

Práctica 4:

Caso Práctico de Análisis y Evaluación de Redes

Linqi Zhu
X6300759R
zhulinqi@correo.ugr.es

Curso 2022-23
Cuarto Curso del Grado en Ingeniería Informática



**UNIVERSIDAD
DE GRANADA**



Índice

1. Introducción y descripción de objetivo	3
2. La estructura diseña para la red	3
3. Tabla de medidas estudiada y análisis de la red	3
4. Visualización de la centralidades y análisis	5
5. Análisis de comunidades	8
6. Visualización de la red compleja	9
7. Bibliografía	10

1. Introducción y descripción de objetivo

La pregunta de investigación planteada es: *¿Cuáles fueron los asuntos más citado en MAMA (mnet Asian Music Awards) durante los días de celebración en la plataforma Weibo?*

Para ello, utilizamos código de scrapping en Python del un repositorio encontrado por internet. Lo que hace es un scrapping de 1000 comentarios aleatorios que llevan la etiqueta #MAMA# durante los días 29 y 30 de noviembre de 2022.

Una vez obtenido los comentarios se realiza filtrado para obtener los key-word de los comentarios que serán los nodo de la red.

Para obtener las aristas con el calculo la matriz de co-ocurrencia que indica el grado de relación entre dos key-word.

2. La estructura diseña para la red

Los elementos y características son:

- Nodos : los key-word de los comentarios
- Enlaces: si aparecen en el mismo comentarios esta relacionados
- Pesos: número de veces que aparecen en el diferente comentarios

Es una red es una red no dirigida y ponderada.

Red Original representado con los algoritmo Fuchterman Rengold :

3. Tabla de medidas estudiada y análisis de la red

A. Tabla con las medidas de la red:

Número de nodos N	3237
Número de enlaces L	50269
Número máximo de enlaces Lmax	263282178354
Densidad	$1.909 * 10^{-7}$
Grado medio	31.059
Diametro maximo	5

Distancia media	2.263
Coeficiente medio de clustering	849
Nº componente conexas	9
Nº nodos componente gigante Ncg	3222
% de nodos componente gigante %Ncg	99,54
Nº aristas componente gigante Lcg	50241
% de aristas componente gigante %Lcg	99,944

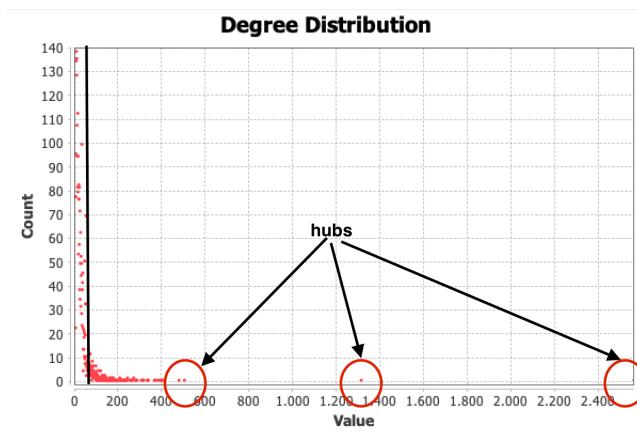
B. Análisis de la red:

Podemos ver que es una red bastante densa con muchos conexiones, y una distancia media pequeña.

Valores de las medidas: $N=3237$ nodos, $L=50269$, Densidad= $1.909 * 10^{-7}$

Grado medio : 31.059 -> un key-word interactúa con otras 31-32.

Distribución de grados:



La distribución de grado indica que hay un nivel de interacción alto. Un 73,8% de los nodos tienen grado menor que 32($k_i < \langle k \rangle$)

Existen unos pocos nodos fuertemente conectados (hubs), el mayor con grado 5558.

Consecuencia de la propiedad libre de escala, muy común en redes reales, presentan una distribución de grados de larga estela o largo cola.

Diámetro: $d_{\max} = 5$. Las distancias entre nodos tiene una media de 2.263.

