



UNIVERSIDAD TECNICA
FEDERICO SANTA MARIA

INF480

REDES COMPLEJAS

Tarea 1

Florencia Ramírez
Sofía Riquelme

1. Introducción

En este informe se presenta un análisis de la red 6 de la lista otorgada por el profesor. Esta red proviene de una red social “tipo Facebook” de estudiantes de la Universidad de California en Irvine; los pesos son cantidad de mensajes entre ellos.

2. Datos básicos

2.1. Cantidad de nodos y aristas

La red cuenta con 1899 nodos y 13838 aristas.

2.2. Dibujo

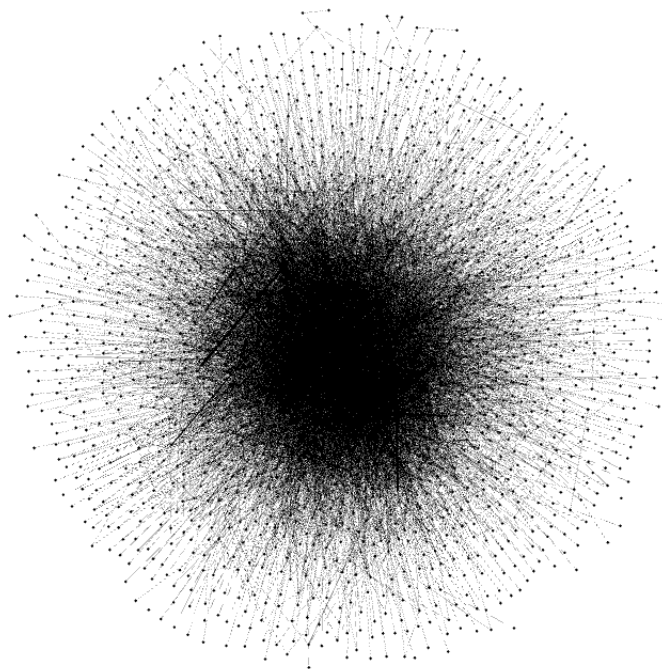


Figura 1: Dibujo de la red

2.3. Conexidad

La red tiene 4 componentes conexas, donde la componente gigante representa un 99,6 % de los nodos.

3. Análisis

3.1. Grados

- Grado mínimo: 1
- Grado máximo: 255
- Gráfico de distribución de grados:

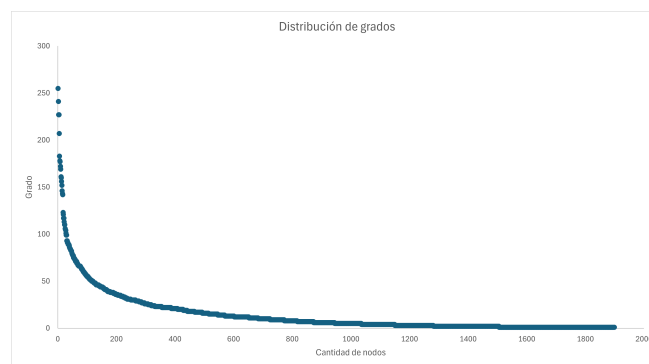


Figura 2: Distribución de grados

Con escala logarítmica:

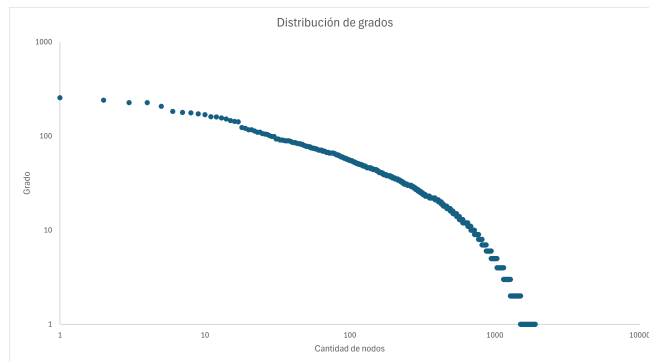


Figura 3: Distribución de grados en escala logarítmica

Se puede observar que si bien se esperaría que fuera una ley de potencia, no lo es.

