Network of Shared Jazz Musicians

Amairani Garcia
Abbey Bethel

05/18/2014

Introduction

In 2003, Pablo Gleiser and Leon Danon of the Fundamental Physics Department at the University of Barcelona in Spain analyzed the collaboration network of popular Jazz musicians.

Using data obtained from The Red Hot Jazz Archive digital database (www.redhotjazz.com), the duo looked at 198 bands from 1912-1940, of which there are 1275 different names of musicians, and created two networks. The first is a network of musicians, where the musicians are the nodes and the edges indicate two musicians having been in the same band (but not necessarily at the same time). Additionally, Gleiser and Danon looked at the band collaboration network in which there are 198 nodes representing bands and 2742 unweighted edges which indicate at least one common musician between two bands. Of the two networks, only the data for the jazz band network was available, so the musician network was not analyzed.

Gleiser and Danon identified the band network as a small world network, since “the average distance between vertices is small, while the clustering of vertices remains high” (Gleiser, Danon). Figure 1 is a Gephi Visualization of the jazz band network; there are three cliques, and the overlap or clustering of nodes supports the notion of this network being a small world network.

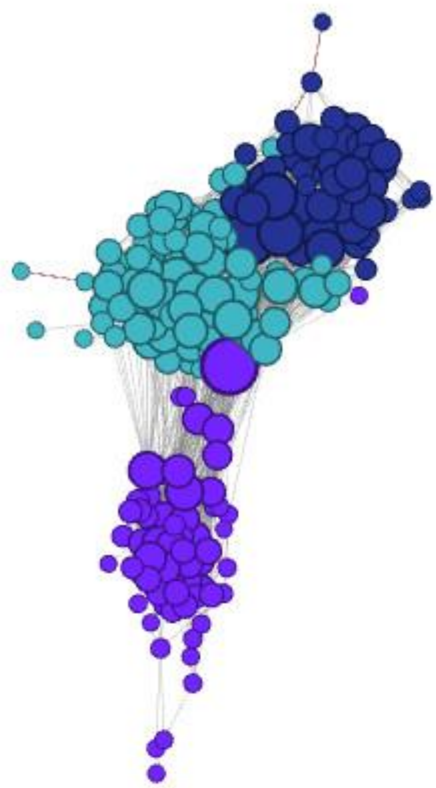

Figure 1: Gephi Visualization of the jazz band network. The nodes are organized by modularity, which expresses three defined cliques with high clustering of nodes in each sub-network.

The degree distribution, clustering coefficient, and average shortest path length are three properties that can support the identification of this network as small world.

Methods and Results

In order to evaluate a small world network, three parameters are needed: *N* (number of nodes), *k* (initial degree of nodes), and *p* (rewiring probability).

The number of nodes, *N*, is 198, the number of bands in the jazz band network. To define *k*, the initial degree, the average degree, $\langle k \rangle$, of the jazz band network was found to be 27.6970. Since *k* is an integer, the average degree was rounded up to 28. Given that N and k are defined, several different rewiring probabilities between 0.01 and 0.1 were implemented; the value p=0.04 provided the most similar clustering coefficient and shortest path length values. 0.04 is an acceptable value that would provide a high clustering coefficient and low shortest path length, properties characteristic of a small world network.

Degree Distribution, P(k) vs k

In a small world network, the degree distribution is approximately binomial. The spread is very small and is centered at the degree *k*, which is a defined parameter of the small world network. The degree distribution for our representative small world network is the left graph in Figure 2. The spread is from 25 to 30, and the center of the distribution is at 28, the *k* value for this network.

The degree distribution for the jazz band network is pictured in the right graph of Figure 2. The distribution has a roughly even spread over the range 0 to 60. There are four degree values which lie outside this spread.

