IDS 564 Final Project Report

Lina Quiceno and Eduardo Herrera

Background and Motivation

In this class, we studied how to analyze and understand networks to provide practical information to drive business decisions. We have studied networks and how they spread information. For our project, we want to look at the spread of musical ideas within a genre of music.

Jazz as a music genre is considered one of the most influential and prominent contributions by the United States. Jazz originated in New Orleans in the late 19th century. One of the most important aspects of Jazz music is improvisation. Our goal is to understand the interconnectedness of Jazz bands within our network and visualize the collaboration between Jazz musicians that led to the spread of Jazz as a popular genre of music as well as an influence on other genres of music that would follow it.

Network Summary Statistics

Our network consists of 198 vertices that represent the jazz musicians and 5,484 Edges that represent their relations. The Jazz musician network is conformed by 1 giant component and that means the collaboration between musicians is very tightly connected. Below in Figure 1, we see the initial plot of our Jazz network using R. For this plot we didn't apply any algorithms and just wanted to visualize the network before continuing with our analysis.

Table 1. Network Description

Vertices	198
Edges	5,585
Giant Component	1

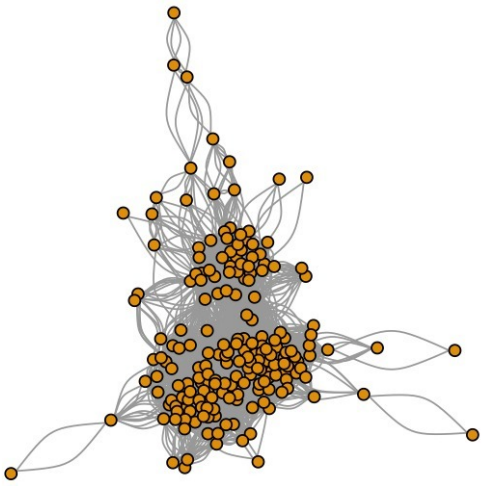

Figure 1: Initial plot of the Jazz network.

Table 2. Network Summary Statistics						
Diameter	**Vertex Connectivity**	**Edge Connectivity**	**Transitivity**	**average path length**	**mean distance**	**avg. degree**
6	1	2	0.52	2.2	2.2	27.69

Based on Table 2 above, we see that the average degree for our network is 27.69. This means for any given node in the network, each individual Jazz band collaborates with about 27 other bands on average. The average path length was 2.2. This means that on average the length between any 2 given Jazz musicians would be 2.2. Already, we can start to see the large amount of collaboration within our network.

Connectivity

The basic statistics we ran in the network show us a well-connected network as there is a short path between every pair of musicians on average. In our network, the mean distance between musicians is 2.2. This means on average any given musician is only 2.2 musicians away from each other. We can also see this by analyzing the diameter of the graph meaning the value of the longest distance between musicians which is 6, or in the transitivity which measures the probability that the adjacent vertices of the musicians are connected, which in this case is 0.52. This is evidence that the Jazz network may be a “small-world” network. Since most nodes can be reached from every other node by a short path and the clustering coefficient, or transitivity is not small.

Degree Distribution

We plot the degree distribution of the Jazz network to see the plot frequency distribution. As we had a weighted network, we preferred to use the vertex strength, as we obtained the plot by adding the weight of the edges to a given vertex. As a result, we got the right skew distribution.

