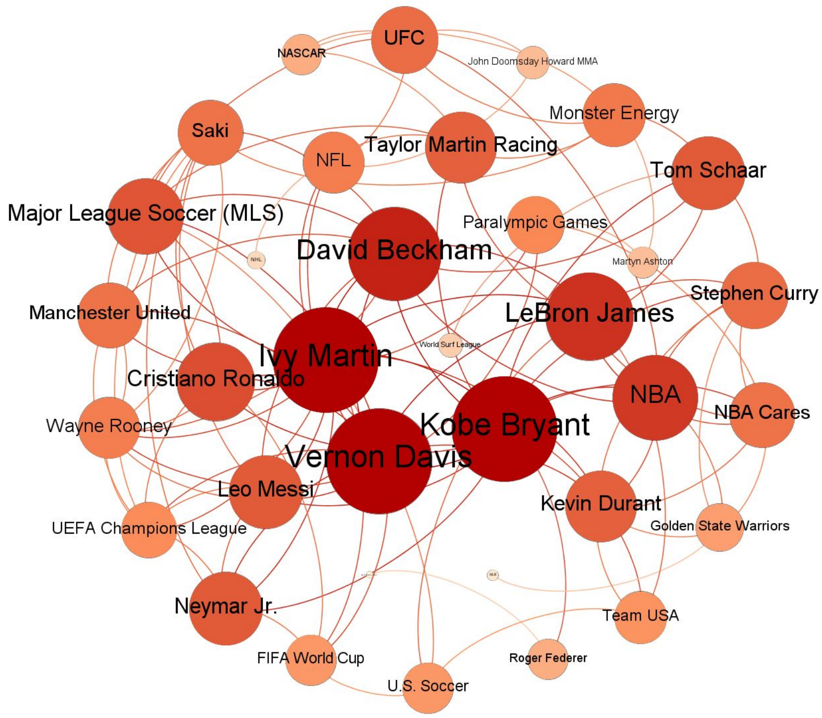

Figure 14. Sports network with closeness centrality (Color and size indicate closeness value. Lighter reds indicate smaller values. Darker reds indicate larger values)

After applying the closeness centrality to the "Internet Industry Partnerships" dataset, it revealed that many companies around the edges have very few partnerships in comparison to the companies in the middle of the graph as shown in figure 15. The companies in the middle have many partnerships with a highest degree of 39. Oftentime, big companies have many partnerships, and as mentioned earlier, these are the companies that hold most of the resources at their disposal. Naturally, the company diversifies the knowledge due to the positions they are in and in most cases helps drive the economy.

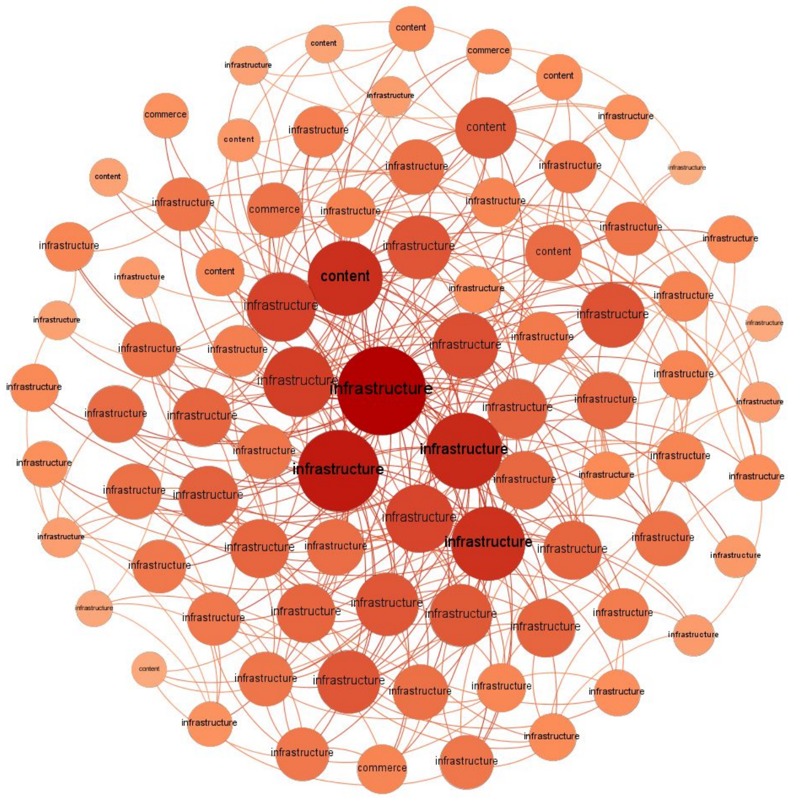

Figure 15. Internet Industry Partnership network with closeness centrality (Color and size indicate closeness value. Lighter reds indicate smaller values. Darker reds indicate larger values)

V. Conclusions

This article shows how different algorithms can be used to discern factors, groups, and other relations on largely different types of networks. PageRank can be used to determine the significance of specific members or terms in a social network or the significance of different types of industries. Modularity shows that we can group terms or members into classes composed of similar terms or individuals. Betweenness Centrality Distribution and Closeness Centrality show which members or terms are the most important for the speed and the flow of information throughout the networks. These different factors are repeatedly useful in discerning some relevant aspect of every network.

