

Creación y estudio de la red de colaboraciones de artistas musicales de Argentina

Delgado Manuel, Rojas Christian Ariel y Waehner Nicolás

Departamento de Física, Universidad de Buenos Aires, Buenos Aires, Argentina. Diciembre de 2022

Resumen

En este trabajo se construyó y estudió una red de colaboración entre músicos argentinos. En primer lugar, se estudió la asortatividad de grado de la red y se concluyó que la red tiene una tendencia disortativa. Además, se encontró que hay homofilia por género y géneros musicales y un valor de clustering medio mayor al esperado por azar. Por otro lado, se analizó la conectividad de la red a partir de la remoción de nodos mediante distintas centralidades comparando con la remoción a partir de la popularidad de los artistas, y se encontró que los artistas más populares no son los más centrales en la conectividad de la red. También se estudió la ruptura de la red según los géneros musicales más multitudinarios. Luego, se detectaron comunidades con el algoritmo de intermedieatz y se difundieron los géneros musicales con propagación de etiquetas. Se encontraron agrupamientos entre algunos géneros musicales con las particiones por comunidades, pero no se encontró una correspondencia clara entre ellos. Finalmente, se predijeron futuras colaboraciones entre artistas no enlazados a partir de una de la similaridad entre ellos y se encontraron cualitativamente colaboraciones razonables.

1. Introducción

En este trabajo se armó y estudió una red de colaboración de artistas argentinos, a través del uso de varias API's, teniendo como objetivos:

- Analizar propiedades generales de la red, tanto topológicas como externas.
- Estudiar la homofilia entre artistas tanto por su género como sus géneros musicales, para analizar si artistas de un mismo tipo tienden a colaborar más entre sí.
- Analizar la conectividad de la red a través de la ruptura según distintos criterios topológicos e información externa.
- Estudiar la formación de comunidades en la red y analizar si hay alguna correspondencia entre las particiones y los géneros musicales.
- Predecir futuras colaboraciones analizando la similaridad entre los nodos.

2. Armado de la red

Se creó una red multienlace en la cual los artistas eran los nodos y se los enlazaba por cada canción en la cual los dos artistas hayan colaborado.

Para armar la red se comenzó por el artista argentino *Bizarrap* debido a que es el artista mas escuchado del país. Con la API de *Spotify* se recorrieron sus álbumes y luego las canciones dentro de cada álbum, guardando a los colaboradores por canción. Luego, se debió verificar que el colaborador sea argentino. Para ello, primero se buscó al artista mediante la API de *MusicBrainz* y se verificó si el país estaba dentro de la información del artista. En caso contrario, se intentó obtener de la misma API el área de procedencia del artista, y se verificó si coincidía con alguna de las provincias de Argentina. Finalmente, en caso de que *MusicBrainz* no tuviese ninguno de los dos datos, se buscó la nacionalidad del artista en *Wikipedia*. Si mediante ninguno de los tres pasos se encontraba información respecto a la nacionalidad, se descartaba al colaborador. De esta forma, se agregó a todos los colaboradores argentinos de *Bizarrap* a la red y se los enlazó con él. Luego se realizó lo mismo para cada uno de

los colaboradores de *Bizarrap*, agregando a cada artista argentino con el que hayan colaborado a la red y buscando las respectivas colaboraciones de estos hasta que no hubiese mas artistas que agregar. Por otro lado, se guardó el género y el género musical de cada artista en caso de estuviera dado por *MusicBrainz*.

Es importante notar que al verificar que el área de procedencia del artista coincida con alguna provincia argentina, se pudo cometer el error de agregar artistas de otro país que provinieran de una localidad con el mismo nombre que una provincia. Similarmente pudo ocurrir con artistas extranjeros que al buscar en *MusicBrainz* o en *Wikipedia* se confundiera con alguno homónimo argentino y se introdujera igualmente a la red. Debido a ello, se descartó manualmente a los artistas extranjeros encontrados en la red. Sin embargo, se considera que el porcentaje restante de ellos en la red es muy bajo.

Filtrado de las canciones

Sobre las canciones se filtraron las versiones en vivo, ya que se consideraron que no eran nuevas colaboraciones en la red ni versiones diferentes de una misma canción. Reversiones/remixes de una misma canción se consideraron como enlaces/colaboraciones diferentes.

Una aclaración importante es que debido a como se armó la red a través de *Spotify*, si un artista interpreta una canción de otro y este lo añade como colaborador, entonces estos dos colaboraron a pesar de que no hayan interactuado. Debido a que este filtrado no fue posible de automatizar se las consideró como colaboraciones de la red.

Filtrado de los artistas

Una vez con la red construida, fue necesario filtrar artistas repetidos. Ejemplos de ello serían "Fito Paez" y "Fito Paéz", o casos como "Aníbal Troilo" y "Aníbal Troilo y su orquesta típica", en los que se encontró diferencias de únicamente una tilde o de artistas con sus orquestas. Para ello, se quitó las mayúsculas, tildes, espacios y caracteres especiales de los nombres de todos los artistas y se verificó para todos los artistas si el nombre de algún otro artista estaba completamente contenido en el suyo. Una vez tenidas estas coincidencias, se verificó manualmente uno por uno que sean artistas efectivamente repetidos y se los descartó de la red, no sin antes pasar los enlaces del repetido al original.

Filtrado de los géneros musicales

Debido a que se obtuvieron los géneros musicales de los artistas de distintas fuentes y que estos estaban catalogados de una manera no convencional, se realizó un agrupamiento y filtrado de los géneros musicales a categorías que se consideraron más representativas. En la figura 1 se muestra un ejemplo de los géneros musicales inicialmente, y como se los unificó/filtro. Por ejemplo, todos los géneros musicales que tengan la palabra "trap", se los unió bajo un mismo género: "trap". Esto se realizó con todos los géneros, para luego elegir según la cantidad de artistas que poseían cada género, 17 géneros musicales distintos.