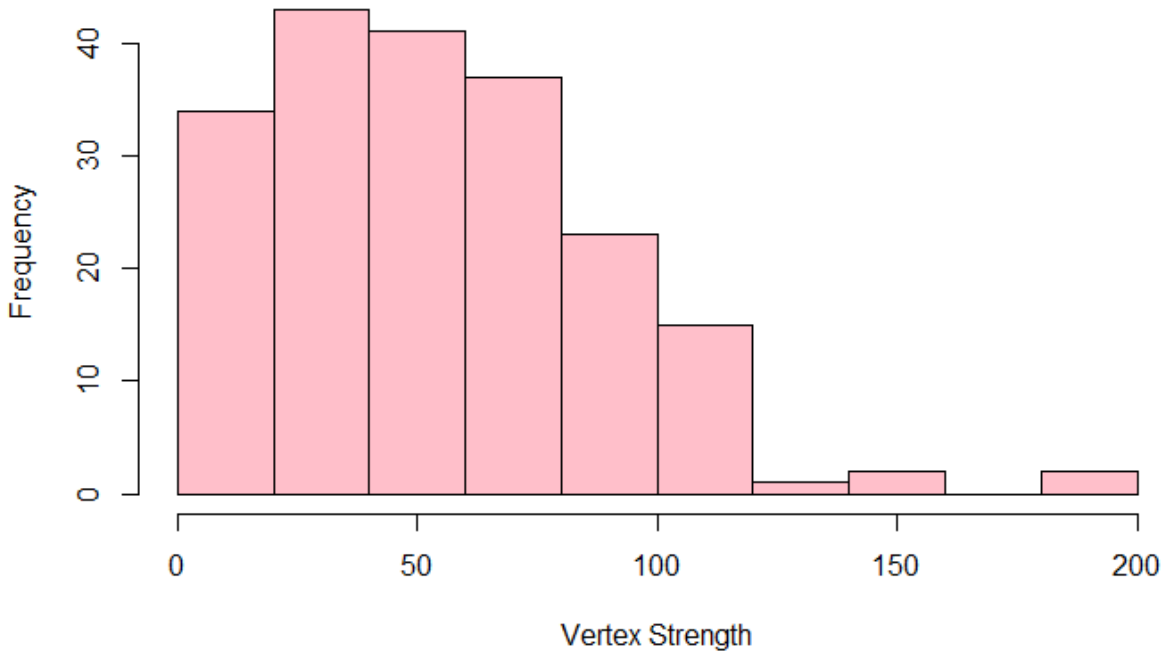

Figure 2: Plot for Degree Frequency using Vertex Strength

As our vertex distribution has a bit of decay shape, we got curious about how it looks with a log-log transformation. We were expecting to get something similar to a linear distribution, but we did not get that result. The log distribution only shows us a tendency to get a log degree greater than 20 with a few lower degrees.

