

Gestión de Información en la Web



UNIVERSIDAD DE GRANADA

PRÁCTICA 3: ANÁLISIS PRELIMINAR Y VISUALIZACIÓN BÁSICA DE UNA RED SOCIAL CON GEPHI

Autor

Juan Manuel Castillo Nievas



MÁSTER PROFESIONAL EN INGENIERÍA INFORMÁTICA 2020/2021

—
Granada, 24 de abril de 2021

Índice

1. Introducción	2
2. Visualización de la red	3
3. Valores de las medidas	4
4. Gráficos	5
5. Análisis	11
6. Bibliografía	12

1. Introducción

El objetivo de esta práctica es conocer y aprender a usar una herramienta de análisis y visualización de redes, así como familiarizarse con los procedimientos de análisis de redes y sus medida. Como herramienta se ha usado **Gephi** [1].

Como conjunto de datos se ha utilizado el **GEMSEM Facebook** [2] publicado en la página de **SNAP** [3], el cual es una colección de datos de las páginas de Facebook en noviembre de 2017. En los **nodos** se representan las páginas de Facebook verificadas y la unión de dos nodos indica que ambas páginas se han dado *me gusta* mutuamente, con lo cual estamos ante un **grafo no dirigido**. Este conjunto de datos incluye 8 categorías de páginas de Facebook, y para este documento se ha elegido concretamente la categoría de **TV Shows**, la cual está formada por **3.892 nodos** y **17.262 aristas**.

El conjunto de datos se llama *tvshow_edges.csv*, con lo cual se debe importar de la forma en la que se indica en la Figura 1.

Pasos

1. Opciones generales de CSV
2. Parámetros de importación:

Opciones generales de CSV (1 de 2)

Archivo CSV a importar:

C:\Users\jumac\OneDrive\Escritorio\P3 Gephi\tvshow_edges.csv

Separador: Coma Importar como: Lista de adyac... Conjunto de caracteres: UTF-8

Previsualización:

node_1	node_2
0	1838
0	1744
0	14
0	2543
1	1009
1	1171
1	1465
1	2069
...	...

< Atrás Siguiente > Terminar Cancelar Ayuda

Figura 1: Importación de la red a *Gephi*

2. Visualización de la red

En la Figura 2 se muestra la visualización de esta red. Para una mejor calidad de visualización de la red, se ha usado el algoritmo de *layout* conocido como **Force Atlas 2**, con un **escalado de 40.0** y activando la opción de **gravedad más fuerte**.

La visualización de la **componente gigante** es exactamente igual, pues solamente hay **1 componente conexa**.