References

[1] R. A. Rossi and N. K. Ahmed, “Dolphins,” networkrepository.com, 2015. [Online]. Available: http://networkrepository.com/soc-dolphins.php. [Accessed: Dec. 12, 2020].

[2] R. A. Rossi and N. K. Ahmed, “Facebook Pages Sport,” networkrepository.com, 2015. [Online]. Available: http://networkrepository.com/fb-pages-sport.php. [Accessed: Dec. 12, 2020].

[3] R. A. Rossi and N. K. Ahmed, “Internet Industry Partnerships,” networkrepository.com, 2015. [Online]. Available: http://networkrepository.com/internet-industry-partnerships.php. [Accessed: Dec. 15, 2020].

[4] Python Software Foundation. Python Language Reference, version 2.7. Available: https://www.python.org.

[5] Shaw, A., 2019. *How A Page Rank Is Calculated In Gephi*. [Online] Strategic Planet. Available: https://www.strategic-planet.com/2019/12/how-a-page-rank-is-calculated-in-gephi/. [Accessed: Dec. 20, 2020].

[6] N. Jenker, "Community Detection with Louvain and Infomap," March, 2020. [Online]. Available: https://www.statworx.com/en/blog/community-detection-with-louvain-and-infomap/. [Accessed Dec. 18, 2020].

[7] Parklize, "Gephi – Clustering layout by modularity," Dec, 2014. [Online]. Available: https://parklize.blogspot.com/2014/12/gephi-clustering-layout-by-modularity.html. [Accessed Dec. 17, 2020].

[8] Gephi, "Gephi Tutorials: Layout," gephi.org, Jun. 13, 2011. [Online]. Available: https://gephi.org/tutorials/gephi-tutorial-layouts.pdf . [Accessed Dec. 8, 2020].

[9] M. Grandjean, "GEPHI – Introduction to Network Analysis and Visualization," Oct, 2015. [Online]. Available: http://www.martingrandjean.ch/gephi-introduction/. [Accessed Dec. 8, 2020].

[10] "Betweenness Centrality," geeksforgeeks.org, 2018. [Online]. Available: https://www.geeksforgeeks.org/betweenness-centrality-centrality-measure/. [Accessed Dec. 18, 2020].

[11] "Closeness Centrality," geeksforgeeks.org, 2018. [Online]. Available: https://www.geeksforgeeks.org/closeness-centrality-centrality-measure/. [Accessed Dec. 18, 2020].

[12] B. Mosse, "A quick tutorial on Gephi layouts — using a Q&A StackOverflow dataset," July, 2020. [Online]. Available: https://towardsdatascience.com/a-quick-tutorial-on-gephi-layouts-daa87fec5a20. [Accessed Dec. 18, 2020].

[13] V. Miz, "Gephi tutorial.Layouts," January, 2020. [Online]. Available: https://blog.miz.space/tutorial/2020/01/05/gephi-tutorial-layouts-force-atlas-circle-pack-radial-axis/. [Accessed Dec. 19, 2020].

[14] P. Oldham, "Network Visualisation with Gephi," May, 2015. [Online]. Available: https://www.pauloldham.net/gephi_patent_network/. [Accessed Dec. 19, 2020].

[15] M. Jacomy, "ForceAtlas2, the new version of our home-brew Layout," June, 2011. [Online]. Available: https://gephi.wordpress.com/2011/06/06/forceatlas2-the-new-version-of-our-home-brew-layout/. [Accessed Dec. 19, 2020].

[16] Ethnographic Machines, "Visual network analysis with Gephi," February, 2019. [Online]. Available: https://medium.com/@EthnographicMachines/visual-network-analysis-with-gephi-d6241127a336. [Accessed Dec. 19, 2020].

[17] B. Augliere, "Life in a Dolphin Pod: Male Social Structure," *The Wild Dolphin Project*, 2014. [Online]. Available: http://www.wilddolphinproject.org/life-in-a-dolphin-pod-male-social-structure/. [Accessed: Dec. 19, 2020].

[18] SeaWorld Parks and Entertainment, "Social Organization," *SeaWorld Parks and Entertainment*, 2020. [Online]. Available: https://seaworld.org/animals/all-about/bottlenose-dolphin/behavior/. [Accessed: Dec. 19, 2020].

[19] A. Kharrazi, "Resilience," in *Encyclopedia of Ecology (Second Edition).* Elsevier, [online document], 2019. Available: Science Direct, https://www.sciencedirect.com/science/article/pii/B9780124095489107511. [Accessed: Dec. 19, 2020].

[20] I. Sutton, "Chapter 6 – Operating Procedures," in *Process Risk and Reliability Management (Second Edition)*. Gulf Professional Publishing, [online document], 2015. Available: Science Direct, https://www.sciencedirect.com/science/article/pii/B9780128016534000060. [Accessed: Dec 19, 2020].

[21] Access Science, "Morbillivirus Infection in Bottlenose Dolphins", 2014. [Online]. Available: https://www.accessscience.com/content/morbillivirus-infection-in-bottlenose-dolphins/BR0904141. [Accessed: Dec. 19, 2020].

[22] J. Shvili, "The Most Popular Sports in the World" 2020. [Online]. Available: https://www.worldatlas .com/articles/what-are-the-most-popular-sports-in-the-world.html. [Accessed: Dec. 19, 2020].