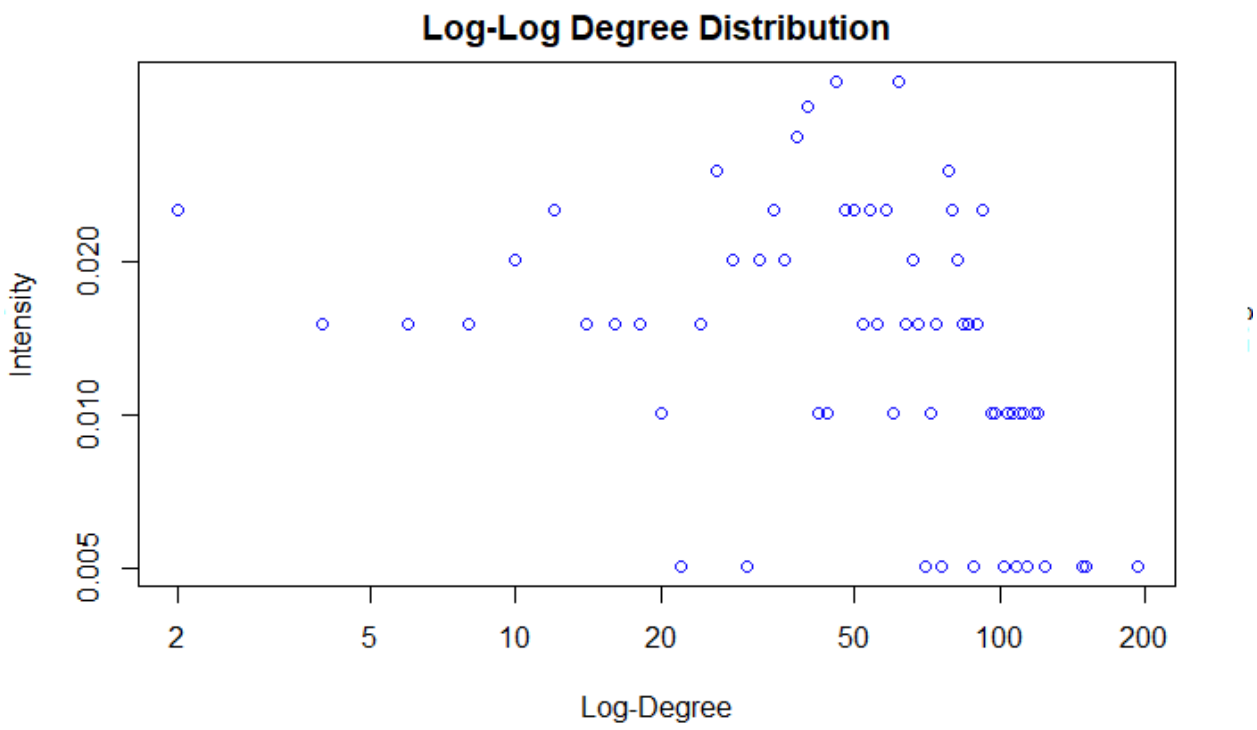

Figure 3: Plot for Log-Log Degree Distribution

Measures of Centrality

We want to look at the distribution of different measures of centralities. Looking at Figure 4 below, we found, in general, betweenness, degree, and strength centralities were right-skewed distributions, and betweenness centrality was the most prominent one. None of the centrality measures are normally distributed.

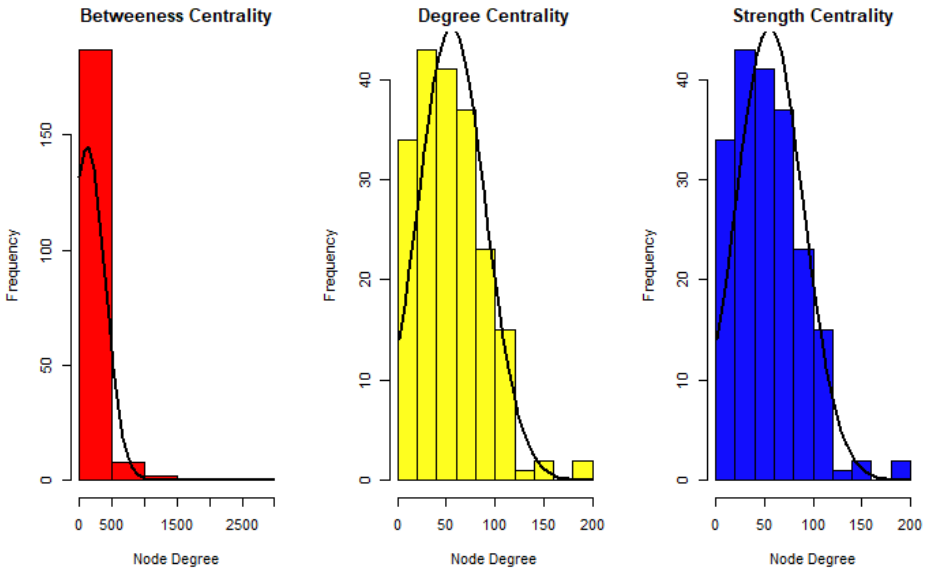

Figure 4: Plots for Betweenness, Degree, and Strength Centrality

Centrality plots

We plotted the network graph in Figure 5 using different centrality measures to get a sense of how the musicians were distributed within the network. We observed a highly interconnected graph but also a big difference between the different centrality measure distributions of musicians. In Figure 5, the blue and yellow vertices are assumed to be the most important musicians in the network using our class text book as a guide (Kolaczyk et al, 2014, 49). As a starting point, we plotted the network with a degree of centrality, which shows us how connected each one of the musicians is. In the closeness centrality plot, we can spot a musician in the middle in yellow, this musician is a central point because it's close to other musicians, this allows us to spot the musicians who can influence the fastest to the rest of the network. As betweenness centrality is based on short paths and summarizes if a musician is "between" other musicians, the plot allows us to find the musicians that are bridges and connect the network, this could also help to analyze how the communication between them flows through the musician's highlight as important in the graph, this shows us who held power in the network (Kolaczyk et al, 2014, 48).

Finally, Eigenvector centrality shows us how central a musician is according to the centrality of its neighbors (Kolaczyk et al, 2014, 48). This shows us the musicians with more prestige, the ones who have more connections and could influence the whole network.