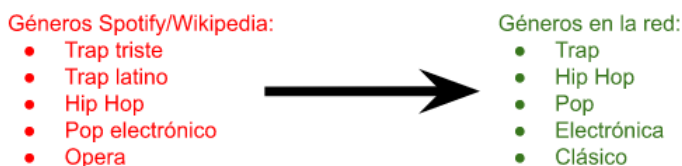


Figura 1 Ejemplo del agrupamiento de los géneros musicales.

Otro proceso de agrupamiento fue la unificación de géneros relacionados a un género principal, pero que en realidad son distintos. Un ejemplo de esto es la opera y la música clásica. Debido a que

en la red no había una cantidad considerable de artistas con este género, se lo combinó con el género clásico, que era más prolífico en la red y con el cual la opera comparte cierta similaridad musical.

En la figura 2 se puede observar la proporción de artistas por cada uno de los géneros musicales de la red luego de realizar el agrupamiento y filtrado de los géneros. Es importante notar que ya que los artistas pueden tener más de un género musical asignado, la suma de los porcentajes de artistas que componen cada género no es uno.

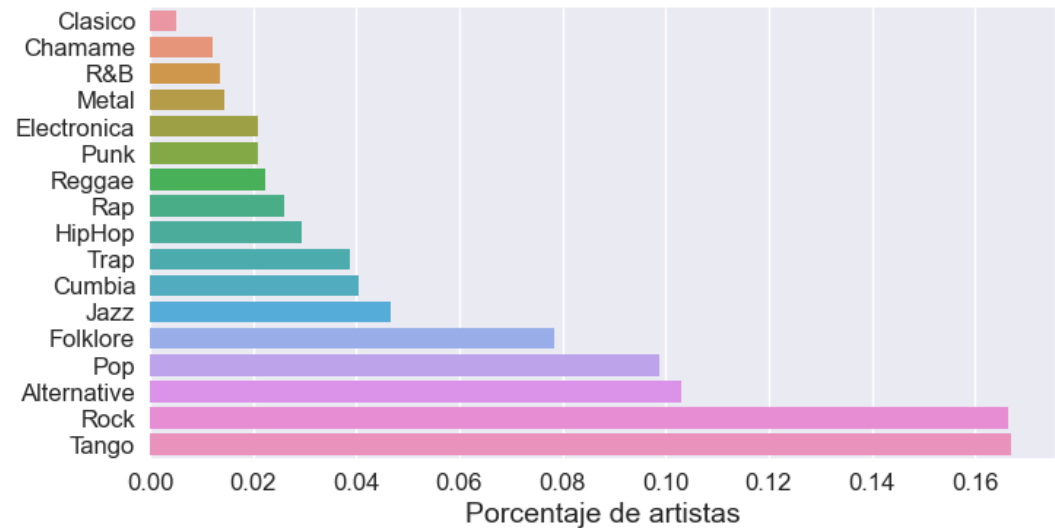


Figura 2 Cantidad de artistas por cada género musical de la red. Debido a que los artistas pueden tener más de un género asignado, la suma de los porcentajes de artistas que componen cada género no es uno.

Un aspecto notable de la distribución de géneros musicales de la red es que a pesar de haber comenzado en un artista como Bizarrap, que se dedica al trap o a la electrónica, no solo se ha llegado a artistas de otros géneros como reggae, clásico o tango, sino que los géneros más multitudinarios de la red son el tango y el rock, géneros que no posee el artista inicial de la red.

Información externa a la red

Sobre la mayoría de los artistas se conoce su género, sus géneros musicales y su popularidad, que es una magnitud que cuantifica que tanto ha sido escuchado recientemente cada artista en *Spotify*.

Sobre las colaboraciones/enlaces de la red se conoce el nombre de la canción y la fecha en la que la canción fue lanzada según *Spotify*.

3. Características de la red y análisis exploratorio

En la tabla 1 se observan las características de la red de artistas argentinos construida.

Tabla 1 *Propiedades topológicas de la red de artistas argentinos.*

Propiedad	Valor
Número de nodos	1388
Número de enlaces	11696
Grado medio	16.85
Coficiente de clustering medio	9.4×10^{-3}
Distancia media entre nodos	5
Diámetro	12

En la figura 3 puede observarse la red, donde se coloreó a cada artista por su género. A simple vista se observa que la mayoría de los artistas de la red son hombres. Esto se ve cuantificado en la tabla de la figura 3 donde se muestra el porcentaje de artistas de cada género. Los nodos con género "NA" fueron artistas cuyo género no se encontró en las diferentes API's, cosa que puede deberse a que el nodo sea un grupo o a que la API no tenga información del género del artista en cuestión.