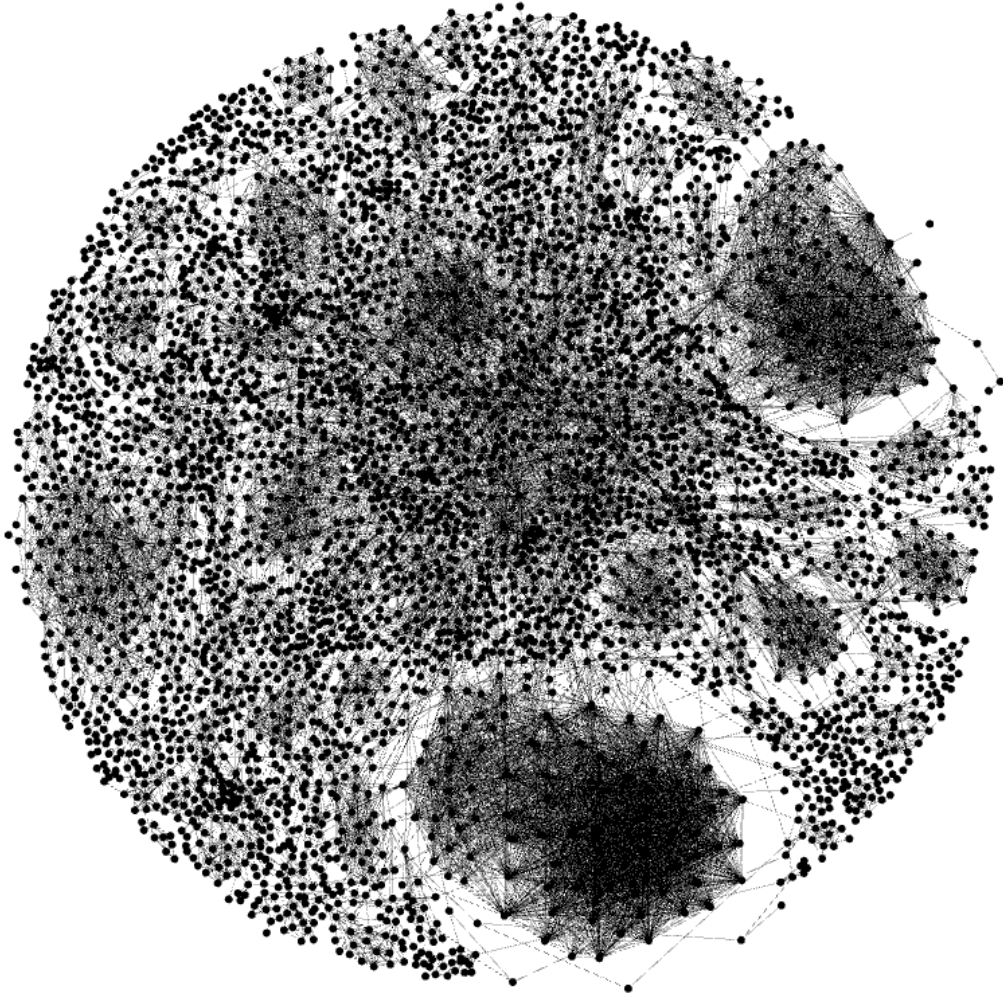


Figura 2: Visualización de la red

3. Valores de las medidas

En la Tabla 1 se muestra la tabla Excel obtenida con las medidas incrustadas.

Medida	Valor
Número de nodos N	3.892
Número de enlaces L	17.262
Número máximo de enlaces L_{max}	7.571.886
Densidad del grafo L/L_{max}	0,002
Grado medio $\langle k \rangle$	8,871
Diámetro d_{max}	20
Distancia media d	6,276
Coefficiente medio de clustering $\langle C \rangle$	0,445
Número de componentes conexas	1
Número de nodos componente gigante (y % %)	3.892
Número de aristas componente gigante (y % %)	17.262