Los nodos no tiene distancia grande, consecuencia de la propiedad de mundo pequeño (small-world), común en redes reales.

Conectividad: la red presenta 9 componentes conexas, 4 de ellas están formados key-word aislado que no interactúa con el resto.

El que queda forma un componentes gigante que agrupa el 99,54 % de los nodos: 3222 de 3237. Las componentes aisladas son muy pequeña, esta fragmentación también es común en redes reales

Distribución de coeficiente de clustering:

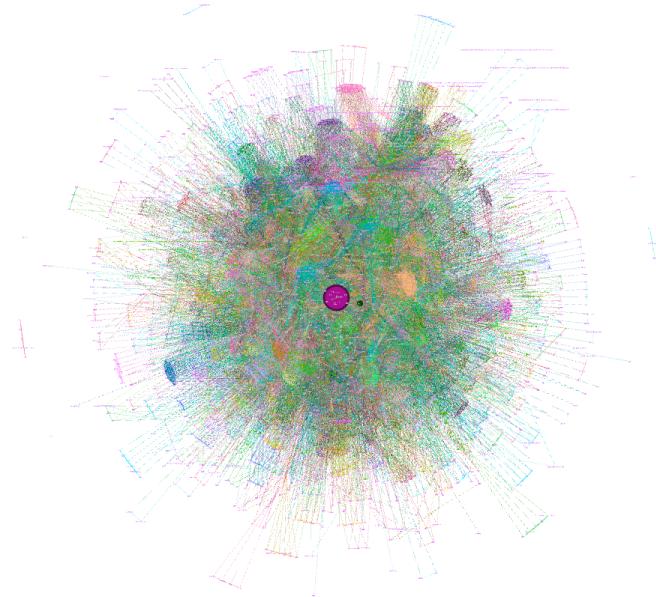


Coeficiente de clustering medio: $\langle C \rangle = 0.849$, es bastante alto, indicando un grado significativo de clustering local y muchos hubs interconectados unos grupos de nodos con otros. Consecuencia de la jerarquía de redes.

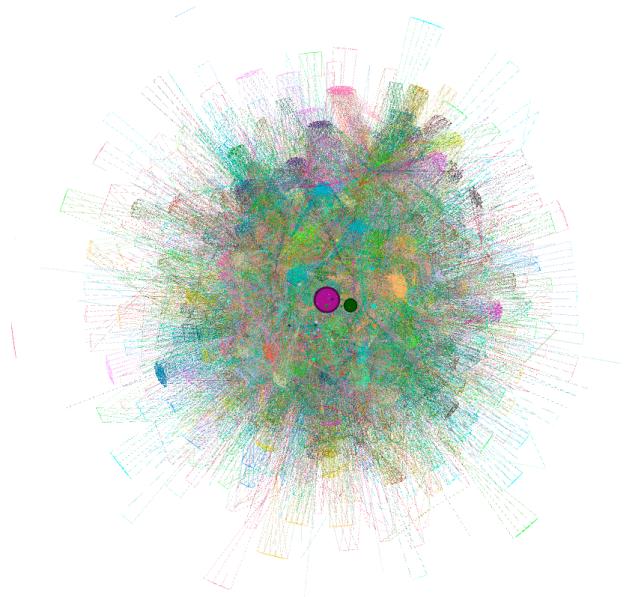
No obstante, aparece algunos aislados, lo cual podríamos hacerlo relacionar erróneamente las bajas distancias con el alto clustering.

4. Visualización de la centralidades y análisis

- Centralidad de grado: nos indica la importancia de cada key-word según el entorno inmediato. Los tres key-word más importantes son MAMA, Escenario y Cooperación



B. Centralidad de Intermediación: favorecen a los nodos por los que pasan muchos caminos mínimos entre otros nodo. Los nodos son: son MAMA, Escenario y Cooperación.