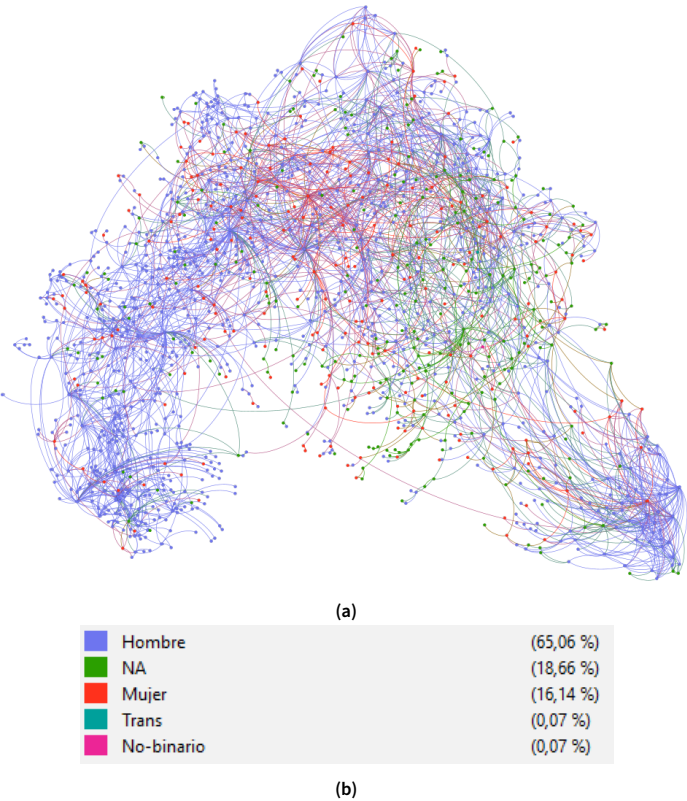


Figura 3 *Visualización de la red realizada con Gephi (a), donde los colores de los nodos representan el género de cada artista. Porcentaje de artistas de cada género en la red (b).*

Por otro lado, en la figura 4 puede observarse cuantos artistas se suman a la red por año, considerando su año de aparición como el año de la primer colaboración. Puede verse que la cantidad

de artistas sigue un comportamiento exponencial a lo largo de los años y que el año en el que se unió la mayor cantidad de artistas a la red fue 2017 con 134 nodos.

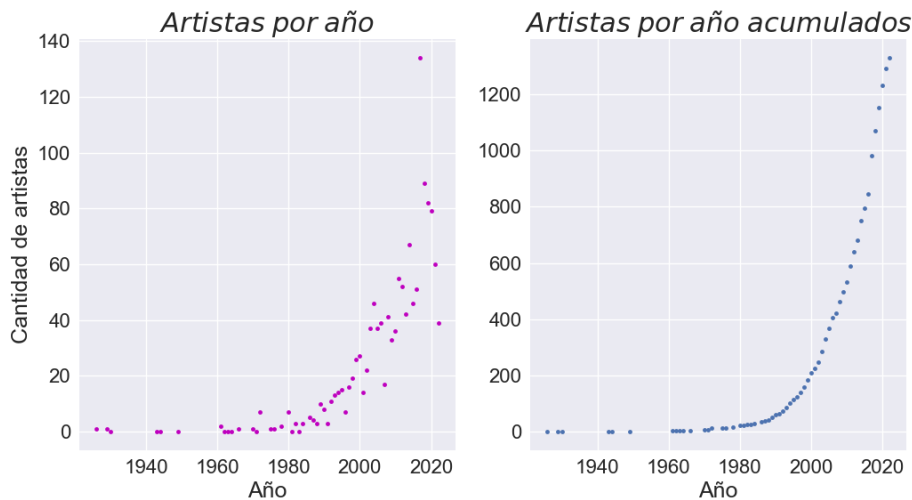


Figura 4 Crecimiento de la cantidad de artistas a lo largo de los años, atravesando un período desde 1925 hasta 2022.

Además, en la figura 5 se muestran las colaboraciones por año. En esta puede verse que las colaboraciones crecen levemente hasta los años 2000, donde la cantidad de colaboraciones empieza a crecer exponencialmente. El año en que hubo la mayor cantidad de colaboraciones fue 2020 con 919.

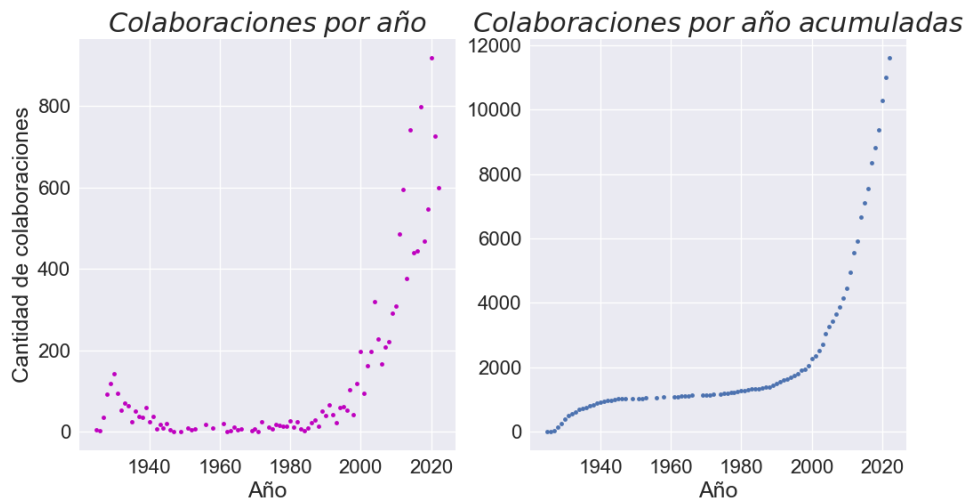


Figura 5 Crecimiento de la cantidad de colaboraciones a lo largo de los años, atravesando un período desde 1925 hasta 2022.

A partir de ambas figuras se observa que la mayor cantidad de información contenida en la red corresponde al último periodo, a partir de los años 2000.

4. Asortatividad de grado

Se analizó la asortatividad de grado para estudiar si artistas de un cierto grado tienden a colaborar con artistas de grado similar. En la figura 6 se puede observar el grado medio de los vecinos según el grado de cada nodo.



Figura 6 Grado medio de los vecinos según el grado, junto con el ajuste dado por el modelo de Barabási.

