MINI PROJECT– PROJECT PROPOSAL "COMMUNITY STRUCTURE IN JAZZ".

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

[[1]](#footnote-0)

***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

El objetivo del proyecto consiste en unas pocas frases con una

declaración ejecutiva de los objetivos del proyecto y lo que espera conseguir.

IV. CASE STUDY

Resuma el estudio de caso que su miniproyecto ha intentado reproducir.

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.

-----------------

Aunque en el paper se dice que hay dos redes diferentes, el archivo que lo acompaña, sólo se compone por un archivo de una red, que no tiene etiquetas para conocer los nombres de los nodos, con lo cual se identifica posteriormente que corresponde al dataset de bandas por sus métricas de comportamiento en el análisis exploratorio por medio de Gephi.

Es por esto que el análisis posterior que se hace para reconstruir este artículo está orientado a reconstruir los resultados que se presentan en el análisis de bandas. Cabe aclarar que esta red está es asimétrica (Los enlaces tienen dirección de un nodo de origen a un nodo destino), pero en el paper están dados los valores de análisis para una red simétrica. Así mismo, no se considera que los enlaces tengan pesos, ya que si una banda comparte un músico o más, igualmente el enlace tendrá valor de 1. Por lo que este es un valor binario que indica existencia (Existe enlace o no existe enlace).

VI. IMPLEMENTED NETWORK SCIENCE APPROACH

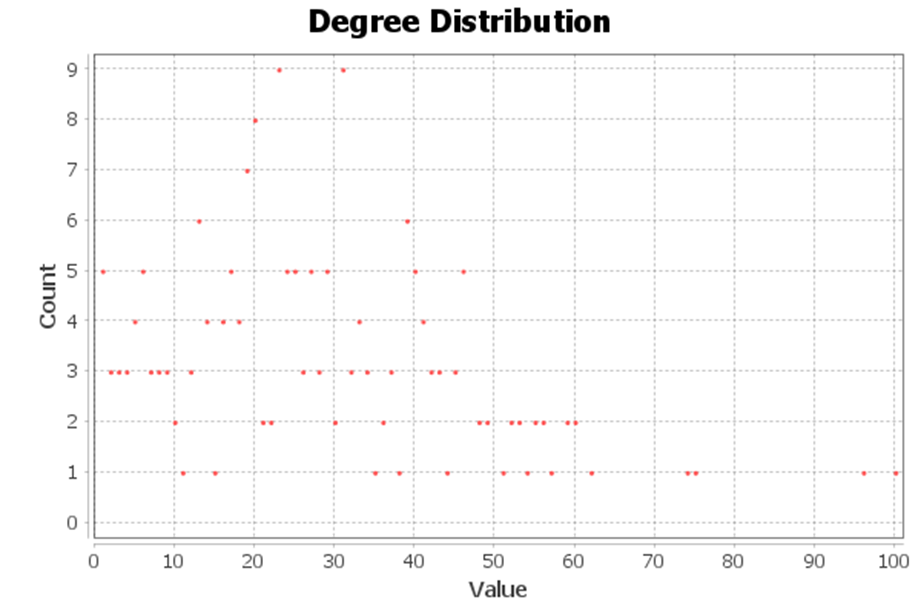
Describa en términos generales el enfoque de la ciencia de redes aplicado. Es necesario mencionar los tipos de modelos y métodos de redes que ha utilizado.

1. Análisis de la topología de la red

El primer acercamiento que se realiza para solucionar el paper consiste en hacer un análisis de la topología de la red, analizando la distribución de grados y average networks degree.

El primer cálculo para entender la topología de la red, es el cálculo de la distancia geodésica media, que por medio de Gephi, corresponde a 2.238 valor cercano al que muestra el paper (2.26), lo cual muestra, como se dice que en el paper que la red tiene comportamiento similar a las redes small world.