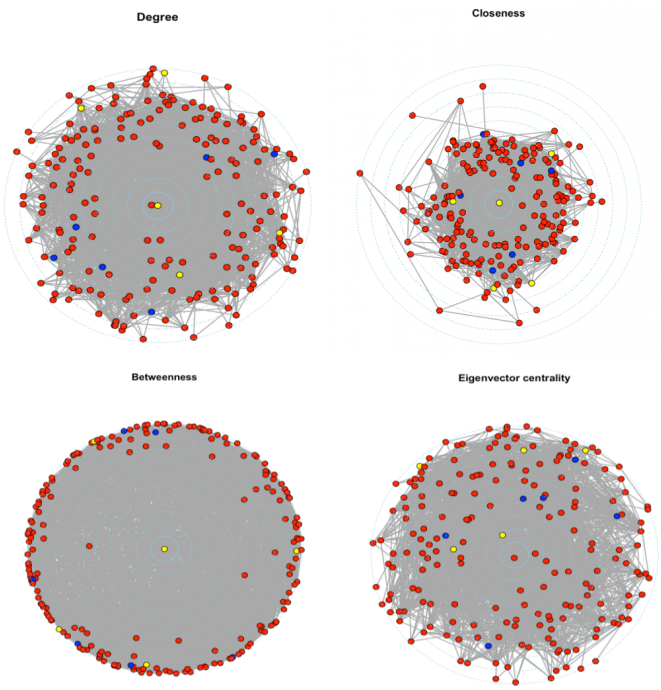

Figure 5: Plots of Centrality Measures for the Jazz network

Community Analysis

First, we created a simplified graph from the overall Jazz network. Then we used the Fast Greedy Algorithm to find 4 distinct communities using R. Taking a look at table 3 below, we found four communities with sizes of 62, 66, 3, and 67 musicians respectively. Figure 6 below visualizes the 4 different communities in the Jazz network using the Fast Greedy Algorithm in R.

Table 3. Community Detection

Community	**1**	**2**	**3**	**4**
# of Nodes	62	66	3	67

Network Visualizations

We used a combination of both R and Gephi to visualize the Jazz network.

Figure 6 below shows the network plot along with community detection using the Fast Greedy Algorithm. We also created a dendrogram that corresponds to the plot in Figure 6. It shows the nodes being split among the 4 communities within the Jazz network.

We also plotted the Jazz network using Gephi. Using Gephi, we applied the Fruchterman Reingold layout to the network and set the size of each node to its betweenness centrality which can be seen in Figure 7. We also used Gephi's Community Detection setting to show that the network has 4 distinct communities similar to the results we saw using R.

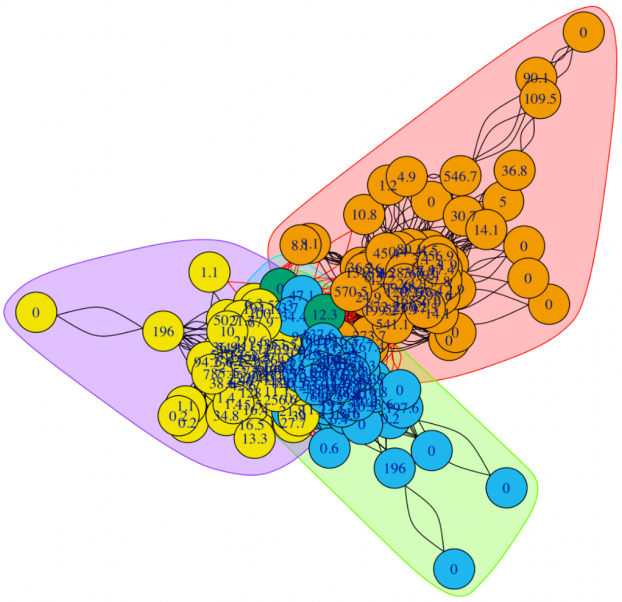

Figure 6: Jazz Network plotted using the Fast Greedy Algorithm using R.