Para estudiarlo cuantitativamente, se utilizó el criterio de Barabási y de Newman. El criterio de Barabási [1], asume que el grado medio de los vecinos sigue la forma $k_{nn} = ak^\mu$ y al realizar un ajuste el valor de μ determina la tendencia asortativa/disortativa si es positivo o negativo respectivamente. Realizando dicho ajuste, que puede verse graficado en la figura 6, se obtuvo un $\mu = -0.37 \pm 0.02$ por lo que la red tiene una tendencia disortativa según el criterio de Barabási.

Calculando el criterio de Newman, el cual está dado por:

$$r = \frac{\sum_{ij} (A_{ij} - \frac{k_i k_j}{2m}) k_i k_j}{\sum_{ij} (k_i \delta_{ij} - \frac{k_i k_j}{2m}) k_i k_j}$$

que también habla de la asortatividad según el signo del coeficiente y tiene un rango de valores posibles entre -1 y 1 , se obtuvo un $r = 0.07$, por lo que según el criterio de Newman la red pareciera ser neutral. Por lo tanto, se estima que la red tiene una tendencia disortativa debido al criterio de Barabási.

5. Estudio de homofilia y clustering en comparación a la hipótesis nula

La homofilia es una magnitud que determina la tendencia a que nodos de una misma clase colaboren más entre sí que con nodos pertenecientes a clases distintas. Para calcularla se recorrieron todos los enlaces de la red y se dividió la cantidad de colaboraciones dadas por dos artistas con la misma clase por el total de enlaces de la red. Se realizó esto tanto para la homofilia por los géneros de los artistas como para sus géneros musicales. En el caso de los géneros musicales se calculó comprobando que los artistas compartan al menos uno de todos sus géneros.

Con respecto al género de los artistas, se comparó la homofilia de la red con la homofilia obtenida bajo dos hipótesis nulas distintas: reasignando el género de cada artistas y recableando todos los enlaces de la red, en ambos casos manteniendo la distribución original. La reasignación de género se realizó 5000 veces distintas, mientras que el recableado se realizó 1000 veces. En la figura 7 se observa el resultado para la homofilia de género original de la red y la distribución de homofilia obtenida bajo las dos hipótesis nulas.

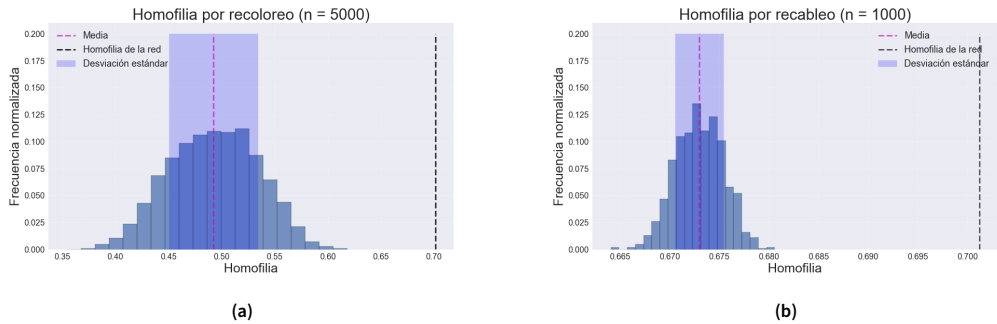


Figura 7 Frecuencia normalizada en función de la homofilia por género para cada hipótesis nula, junto a la homofilia de la red original. "n" indica la cantidad de iteraciones realizadas bajo cada hipótesis.

Se observa en la figura 7 que la red original presenta una homofilia mayor que la esperada por azar bajo estas dos hipótesis nulas, con un p -valor despreciable en ambos casos. Por lo tanto, se puede estimar que las colaboraciones entre artistas se ven influenciadas por el género de los artistas.

En el caso de los géneros musicales de los artistas, se comparó la homofilia de la red únicamente con las obtenidas al recablear todos los enlaces de la red 1000 veces. En la figura 8 se observa el resultado para la red original y la distribución obtenida bajo la hipótesis nula dada por el recableo.

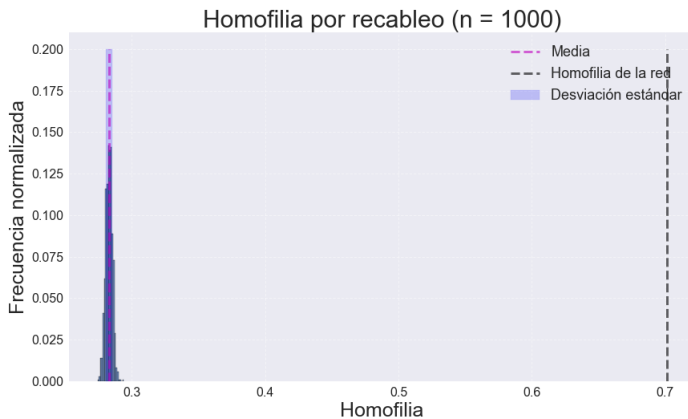


Figura 8 Frecuencia normalizada en función de la homofilia por género musical. Se observa la distribución dada al recablear junto a la homofilia de la red original. "n" indica la cantidad de recableos de la red.

Nuevamente, se observa de la figura 8 que la red presenta homofilia por género musical al ser comparada con la obtenida por recableo. Esto era esperable, ya que los artistas de un género musical tienden a colaborar con otros del mismo.

Finalmente, se estudió el coeficiente de clustering medio de la red. El clustering cuantifica que tan agrupado están los vecinos de un nodo entre si. Para calcularlo, se convirtió la red a una red pesada ya que el cálculo del coeficiente no está implementado para redes multienlace en *NetworkX*. Se comparó el clustering medio de la red con los obtenidos al recablear de mil maneras diferentes la red, cuyo resultado puede verse en la figura 9. Se observa que el clustering promedio de la red es mayor que el obtenido al recablearla, con un p -valor despreciable. Por ende, se encuentra que los artistas que tienen un colaborador en común tienden a colaborar entre sí más que lo esperado por azar.