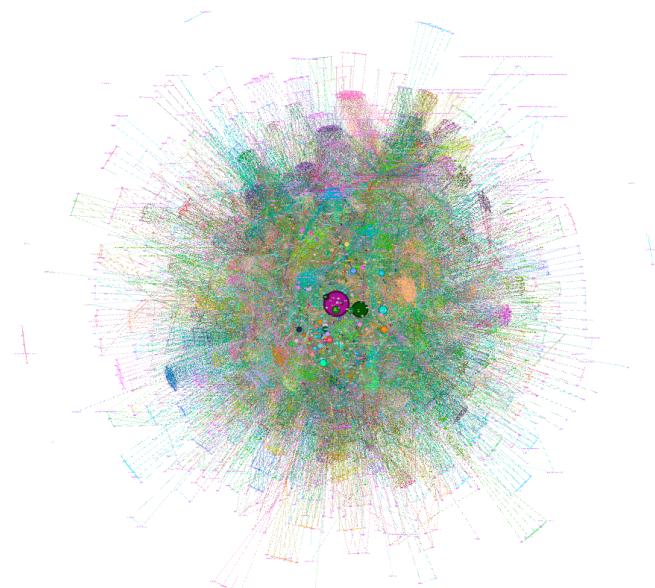
Los hubs tienen facilidad para obtener valores altos en la intermediación.

C. Centralidad de Cercanía: valora la capacidad de estar bien situado en la red, que las distancias a los demás nodos sean más corto.



Podemos ver que hay algunos nodo que están alrededor del componente gigante, son los componentes conexos, al estar aislados y menos nodo conectado tiene un valor alto. Podemos ver que casi todos los nodos tienen el tamaño parecido o igual, significa que no varia mucho el valor de cercanía de los nodos.

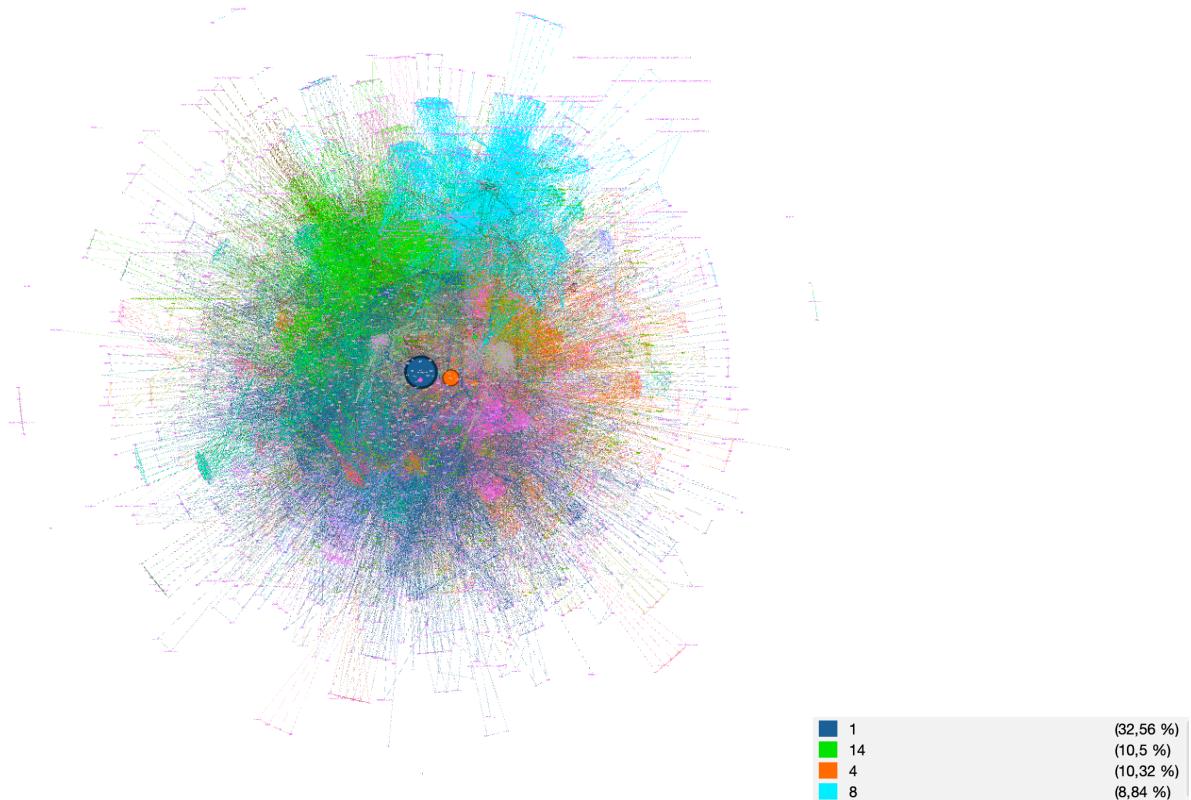
D. Centralidad de vector propio: que key-word son más "importantes" dependiendo de sus vecinos.