Tabla 1: Valores de las medidas

4. Gráficos

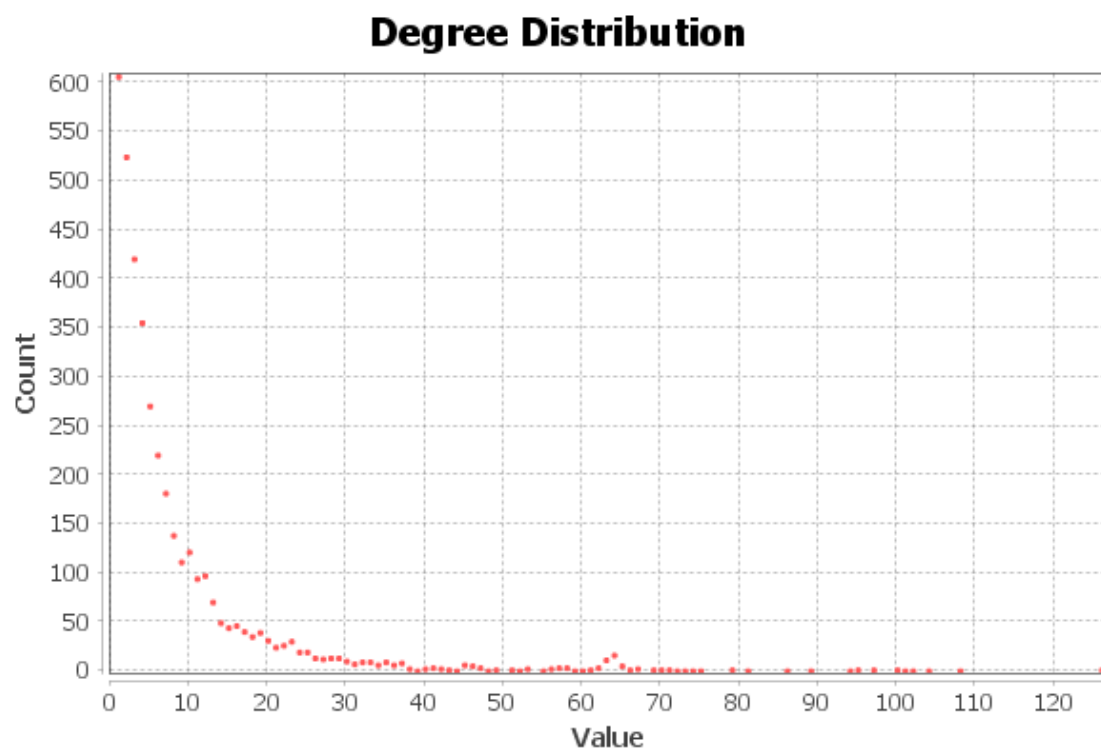


Figura 3: Distribución de grados de la red

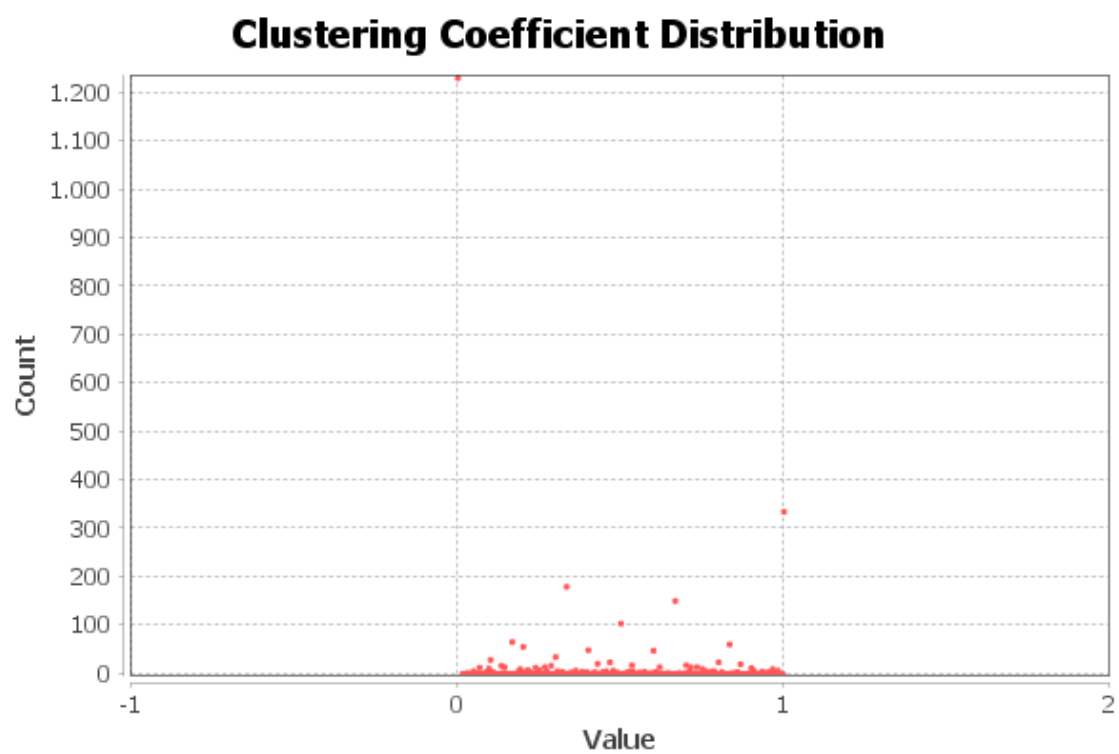


Figura 4: Distribución de coeficientes de clustering de la red

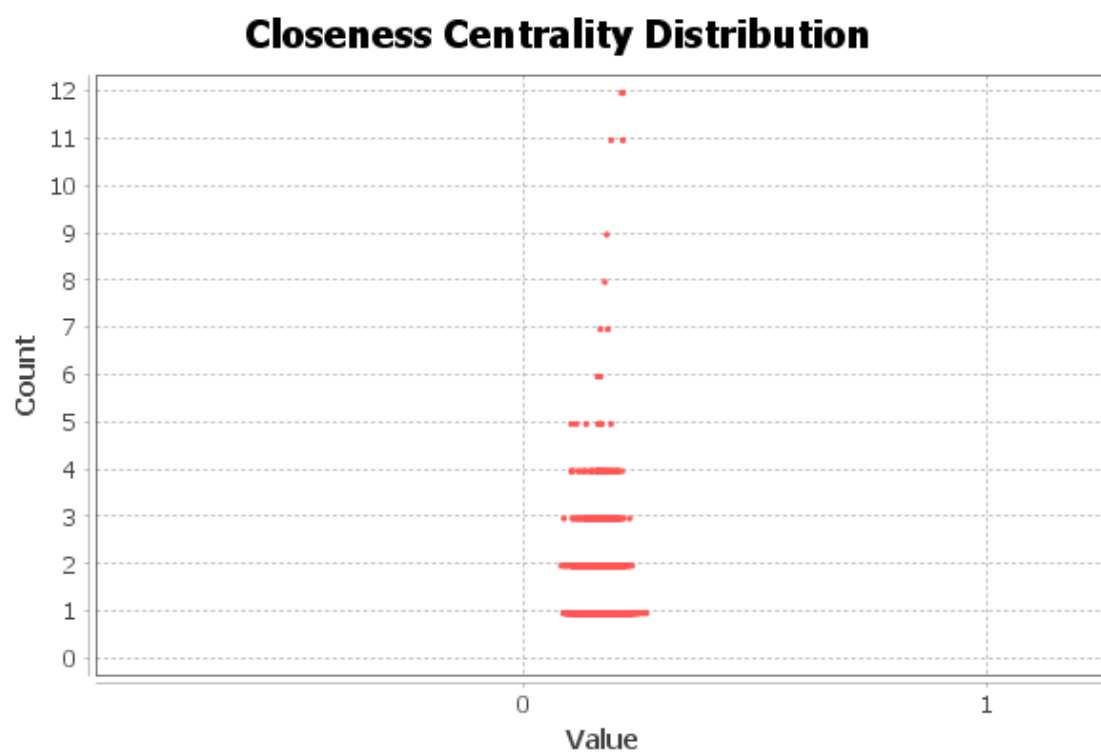


Figura 5: Distribución de centralidad de intermediación de la red

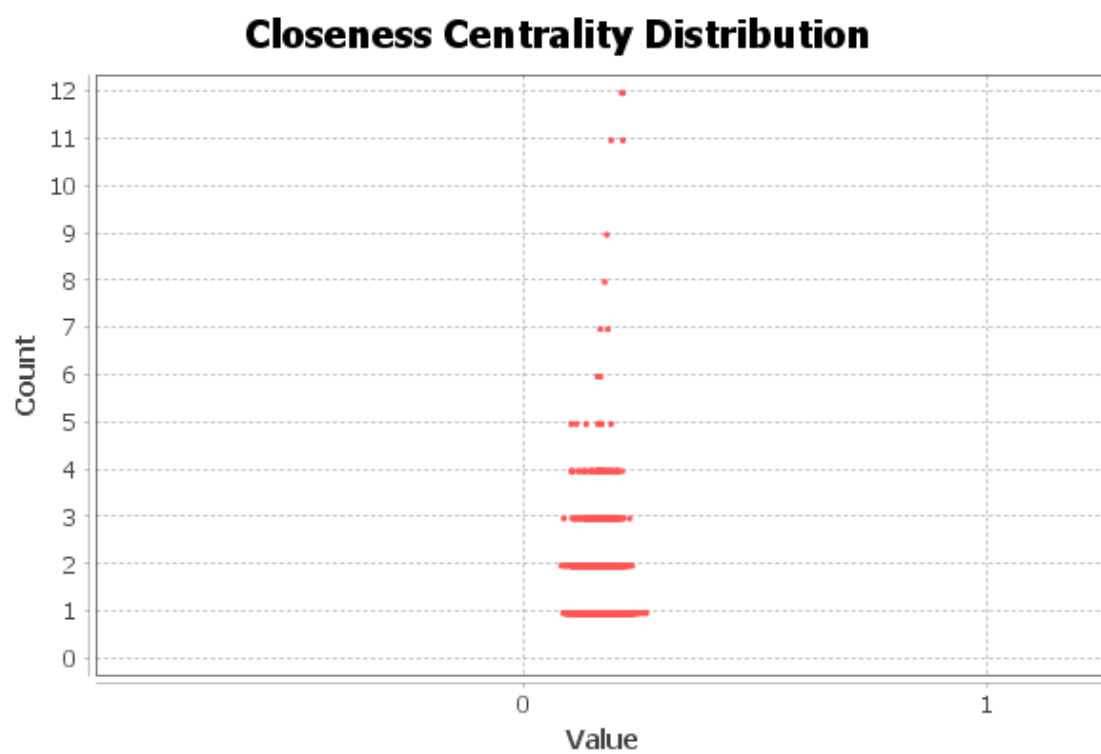


Figura 6: Distribución de centralidad de proximidad de un nodo

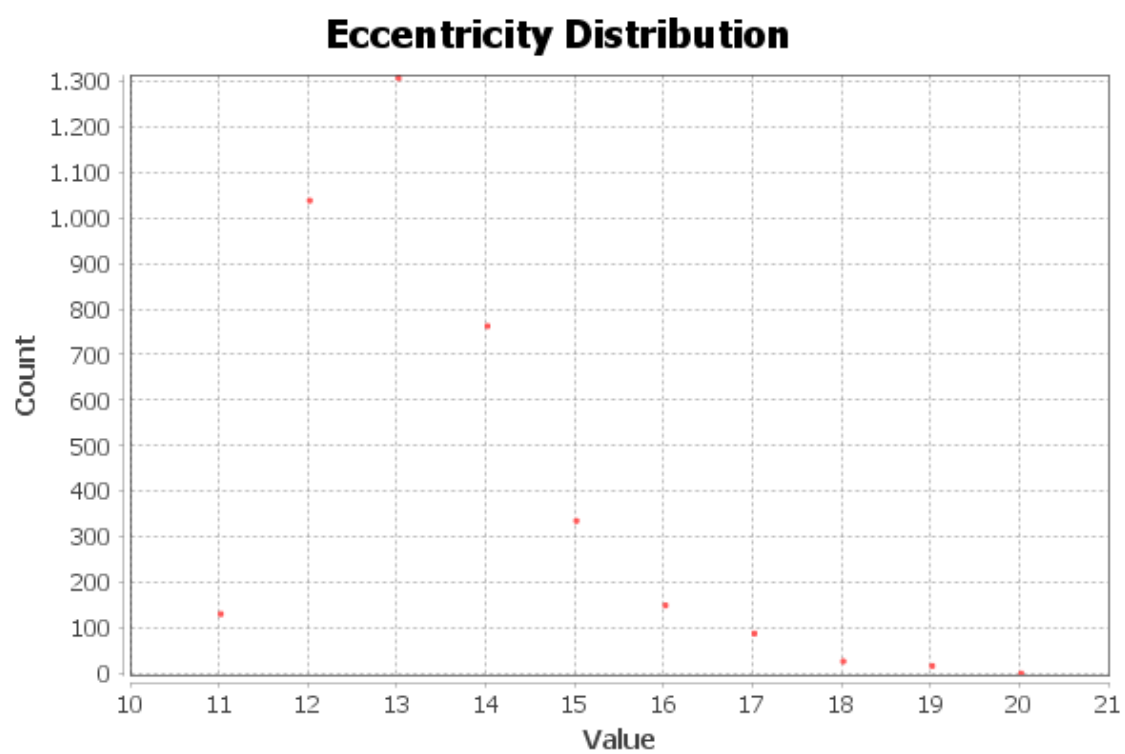


Figura 7: Distribución de excentricidad de la red

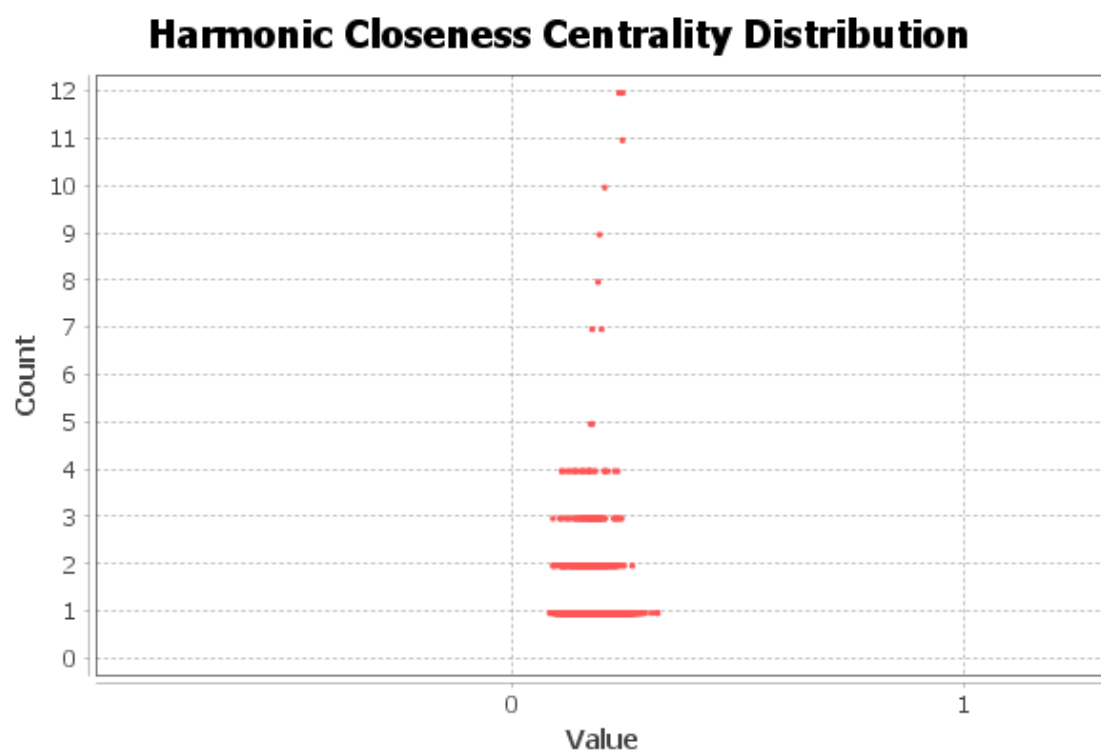