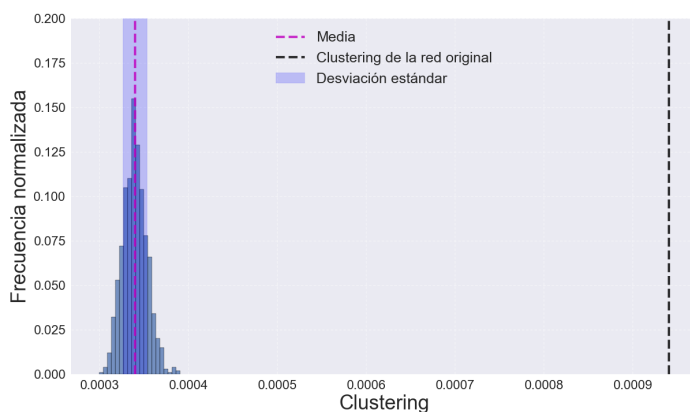


Figura 9 Frecuencia normalizada en función del clustering medio. Se observa la distribución dada al recablear la red de 1000 maneras diferentes junto al clustering de la red original.

6. Análisis de vulnerabilidad

El objetivo de esta sección es estudiar el rol de los artistas en la conectividad de la red; por un lado en función de la popularidad para compararlo con distintas medidas de centralidad, y por el otro a partir de los géneros musicales. Para ello, en primer lugar, se examinó la topología del grafo usando diferentes heurísticas. En particular, se calcularon las centralidad por grado, por cercanía y por intermediezt para cada nodo.

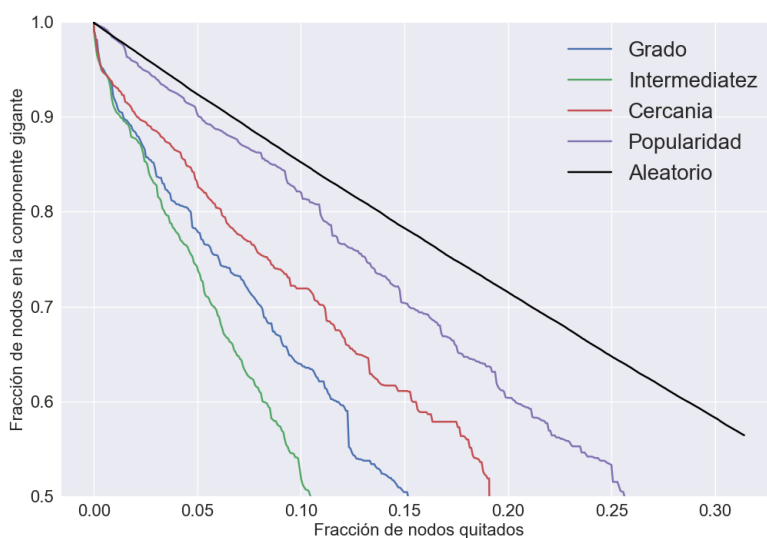


Figura 10 Cada curva muestra la fracción de nodos en la componente gigante como función de la fracción de nodos removidos de la red.

A partir de esto, se fueron removiendo nodos de la red según su centralidad y se calculó la fracción

de nodos en la componente gigante en función de la fracción de los nodos sacados en cada red. Lo mismo se hizo con la popularidad: se fue removiendo uno por vez a los artistas más populares de la red. A su vez, se repitió el proceso removiendo nodos aleatoriamente. Esto último se hizo para 1000 iteraciones diferentes y se promediaron los resultados de las distintas repeticiones. La figura 10 muestra las curvas obtenidas para todos los casos.

Se puede observar que los nodos con mayor centralidad por intermediaez y por grado resultaron ser más disruptivos en comparación a la remoción de nodos por cercanía y por popularidad. En particular, si bien el impacto de remoción en este último caso es mayor que el aleatorio, resulta ser menor que la remoción mediante cualquiera de las otras propiedades topológicas, por lo que la popularidad no resulta ser esencial en la conectividad de la red.

También se estudió el solapamiento de artistas más importantes entre todas las centralidades, tomando el 2% y el 10% de los artistas totales de la red. Este resultado puede observarse en la figura 11. En ambos se observa que los artistas populares tienen un menor solapamiento que el resto de centralidades, lo que reafirma que los artistas más populares no son los esenciales para mantener la topología de la red.

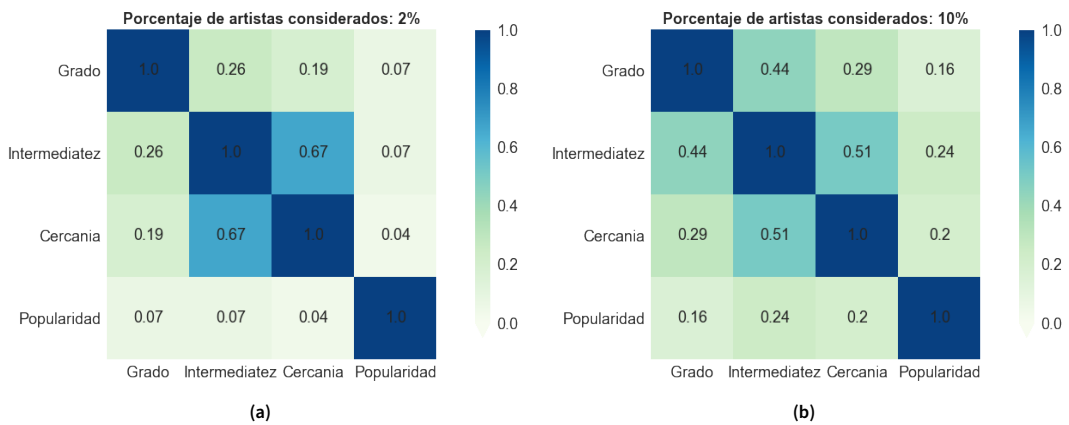


Figura 11 Solapamiento de artistas entre las distintas centralidades tomando como umbral dos porcentajes distintos de los artistas totales de la red: 2% (a) y 10% (b).