5. Análisis de comunidades

1. Detección de comunidad con la determinación de la estructura modular:

Aparece 28 comunidades diferentes. Cada comunidad se trata de asunto diferentes.



Las cuatro comunidades más destacadas:

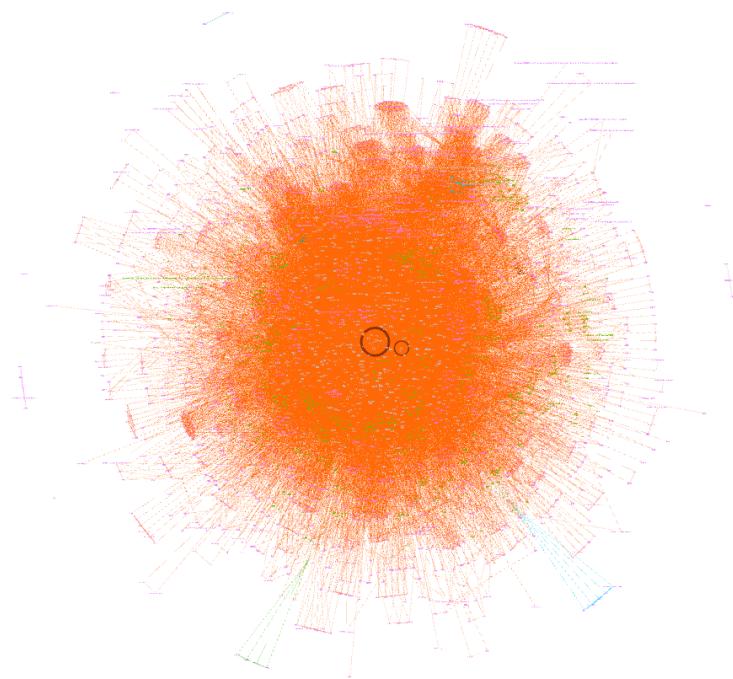
Modularity Class = 1 : Comentarios relacionados con los artistas que han actuado en la noche, key-word sobre forma de actuación

Modularity Class = 14 : Comentarios relacionados con los artistas que han obtenido el premio más importante

Modularity Class = 4 : Comentarios relacionados con una actuación colaborativa de la noche, fue la actuación con más visualizaciones