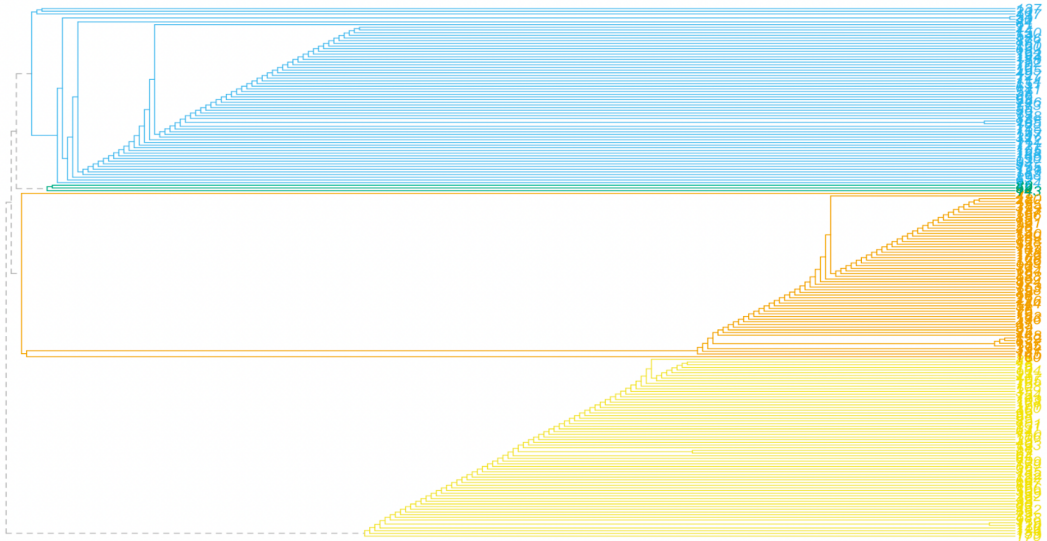

Figure 7: The corresponding dendrogram for the network graph in Figure 6 using R.

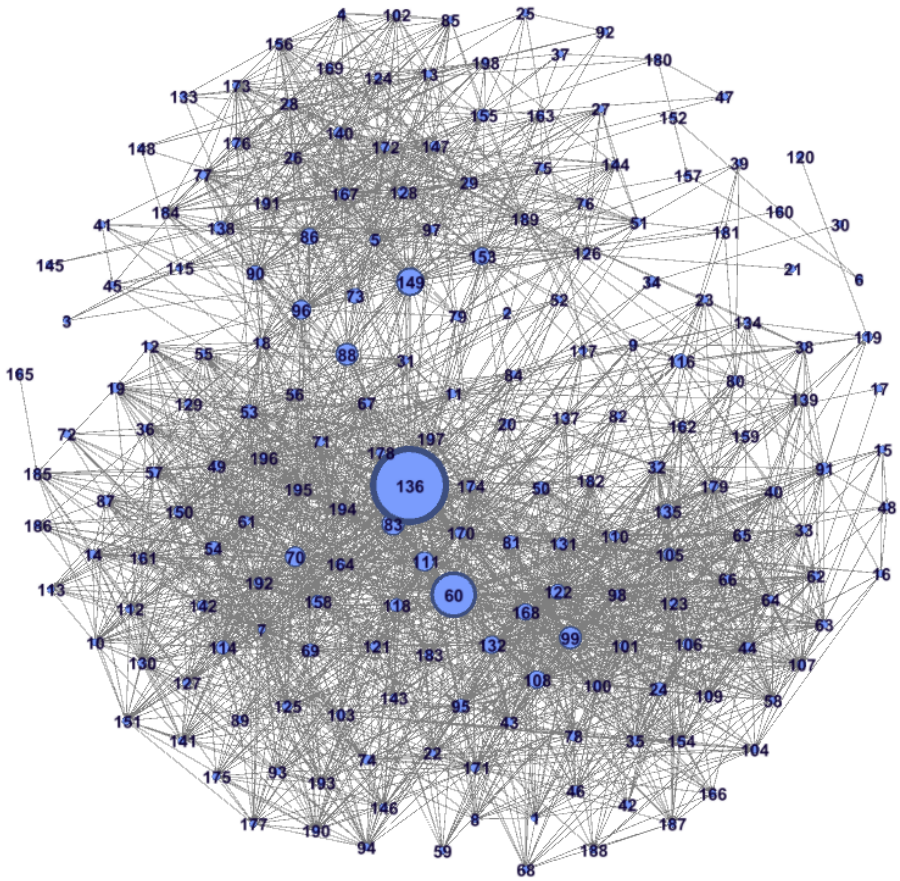

Figure 8: The Jazz Network plotted using the Fruchterman Reingold Algorithm using Gephi.