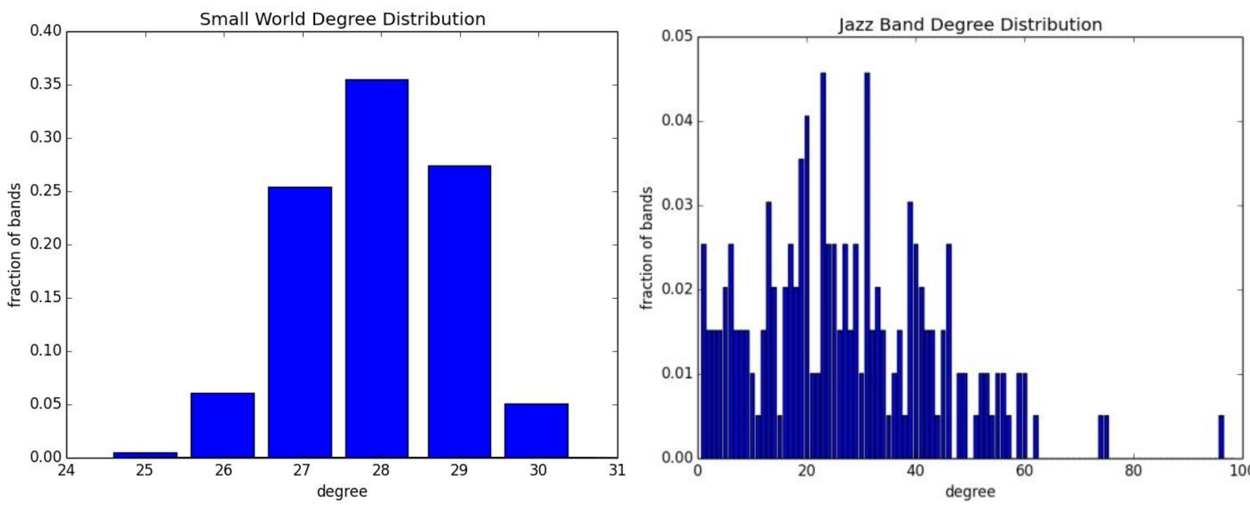

Figure 2: Left graph represents degree distribution for small world network-- roughly binomial, and centered around degree *k*, which is a parameter of the network. Right graph is the degree distribution for the jazz band network, which has an average degree of 27.6970. The bands spread is roughly even over 0 to 60 degrees, with four “outliers.”

Table 1 lists the ten nodes with the highest degrees. The first four nodes are the outlying nodes with noticeably higher degrees. These nodes could represent bands that had great notoriety among the jazz community, bands that existed during the entire 28 year period, or bands that had high turnaround as a transitional band for up-and-coming musicians.

Table 1: Ten nodes with the highest degrees. Highlights four outlying nodes.

Node Number	**Degree, k**
136	100
60	96
132	75
168	74
70	62
99	60
108	60
83	59
158	59
7	57

While the degree distribution for the jazz band network was not extremely similar to the degree distribution for a small world network of same size and average degree, it is understandable. In a small network, all nodes start with the same number of edges, which are rewired with probability, *p*. The jazz band network nodes started with zero edges and over time, nodes increased in degree. Some nodes developed many edges, while other nodes only gained a few edges.

Clustering Coefficient, CC

The clustering coefficient is a fraction, between 0 and 1, that represents the total number of possible edges under the actual number of edges present. If a network is highly connected, as small world networks are, the clustering coefficient would be between 0.5 and 1. For the comparison small world network, the clustering coefficient fell in the range of 0.60 to 0.66. This range is a result of the rewiring probability changing slightly in subsequent iterations of the small world network. In our jazz band network, the clustering coefficient was 0.617451. Table 2 compares the clustering coefficient for the small world network and the jazz band network.

Table 2: Clustering Coefficient for jazz band network and equivalent small world network.

	Jazz Band	Small World
Clustering Coefficient, CC	0.617451	0.60 - 0.66

As seen in Figure 1, the Gephi Visualization of the jazz band network, the high clustering coefficient is responsible for the formation of the three cliques.

Average Shortest Path Length <d>

The average shortest path length represents the number of musicians a band would need to go through in order to get in touch with a new band. If a network is highly connected, as small world networks are, the average shortest path length would be a relatively low number. For the comparison small world network, the average shortest path length fell in the range 2.30 to 2.40. This range is a result of the rewiring probability slightly altering the network, and therefore the shortest path length, in each iteration. In our jazz band network, the average shortest path length was 2.235040. Table 3 compares the average shortest path length for the small world network and the jazz band network.

Table 3: Average shortest path length for jazz band network, equivalent small world network.

	Jazz Band	Small World
Average Shortest Path Length, <d>	2.235040	2.3 - 2.4

As seen in Figure 1, the Gephi Visualization of the jazz band network, the low shortest path length is responsible for the overlap or proximity of nodes within each clique. An average shortest path length of 2.235040 means that a band would have to go through approximately two musicians in order to connect with a new band, which is an impressively low number.

Challenges and Conclusions

The significant challenge in analyzing the given data was the incompleteness. The data included 198 band-nameless nodes and 2742 unweighted edges connecting the nodes. If there was data on the musician network, names for the 198 bands, or weights given to the links, a more complete network story would have been given.

Gleiser and Danon evaluated two networks, a collaboration network of jazz musicians and a collaboration network of jazz bands. They represented the same data in two different regards, and were able to draw interesting information from the two networks. Unfortunately, the only data currently available regards the jazz band network.

Of the jazz band network, the nodes were numbered 1 to 198. While Gleiser and Danon infer that racial segregation and band recording location may have played a role in the clustering of this network, without the names of the bands, there is no easy way to verify such inferences.

The edges of this network indicate two bands have had at least one musician in common over the 28 year period from 1912-1940. This information is misleading. Two bands could have had 30 musicians in common or just one musician in common, but both cases are represented with an unweighted edge. If the edges were weighted to correspond to the number of musicians in common among two bands, a more interesting network story would be apparent.

The jazz band network resembled a small world network. The degree distribution, clustering coefficient, and average shortest path length were analyzed, since these properties can indicate whether a network is small world. The clustering coefficient was sufficiently high and the average shortest path length was sufficiently low, which agrees with the qualities inherent in small world networks.

The highest degree nodes could have been particularly well known bands or bands with high turnover. The clustering of the three sub-networks could have been a result of recording studio location or racial segregation. The average shortest path length demonstrates how few musicians were needed in order to connect two unconnected bands. With more information on the network, a more colorful network story can be illustrated.