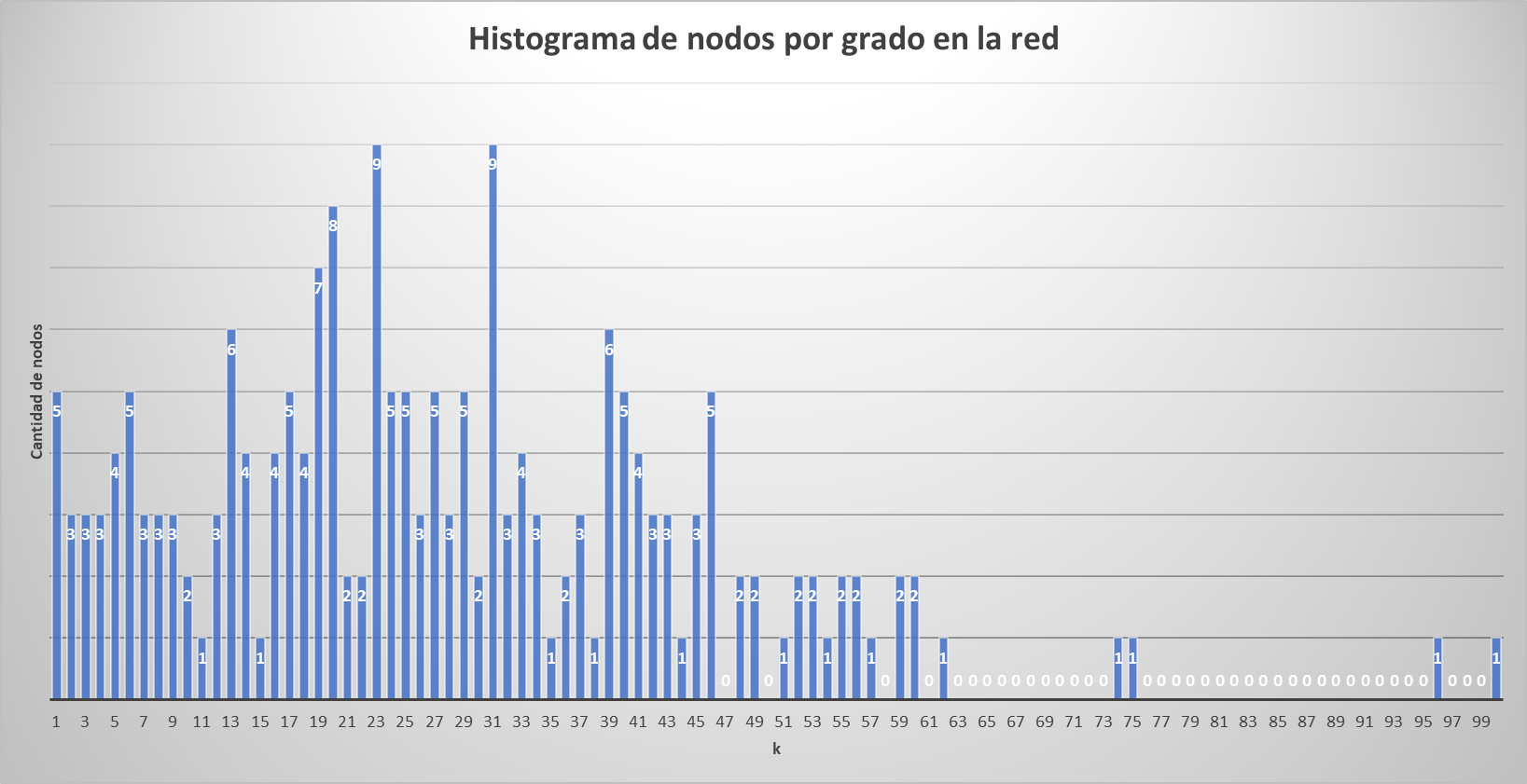
Luego se busca hacer el grado de distribución de los nodos. Para llegar a este resultado, primero se debe hacer el cálculo del grado medio de los nodos de la red. Por medio de Gephi, se obtiene la gráfica de la distribución de grado, como se observa en la Figura x. En el eje x de este gráfico se observa el valor K, es decir, el número de conexiones que puede tener un nodo. En el eje Y se observa un conteo, correspondiente a cuántos nodos tienen un número de grado específico.



