MINI PROJECT - COMMUNITY STRUCTURE IN JAZZ".

Sergio Iván Arguello A, Sebastian Campiño Figueroa, and Valentina Ruiz Nova

1

Abstract—**The objective of this paper is to present the mini project - project proposal for the Analysis course of the Network Science for Data Analytics course.**

Index Terms—**Collaborative Jazz Music, Collaborative Network, Community Structure Analysis**

I. PAPER CITATION

IN this document we will discuss a summarize a little about the next paper:

GLEISER, PABLO M.; DANON, LEON (2003). COMMUNITY STRUCTURE IN JAZZ. Advances in Complex Systems, 6(4), 565–573. doi:10.1142/S0219525903001067

II. PROBLEM DESCRIPTION

The article focuses on collaboration networks within the context of jazz music. Specifically, it examines two levels of networks: one for individual musicians and another for bands. These networks are connected if they share musicians. The problem addressed is to understand social interactions and collaborative patterns in the history of jazz music.

III. PROJECT GOAL AND SCOPE

Apply some methods studied in class to reproduce the paper and understand how they were able to do it through network science and various analyses. This will help us understand the relationships between the musicians in the bands."

IV. CASE STUDY

We are going to study some relationship between the musicians and bands, duenot having the same tools/software and additional information from the network, an approach is made from the knowledge one has and the tools available.

V. NETWORK DATA SET

Data were obtained from The Red Hot Jazz Archive digital database. Included in the analysis were 198 bands that performed between 1912 and 1940, with the majority of bands in the 1920s. The database lists the musicians who played in each band without distinguishing which musicians played at different times. The bands contain 1275 different names of musicians.

Two levels of collaborative networks are examined: a network of individuals (musicians) and a network of bands, connected if they share musicians.

Although the paper mentions the existence of two different networks, the accompanying file consists only of one network file, lacking labels to identify the node names. Consequently, it was later identified that this corresponds to the bands dataset based on its behavioral metrics in the exploratory analysis conducted through Gephi.

Therefore, the subsequent analysis aimed at reconstructing this article is focused on reconstructing the results presented in the bands analysis. It should be noted that this network is asymmetric (links have direction from a source node to a destination node), whereas the paper provides analysis values for a symmetric network. Additionally, it is assumed that the links do not have weights, as if a band shares one or more musicians, the link will still have a value of 1. Thus, this is a binary value indicating existence (whether a link exists or not).

VI. IMPLEMENTED NETWORK SCIENCE APPROACH

a) Topology Analysis of the Network

The initial approach taken to address the paper involves conducting a topology analysis of the network, focusing on degree distribution and average network degree.

The first step in understanding the network's topology is calculating the average geodesic distance, which, through Gephi, yields a value of 2.238, close to the value reported in the paper (2.26). This indicates, as mentioned in the paper, that the network exhibits behavior similar to small-world networks.

Next, the degree distribution of the nodes is examined. To achieve this, the average degree of the nodes in the network must first be calculated. Using Gephi, the degree distribution graph is obtained, as shown in Figure 1. The x-axis of this graph represents the value K, i.e., the number of connections a

COURSE: Network Science for Data Analytics 2024

node can have, while the y-axis displays a count indicating how many nodes have a specific degree.
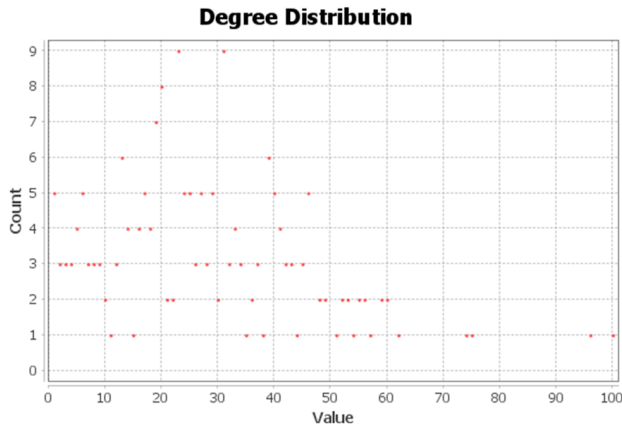

Fig. 1. Degree Distribution in Gephi.

For practical purposes, and because Gephi does not generate custom statistical graphs, the data laboratory section is exported to a CSV file. With this, it becomes possible to create statistical graphs that better represent the data presented in the paper.