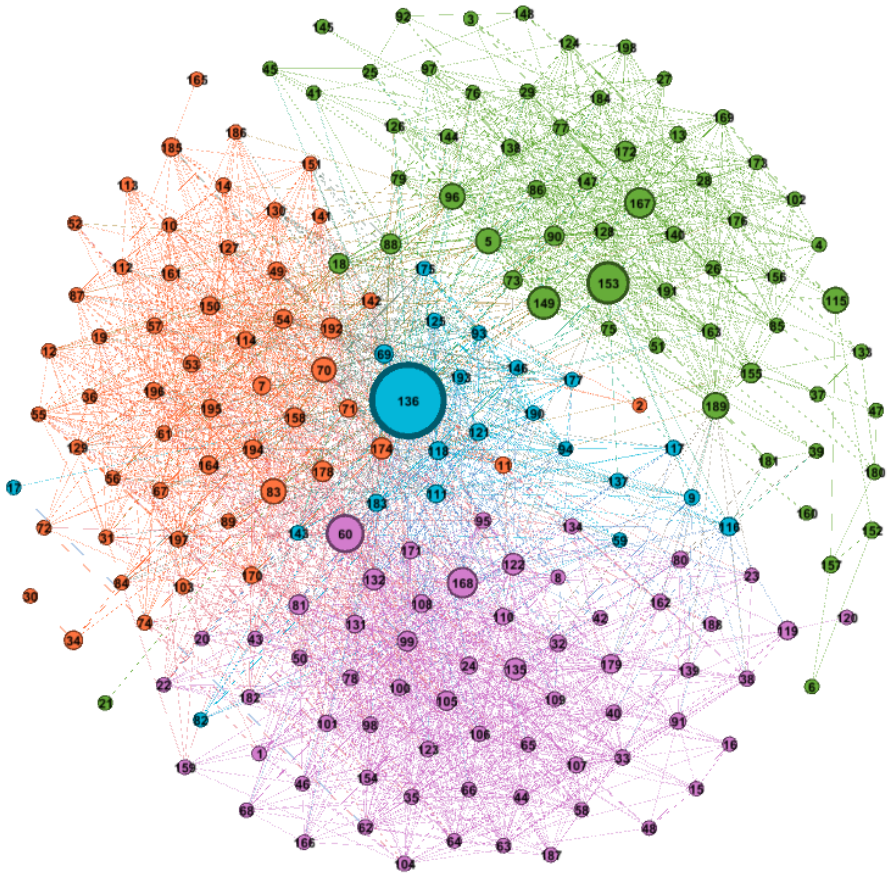

Figure 8: Community Detection for the Jazz Network using Gephi using Gephi

Conclusions

Overall, we found the network to behave in the way as a “small-world” network would. Jazz musicians collaborated with one another frequently. As we saw there were a few musicians with a high number of connections between them through the various plots of centrality that we showed previously and the degree centrality plot being right skewed. We see evidence of the improvisation jazz musicians displayed which led to inspiring other musicians to improve upon and spread the genre further.

This leaves us with questions such as, is this phenomenon observed in the music industry in general? Or is this unique to Jazz alone?

Bibliography

Easley, David, and Jon Kleinberg. Networks, crowds, and markets: Reasoning about a highly connected world. Cambridge University Press, 2010.

Kolaczyk, Eric D., and Gábor Csárdi. Statistical analysis of network data with R. Vol. 65. Springer, 2014.

Jackson, Matthew O. Social and economic networks. Princeton: Princeton University Press, 2008.

Pablo M. Gleiser., and Leon Danon. Community Structure in Jazz. Departament de Física Fonamental, Universitat de Barcelona, 2003.

Data Source: https://deim.urv.cat/~alexandre.arenas/data/welcome.htm