Por último, se rompió la red removiendo artistas según su género musical, buscando cual de estos es más central en la estructura de la red. Esto se hizo para los ocho géneros musicales mas multitudinarios, ya que la proporción de artistas en los demás géneros era muy reducida como para analizar la ruptura de la red. Se fueron removiendo aleatoriamente artistas de un mismo género y se realizó un promedio de estas remociones. Esto se hizo 100 veces para cada género musical y el resultado puede observarse en la figura 12.

En la mayoría de casos se observa de la figura 12 que la ruptura por géneros musicales es más disruptiva que la remoción aleatoria de nodos, lo que indica que los artistas de un mismo género musical poseen un cierto rol en el mantenimiento de la conectividad de la red. En cambio, remover artistas del género trap pareciera seguir un comportamiento similar a lo aleatorio, por lo que se estima que los artistas de este género tienden a colaborar únicamente con otros del mismo. Sin embargo, ninguno de los géneros alcanzó a romper en mayor medida la red comparado a las centralidades analizadas en la figura 10.

Un problema de este análisis es que, como se observó en la figura 2, en la red hay géneros más representados que otros, por lo que no se puede asegurar el comportamiento al romper la red en géneros con pocos artistas.

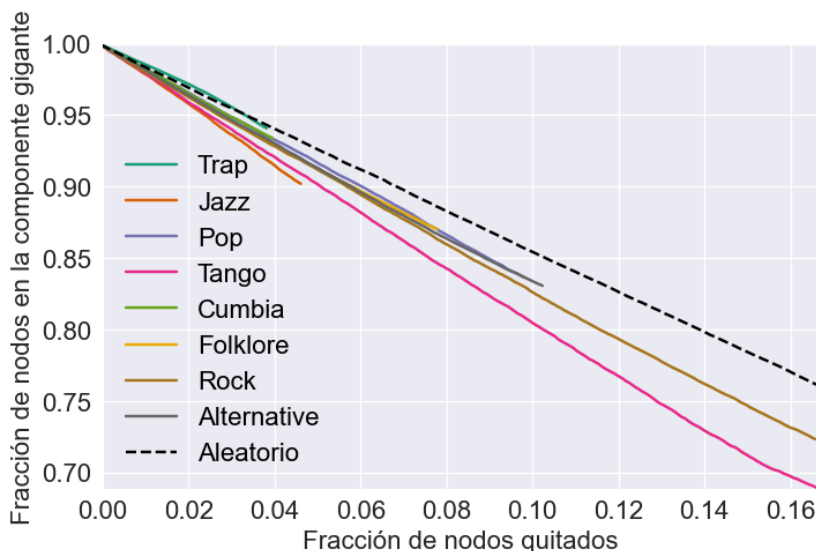


Figura 12 Ruptura de la red según la remoción de artistas de un mismo género musical.

7. Detección de comunidades y difusión de géneros musicales

El objetivo de esta sección es detectar comunidades en la red de artistas utilizando el algoritmo de *intermediatez*, y a su vez estudiar si existe algún tipo de correlación entre las comunidades detectadas y los géneros musicales de los artistas. Para ello, como únicamente alrededor del 53% de los artistas tenía al menos un género musical, se difundieron los géneros musicales sobre la red para los artistas sin género musical alguno.

Para cada género musical por separado, se utilizó el siguiente algoritmo de difusión basado en *Label propagation Zhu (2002)*:

- Se calculó la matriz de adyacencia W y la matriz diagonal D con los grados de cada nodo.
- Se generó un vector inicial Y^0 de longitud igual a la cantidad de artistas, con un 1 si el artista tenía ese género musical, un -1 si no lo tenía y un 0 si el artista no tenía ningún género.
- Se iteró hasta lograr la convergencia del vector $Y^{t+1} = D^{-1}WY^t$, reasignando en cada paso los valores de 1 y -1 en los vectores para los artistas que originalmente tenían dichos valores.
- Se asignó el género musical a los artistas sin género musical alguno según si el signo del elemento correspondiente al artista del vector Y^{t+1} era mayor a cero.

Es importante aclarar que los géneros musicales se agregaron al finalizar la difusión de todos los géneros, de manera que los artistas sin géneros musicales fueran siempre los mismos en cada difusión.

Para evaluar el desempeño del algoritmo se removieron los géneros musicales del 10% de los artistas previamente etiquetados y se predijeron los géneros musicales de los mismos, obteniendo una precisión de 0.75. Esta se midió triplicando el peso de los verdaderos positivos frente a los verdaderos negativos, ya que se priorizó la eficacia del algoritmo a obtener correctamente el género musical frente a que obtenga que un artista no tiene cierto género. Esto debido a que la cantidad de artistas con un género musical determinado es mucho menor que la cantidad de artistas que no lo tienen.

Por otro lado, el algoritmo de detección de comunidades por *intermediatez* de enlaces consiste en remover enlaces que conecten nodos de baja similaridad e ir rompiendo la red hasta remover el último enlace [1].