Another way to represent the degree distribution is through a histogram, as shown in Figure 2.
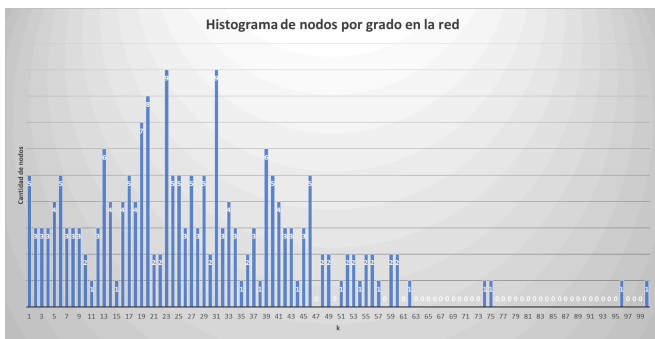

Fig. 2. Histogram

By normalizing this histogram, dividing it by the number of nodes in the network, we obtain the degree distribution P(k). Subsequently, the cumulative degree distribution F(k) is calculated by summing the probabilities P(k') for all degrees k' that are greater than or equal to k. This yields the graph shown in Figure 3.
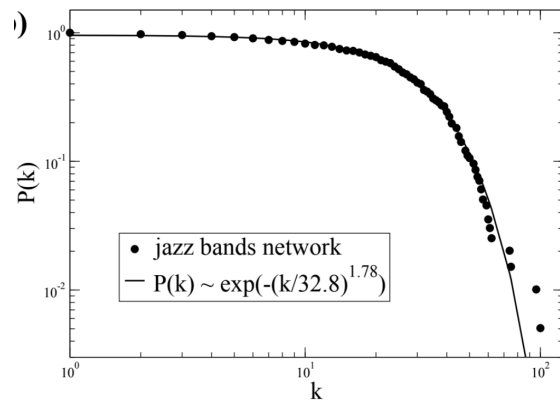

Fig. 3. Degree distribution P(k)

In the paper, the authors approximate the behavior of the nodes to a negative exponential function. Additionally, the graph shows that the majority of nodes are concentrated in values of k between 10 and 60.
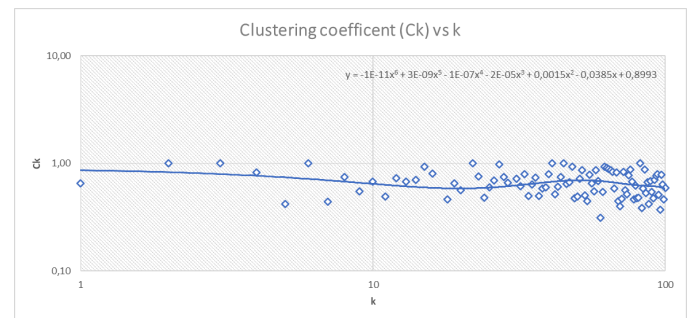

Fig. 4. Behavior of the clustering coefficient

In Figure 4, the behavior of the clustering coefficient as a function of node degree (k) is depicted. As mentioned in the paper, there is a trend towards the maximum coefficient with k > 30, followed by a slight decline after k > 60.

b) Community Structure Analysis in the Network

For the network described in the paper, four communities corresponding to the cities where bands record their albums are identified. Through this graph and by analyzing the labels, the paper establishes a relationship between the location of the bands and racial segregation among ethnically white and black bands, which is one of the research questions addressed in the paper.
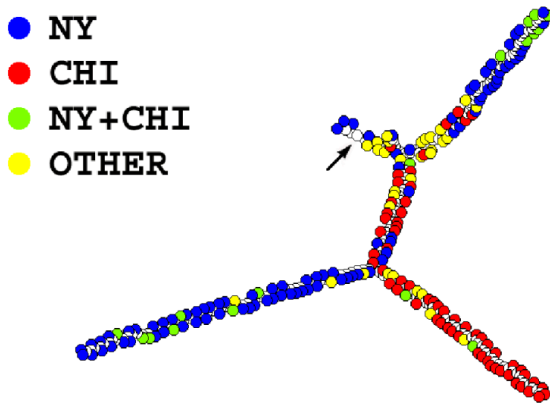
Fig. 5. Taken from: [1]. Communities in the jazz bands network. The arrow indicates the root of the tree. The different colors correspond to cities where a band has recorded: New York (blue), Chicago (red), both in New York and Chicago (green) and other cities (yellow).