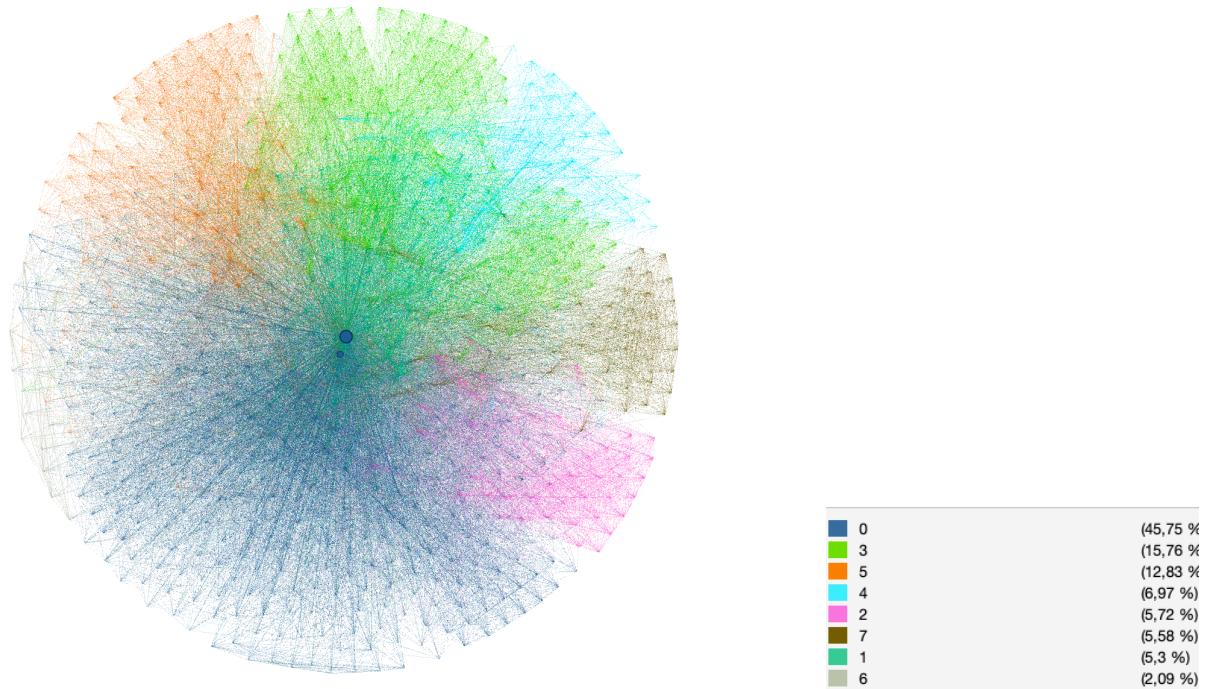
Modularity Class = 8 : Comentarios de diferentes cuentas de fandom que comparten video de escenarios, fotos, noticias relacionadas a su artista

2. Detección de comunidad con Girvan-Newman:



Gráfica obtenida después de 10 minuto de ejecución, el método de Girvan-Newman. No es muy significativo comparando con el método anterior, el mayor comunidad esta formada por el 99,2% de nodos.

6. Visualización de la red compleja



Para obtener una visualización mejor se elimina todo los nodos que tiene grado con peso menor que la media de ellos.

Podemos con claridad las diferentes comunidades que forma, y podemos responder la pregunta planteada al principio.

Los diferentes asuntos (comunidades) tratados según porcentaje de mayor a menor elemento son:

- Modularity class = 0: Los artistas que han realizado actuación en la ceremonia.
- Modularity class = 3: Comentario de lugar de ceremonia , red carpet, salida de aeropuerto etc ...
- Modularity class = 5: El artista que se llevo el premio mas importante de la ceremonia (BTS).
- Modularity class = 4: comentarios de cuentas de fandom que comparte videos, fotos ,noticias relacionado
- Modularity class = 2: comentarios de grupos Rookies del año
- Modularity class = 7: comentarios de quien puede obtener los premios
- Modularity class = 1: comentarios de recordatorio de actuaciones y artistas de años anteriores.
- Modularity class = 6: compara un poco la situación en China en este tema.

7. Bibliografía

<https://github.com/eastmountyxz/Wuhan-data-analysis/blob/master/blog05-knowledge-graph>

<https://github.com/dataabc/weibo-search>

<https://www.mwave.me/en/mama>

<https://s.weibo.com/topic?>

[q=mama&pagetype=topic&topic=1&Refer=weibo_topic](https://s.weibo.com/topic?q=mama&pagetype=topic&topic=1&Refer=weibo_topic)