Utilizando este método se fijó el número de particiones en 17 de forma tal que coincida con

el número total de géneros musicales, para así estudiar si hay una correlación entre estos y las comunidades. En la figura 13 puede observarse la red coloreada según las comunidades determinadas por el algoritmo. Para esta partición se obtuvo una modularidad de 0.72, mientras que la máxima modularidad alcanzada con este método fue de 0.81 con 104 comunidades, lo cual es mucho mayor al número total de géneros, motivo por el cual no se estudió esta partición.



Figura 13 Red coloreada según las comunidades detectadas con el algoritmo de intermediaridad forzando a que las comunidades sean 17, igual a la cantidad de géneros musicales

En la figura 14 se muestra la matriz de correlación entre las comunidades/clusters y los géneros musicales de los artistas luego de aplicarle la difusión a la red. Cada fila de la matriz corresponde a una comunidad en particular, y cada valor de esa fila indica el número de artistas de esa comunidad que tienen como género musical al indicado para toda esa columna. En el caso en que cada comunidad correspondiese a un género musical, la matriz debería tener mayoritariamente artistas de un solo género, cosa que no se ve representada en las comunidades de la red.

Se observa de la figura 14 que si bien si hay géneros que parecen tener grandes comunidades donde son el género principal, como trap o tango, hay géneros que están repartidos en la red, como el rock, que tiene varias comunidades donde es el género mayoritario. Esto tiene sentido ya que los artistas participan en más de un género musical y que la división de los géneros musicales en 17 fue arbitraria. Todo esto conlleva a pensar en que la partición de comunidades de la red no está representada por los géneros musicales.

En el apéndice, en las figuras 17 y 18 pueden observarse nubes de palabras para los géneros más multitudinarios y los artistas de mayor grado de cada una de las comunidades.

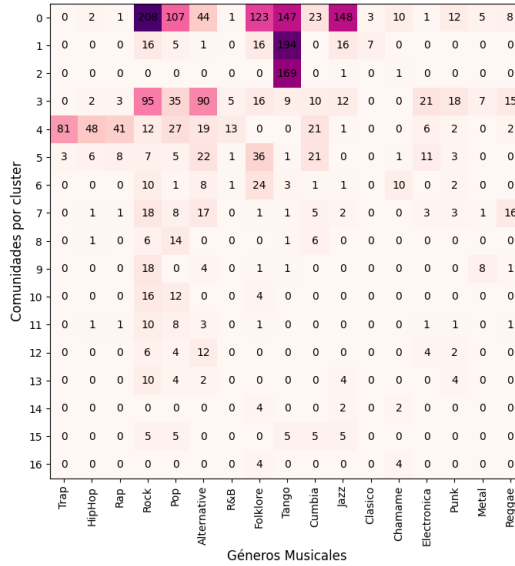


Figura 14 Matriz de correlación entre las comunidades obtenidas y los géneros musicales obtenidos a través de difusión.

8. Predicción de futuras colaboraciones a través de un estudio de la similitud entre nodos

El objetivo de esta sección es predecir futuras colaboraciones entre artistas de la red aún no enlazados. Para ello, en primer lugar se utilizó el algoritmo *Node2Vec* para pasar la red a un espacio métrico de 17 dimensiones (a modo de igualar con la cantidad de géneros musicales de la red). Luego, en segundo lugar se calculó la similitud coseno en este espacio para todos los pares de nodos conectados y no conectados, bajo la hipótesis de que los nodos conectados (artistas que colaboraron) resultarían mas similares entre sí que los que no. En la figura 15 se observan los resultados para la similitud coseno entre todos los pares de nodos enlazados y no enlazados.

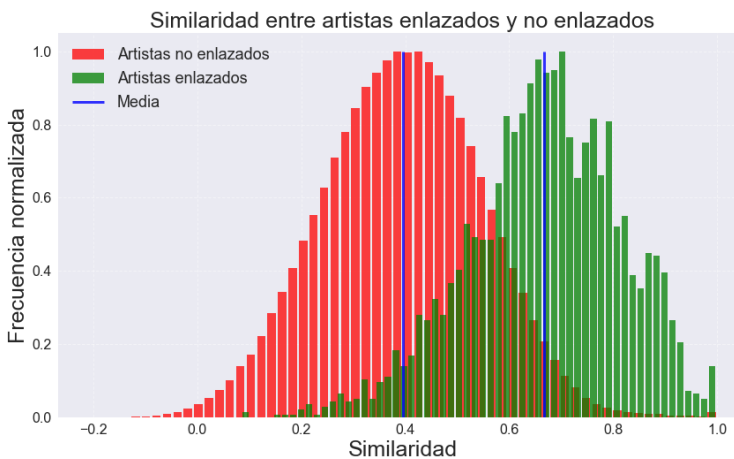


Figura 15 Frecuencia normalizada en función de la similitud coseno entre todos los pares de artistas enlazados y no enlazados.

Se puede observar de la figura 15 que ambas distribuciones se asemejan a distribuciones gaussianas, siendo la de artistas enlazados con una media mayor que la de artistas no enlazados. Por lo tanto, se tomó el criterio de que todos los pares de artistas no enlazados con un valor de similaridad mayor al de la media de los artistas enlazados podrían ser los futuros nuevos enlaces de la red. Cabe destacar que estos pares de artistas no enlazados con una similaridad mayor a la de los enlazados representan el 3% (alrededor de 30000 pares de nodos) del total de pares de nodos no enlazados. De este total de 30000 pares de artistas se seleccionó algunos artistas conocidos a modos de ejemplo. Se observan en la figura 16 posibles futuras colaboraciones entre artistas.



Figura 16 Futuras posibles colaboraciones establecidas según la similaridad entre los artistas.