Figura 8: Distribución de centralidad de proximidad armónica de la red

5. Análisis

A continuación se muestran los diferentes puntos realizados para este análisis de la red:

- Tal y como se puede visualizar en la red en la Figura 2, sólo hay una componente conexa. No existen nodos aislados.
- El **grado medio de la red es 8,871**, que indica que cada nodo (página de Facebook) se da *me gusta* mutuamente con casi 9 páginas.
- En la **distribución de grados de la red** en la Figura 3 se puede ver que se da la **propiedad de libre escala**: hay pocos nodos que están altamente conectados. De hecho, sólo hay un nodo que comparte *likes* con más de 120 páginas y la mayoría de los nodos están conectados con menos de 10 páginas.
- El **diámetro máximo es de 20** y la **distancia media es de 6,276**. En **distribución de centralidad de la proximidad de un nodo** en la Figura 6 se ve que realmente no existen distancias entre nodos tan grandes en la red, con lo cual estamos ante un **mundo pequeño**.
- El **coeficiente medio de clustering es de 0.445**, un valor que es alto y que indica que hay una probabilidad alta de que los vecinos de un nodo sean también vecinos entre sí. De hecho, en la visualización general del grafo en la Figura 2 se puede ver que las páginas que están conectadas con pocas páginas están situadas en vecindarios con mucha densidad.

6. Bibliografia

- [1] Gephi 0.9.2. <https://gephi.org/>, 2017.
- [2] Benedek Rozemberczki, Ryan Davies, Rik Sarkar, and Charles Sutton. Gemsec: Graph embedding with self clustering. In *Proceedings of the 2019 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining 2019*, pages 65–72. ACM, 2019.
- [3] Graph embedding with self clustering: Facebook. <http://snap.stanford.edu/data/gemsec-Facebook.html>, 2018.