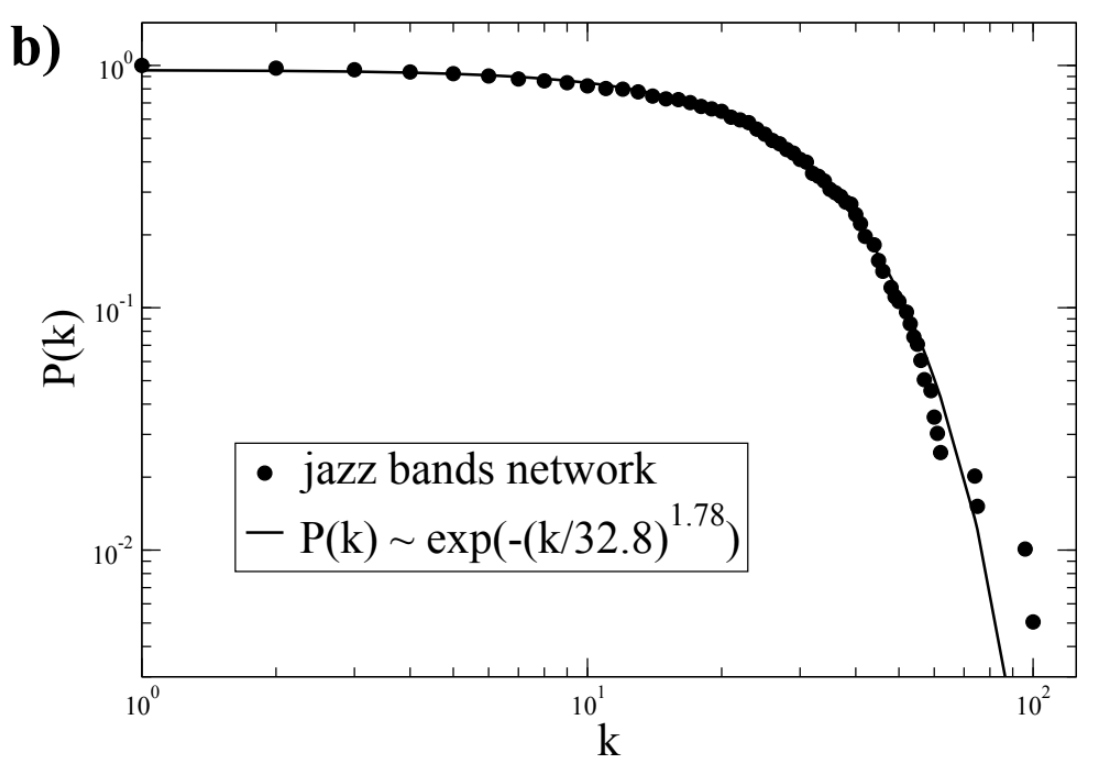
**Fig. 1.** Degree Distribution in Gephi.

Para fines prácticos y debido a Gephi no realiza gráficos estadísticos personalizados, se utiliza la exportación de la sección de Laboratorio de Datos, hacia un archivo csv, con lo cual, ahora sí se pueden realizar gráficas estadísticas que representen de mejor manera los datos que se muestran en el paper.

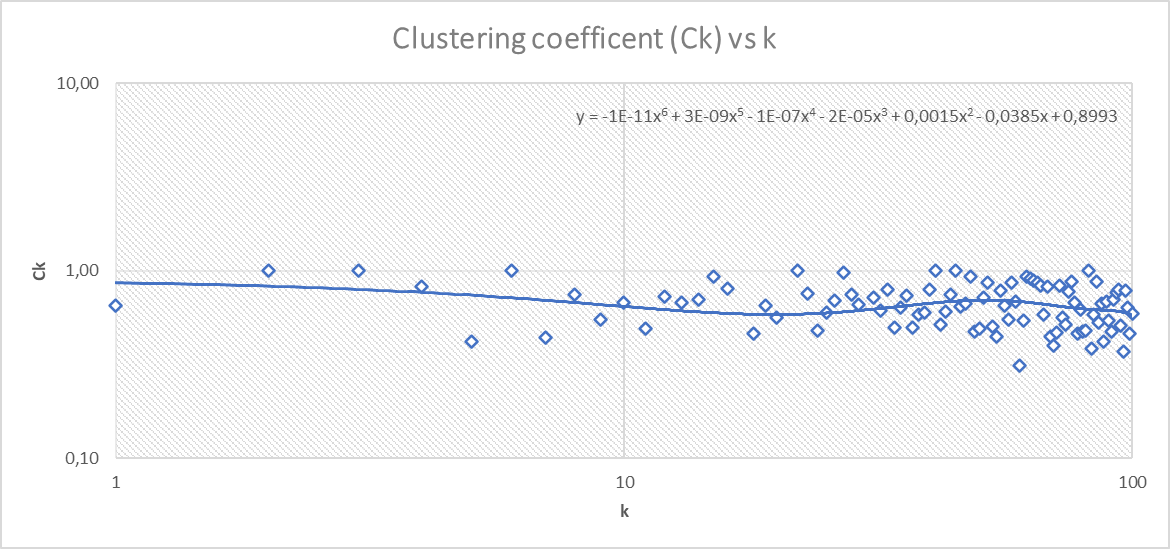
Otra forma de representar el grado de distribución, es por medio de un histograma, como se muestra en la Figura X.