To reconstruct this, within Gephi using its graphical interface (Overview), applying the exploratory network analysis led to the same result in searching for communities within the network. In Figure 6, the network is depicted with communities highlighted in four different colors, and nodes of varying sizes representing those with higher degrees.
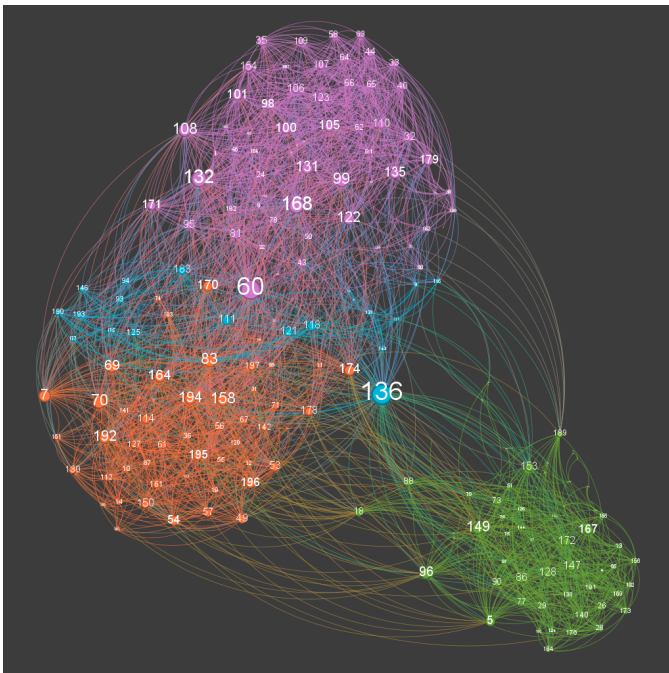

Fig. 6. Communities in the jazz bands network

VII. CONCLUSIONS

This mini project helped us understand the process used in the paper. Replicating the steps with the methods learned in class

allowed us to expand our knowledge and gain a deeper understanding of the relationships depicted in the networks.

Additionally, the community division demonstrates a robust association with the geographical locations of band recordings. These findings highlight that musician and band networks encapsulate crucial aspects of the collaborative jazz musician network.

Another tool called graph tool was explored, but the results weren't better than the obtained before, in the github repository there is a Notebook with this approach.
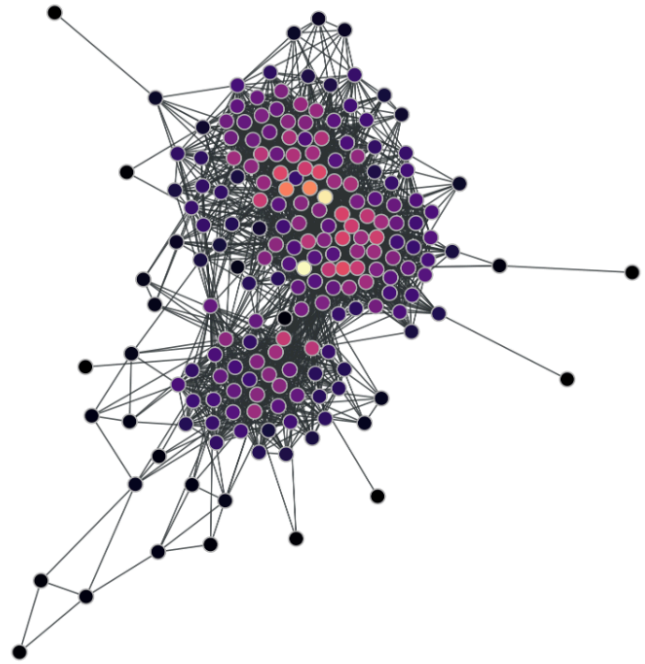

Fig. 5. Degree of each node represented using Graph Tool

VIII. WEB LINKS TO SOURCE CODE AND EXPLANATORY VIDEO

HTTPS://WWW.CANVA.COM/DESIGN/DAF_Z5OSRLO/84W_ZORZAARXJIPG3YIDDG/VIEW?UTM_CONTENT=DAF_Z5OSRLO&UTM_CAMPAIGN=DESIGNSHARE&UTM_MEDIUM=LINK&UTM_SOURCE=RECORDING_VIEW

HTTPS://GITHUB.COM/SHCAMPINOF/NS4DA-MPFR-3

IX. TEAM MEMBERS CONTRIBUTIONS

Team member	Role	Activities/ Contributions
Sergio Iván Arguello A	Resource Booster/Researcher	Preparation of the document DataSet exploration (Gephi)and excel help
Sebastian Campiño Figueroa	Cohesive / Specialist	Literature review Preparation of the document and apply of methods (Gephi)
Valentina Ruiz Nova	Finisher/monitor	Preparation of the slides and video Review Case Study and problem definition

REFERENCES

[1] GLEISER, PABLO M.; DANON, LEON (2003). COMMUNITY STRUCTURE IN JAZZ. Advances in Complex Systems, 6(4), 565–573. doi:10.1142/S0219525903001067.