Cabe destacar que se encontraron artistas que realmente colaboraron dentro de estos pares de nodos no enlazados con una similaridad mayor a la media de los artistas enlazados. Ejemplo de ello son Duki con Rei y Brian Taylor. Se estima que esto es debido a que únicamente se pudo recorrer las canciones dentro de álbumes para cada artista (y no sencillos, por ejemplo), por lo que algunas canciones con varias colaboraciones no fueron incluidas en el armado de la red. Sin embargo, es un factor positivo encontrar que estas falencias se encuentran en las predicciones.

9. Conclusión

En este trabajo se construyó una red multienlace a partir de las colaboraciones entre artistas de nacionalidad argentina. A cada uno de los artistas de la red se les agregó como atributos su género, su popularidad según *Spotify* y, en la mayoría de los casos, los géneros musicales del artista.

En primer lugar se estudió la asortatividad de grado para analizar si los artistas tienden a colaborar con artistas de su mismo grado. A través de los criterios de Barabási se encontró que la red tiene una tendencia disortativa, pero a través del criterio de Newman se encontró que la red es neutral.

También se calculó la homofilia de la red para el género del artista y los géneros musicales y el coeficiente de clustering medio y se los contrastó con una hipótesis nula a través de recolorar y/o recablear la red. A partir de los resultados obtenidos se concluyó que existe una tendencia a que los artistas colaboren más con artistas de su mismo género o con los que compartan un mismo género musical y que es más probable que artistas que hayan colaborado con otro artista, colaboren a su vez entre sí.

Por otro lado, se realizó un análisis de la vulnerabilidad de la red removiendo nodos según distintas centralidades, tales como el grado, la intermediezt y la cercanía, contrastando con la remoción de artistas a partir de su popularidad. Los nodos con mayor centralidad por intermediezt y por grado resultaron ser más disruptivos en comparación a la remoción de nodos por cercanía y por popularidad. En particular, el impacto de remoción en este último caso resultó ser menor que la remoción mediante cualquiera de las otras propiedades topológicas pero más que la ruptura por remoción aleatoria de nodos.

Estudiando el solapamiento entre un cierto porcentaje de artistas para dos centralidades distintas se observó que varios artistas que presentaban mayor centralidad bajo alguno de los criterios, también

presentaban una centralidad alta para otro. Sin embargo, se encontró que los artistas con mayor popularidad no se correspondían con los artistas más centrales de la red.

Con respecto a la ruptura de la red según los géneros musicales se encontró en la mayoría de casos que la ruptura por géneros musicales es más disruptiva que la remoción aleatoria de nodos, indicando que los artistas de un mismo género musical poseen cierto rol en el mantenimiento de la conectividad de la red. Sin embargo, ninguno de los géneros alcanzó a romper en mayor medida la red comparado a las centralidades analizadas.

Por otra parte, se estudió la detección de comunidades en la red a partir del método de intermediariedad para estudiar si existía una relación entre las comunidades de la red y los géneros musicales. Para esto se forzó al algoritmo a hallar la misma cantidad de comunidades que géneros musicales. Como sobre aproximadamente un 47% de los artistas no se conocía sus géneros musicales, se los calculó a través de un proceso de difusión de etiquetas. De esta manera se obtuvieron los géneros musicales de todos los artistas de la red que se utilizaron para comparar con la partición en comunidades.

Analizando la matriz de correlación entre las comunidades y los géneros se encontró que no existe una correspondencia uno a uno para estos, ya que si bien se encontraron comunidades que están mayormente representadas por un género, como la comunidad de trap, la mayoría está compuesta por artistas que pertenecen a géneros musicales variados que no comparten similitudes o un género es predominante en varias comunidades, como el rock.

Finalmente se buscó predecir futuras colaboraciones a partir del estudio de la similaridad entre artistas no enlazados. Para ello, se paso a un espacio métrico con *Node2Vec* y se encontró que artistas que colaboraron resultan mas similares entre sí que artistas que no. Con ello se estableció una similaridad límite para artistas que aún no colaboraron para determinar si lo harán futuramente, y se ejemplificó para cuatro artistas. Se encontraron futuras colaboraciones razonables de esta manera, e incluso algunas ya existentes pero que por falencias de la red no están presentes.

Adicionalmente, se sugiere para futuras líneas de trabajo:

- Utilizar la información temporal de los enlaces para estudiar la red como una red que evoluciona en el tiempo.
- Cuantificar la precisión del método para predecir enlaces futuros removiendo enlaces de la red e intentar que la similaridad los prediga.
- Profundizar en el estudio de comunidades saliendo de la hipótesis de que se corresponde con la cantidad de géneros.
- Estudiar el coeficiente de clustering en comparación a una hipótesis nula por género musical, para estudiar si artistas que comparten un género musical tienden a formar triángulos de colaboraciones.

Apéndice

En la página <https://github.com/nitrox639/Redes-complejas> se encuentra el repositorio utilizado durante este trabajo tanto como para crear la red como para analizarla, para que pueda ser utilizado para futuros análisis. También se encontró otro repositorio trabajando sobre red de colaboración de artistas internacionales que puede ser de utilidad: <https://github.com/akashlevy/Musical-Collab-Graph>.

En las figuras 17 y 18 pueden verse las nubes de palabras para todas las comunidades detectadas con el algoritmo de intermediariedad. A simple vista la mayor cantidad de comunidades es de algún tipo de rock, combinado con otros géneros.



Figura 17 Nubes de palabras con los géneros de cada comunidad por intermedietez.



Figura 18 Nubes de palabras con los artistas de cada comunidad por intermedieatz.

Referencias

- [1] Albert-László Barabási et al. *Network Science*. Cambridge university press, 2016.