Al normalizar este histograma dividiéndolo por el número de nodos que hay en la red, se obtiene la distribución de grado P(k), para después calcular la distribución acumulativa de grados F(k) sumando las probabilidades P(k') para todos los grados k' que son mayores o iguales a k. Obteniendo la gráfica de la figura X.



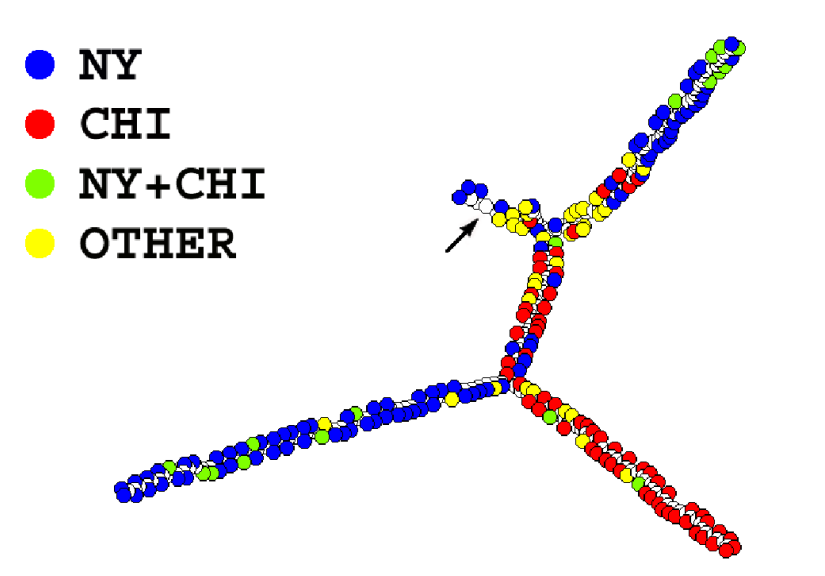
En el paper, los autores aproximan el comportamiento de los nodos a una función exponencial con exponente negativo, en el gráfico se observa además, que la mayoría de los nodos se concentran en valores k entre 10 y 60.



En la figura X, se puede ver el comportamiento del coeficiente de clustering, en función del grado de los nodos (k). Se observa lo mencionado en el paper, donde hay una tendencia hacia el máximo coeficiente con un k>30 y se presenta una leve caída después de k>60.

1. Análisis de la estructura de comunidad en la red

Para la red del paper, se encuentran 4 comunidades correspondientes a la ciudad donde las bandas graban sus discos. En el paper por medio de este gráfico y analizando las etiquetas, encuentran una relación entre la locación de las bandas y la segregación racial entre bandas etnicamente blancas y negras, que es una de las preguntas de búsqueda que se tienen como objetivo en el paper.

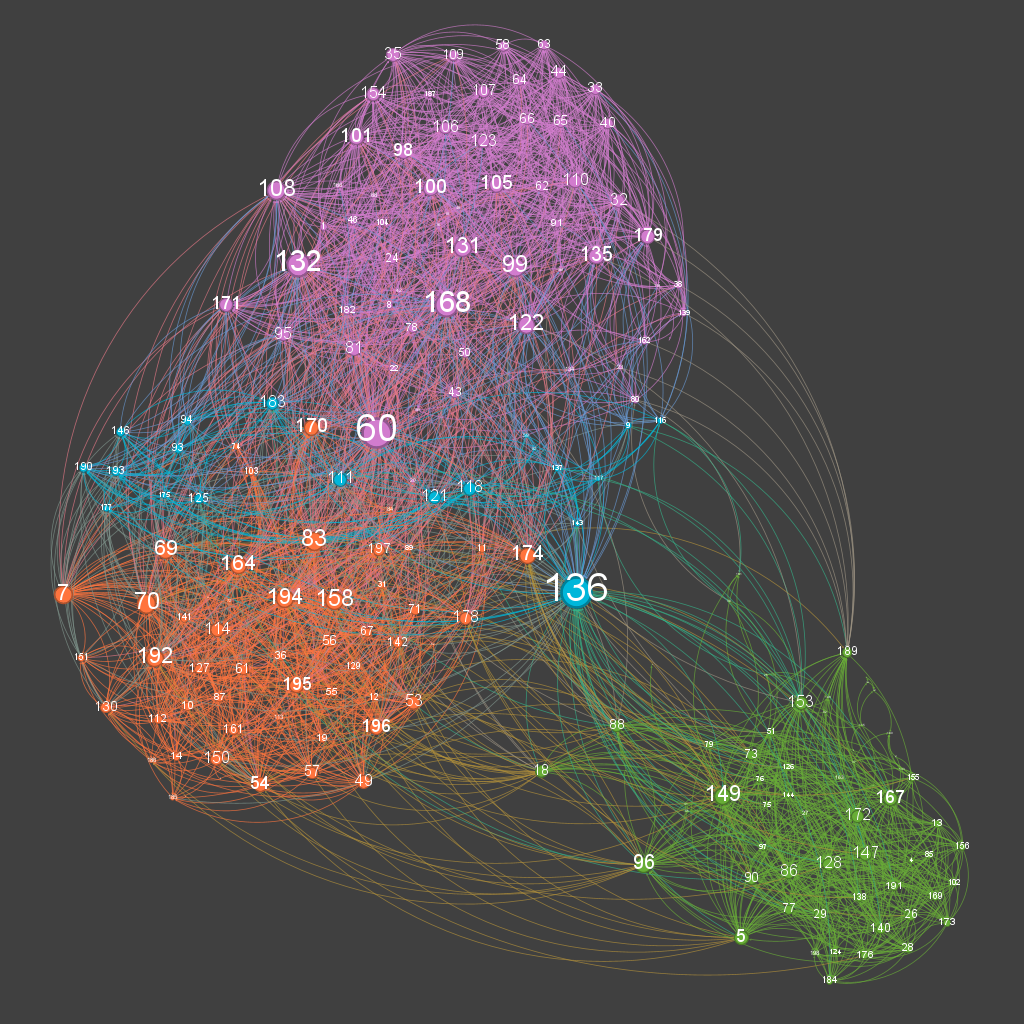


**Fig. 1.** Taken from: [1]. Communities in the jazz bands network. The arrow indicates the root of the tree. The

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).

Para reconstruir esto, dentro de Gephi con su interfaz gráfica (Vista general), aplicando el análisis exploratorio de la red, se pudo llegar al mismo resultado de la búsqueda de comunidades dentro de la red. En la figura X, se observa la red, con 4 colores distintos las comunidades dentro de la red y con diferentes tamaños de nodo, para aquellos nodos que tienen mayor grado.



VII. CONCLUSIONS

In summary, the results obtained show that the network of musicians is varied in degree and reflects the racial segregation among them. Furthermore, the division into communities presents a strong classification with the geographical locations where the bands recorded. These results demonstrate that networks of musicians and bands capture essential ingredients of the collaborative network of jazz musicians. With this in mind, it was quantitatively characterize the community structure. They calculate the cumulative community size distribution P(s), which gives the probability that a community has a size greater than or equal to s (Fig. 2). This distribution shows a significant skew, with a slow decline for community sizes up to s ∼ 200, followed by a faster decline and a cutoff corresponding to system size at s ∼ 1000.

VIII. WEB LINKS TO SOURCE CODE AND EXPLANATORY VIDEO

IX. TEAM MEMBERS CONTRIBUTIONS

| Team member | Role | Activities/ Contributions |
| --- | --- | --- |
| Sergio Iván Arguello A | Resource Booster/Researcher | Review and selection of data set- (network)  Preparation of the document  DataSet Explanation |
| Sebastian Campiño Figueroa | Cohesive / Specialist | Literature review  Preparation of the document and  Results explanation |
| Valentina Ruiz Nova | Finisher/monitor | Preparation of the slides  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.

1. [↑](#footnote-ref-0)