3.2. Distancias

Determine la distancia promedio entre los nodos, y el diámetro de la red. ¿Se justifica hablar de efecto small world?

La distancia promedio entre los nodos es de 3.055 y el diámetro de la red es 8. Para justificar el efecto small world, la distancia promedio y $\log_k N$, que es igual a 2.813, deben ser del mismo orden de magnitud, como ambos valores son del mismo orden, se justifica hablar de small world.

3.3. Transitividad

Determine el coeficiente de clustering local de la red. ¿Hay transitividad alta?

El coeficiente de clustering promedio de la red es 0.11 y la transitividad de la red es 0.05. Esto indica que no hay un alto nivel de clustering ni tampoco una transitividad alta

3.4. Centralidad

Determine la centralidad de intermediación, PageRank y cercanía de cada nodo. Expórtelas a alguna herramienta de análisis y explore qué tan relacionadas están entre sí, y con el grado del nodo (en particular, haga gráficos de dispersión de grado y PageRank, PageRank e intermediación, etc.). Identifique las mayores desviaciones (nodos con centralidad alta según una métrica y baja según otra) e intente explicarlas en función de la estructura de la red. Obtenga imágenes de la red coloreada según los valores de las distintas centralidades. [Nota: Gephi es un poco críptico. Para calcular intermediación o cercanía hay que decirle que calcule “diámetro de la red”, que no tiene nada que ver.]

Luego de determinar la centralidad de intermediación, cercanía y PageRank de cada nodo, al hacer un gráfico de dispersión de grado y PageRank se obtiene lo siguiente:

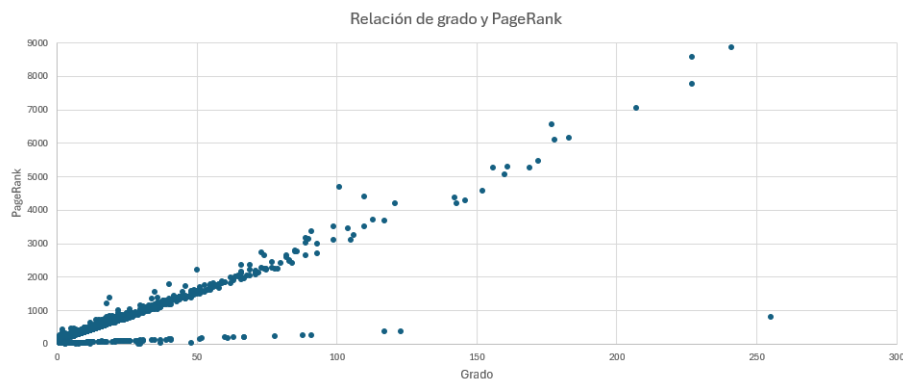


Figura 4: Relación de grado y PageRank

Se puede observar como en la mayoría de los casos, la cantidad de conexiones de un nodo se relaciona con la importancia que tiene este en la red. Lo que no ocurre con una cierta cantidad de estos, que a pesar de tener una cantidad considerable de conexiones, no son considerados como importantes según PageRank, ya que PageRank considera también la importancia de los nodos conectados.

Siguiendo con el gráfico de la relación entre intermediación, o betweenness, de cada nodo y su PageRank se tiene:

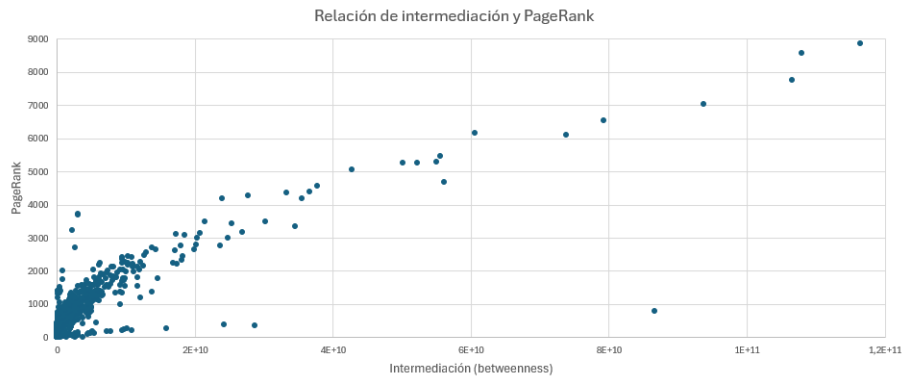


Figura 5: Relación de intermediación y PageRank

La centralidad de intermediación de un nodo nos indica la importancia de un nodo como puente entre otros dos nodos, y se puede observar como los nodos que son considerados menos importantes según PageRank, corresponden a los nodos con menores valores de intermediación, pero también existe un pequeño conjunto de nodos con un mayor valor de intermediación pero son considerados menos importantes.

Finalmente, al graficar la relación entre el grado e intermediación de un nodo se obtiene:

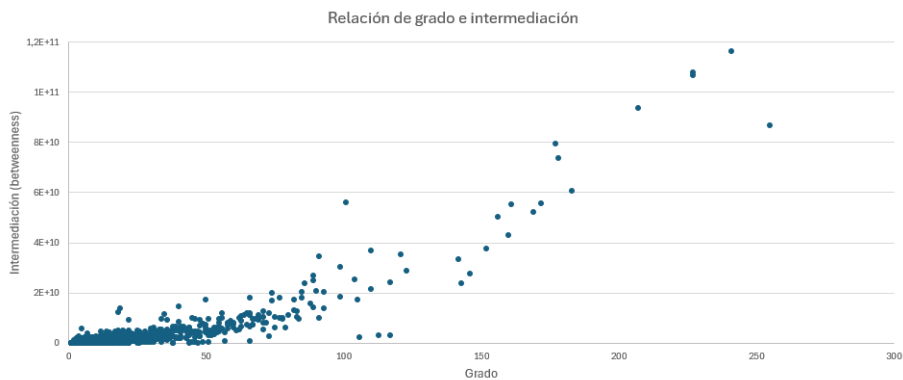


Figura 6: Relación de grado e intermediación

Se puede notar como, para la mayoría de los casos, el grado del nodo

está directamente relacionado con la intermediación de este. Sin embargo, se puede notar la existencia de algunas excepciones, existen nodos con menor grado que tienen un alto valor de intermediación, mientras que algunos nodos con mayor grado no tienen grandes valores de intermediación, lo que no garantiza que al tener una gran cantidad de conexiones se tenga también un gran valor de intermediación.

3.5. Núcleo/Periferia

Determine la profundidad de k -cores (el máximo k para el cual el k -core es no vacío), y la cantidad de nodos en cada k -shell (cada “capa de la cebolla” en k -cores).

La profundidad de los k -cores es 20, y para cada k -core, a continuación se adjunta una tabla con su cantidad de nodos:

| k | Cantidad de Nodos |
|-----|-------------------|
| 0 | 0 |
| 1 | 395 |
| 2 | 228 |
| 3 | 141 |
| 4 | 118 |
| 5 | 100 |
| 6 | 75 |
| 7 | 54 |
| 8 | 64 |
| 9 | 59 |
| 10 | 44 |
| 11 | 57 |
| 12 | 39 |
| 13 | 32 |
| 14 | 63 |
| 15 | 54 |
| 16 | 29 |
| 17 | 53 |
| 18 | 55 |
| 19 | 32 |
| 20 | 201 |

Cuadro 1: Cantidad de nodos en cada k-shell

3.6. Comunidades

Aplique el algoritmo de Lovaina de detección de comunidades (Gephi lo trae por defecto). Si usa parámetros distintos al default, especifique. Reporte la cantidad y el tamaño de las comunidades encontradas, y el valor de modularidad conseguido. Use las comunidades obtenidas para colorear la red. ¿Se aprecia en el gráfico la estructura de comunidades? ¿Interactúan todas con todas, o cada comunidad se relaciona con unas pocas de las demás? [Aquí si hace falta puede jugar un poco más con las opciones de layout, para buscar un buen dibujo de la red.]

Se obtuvieron 15 comunidades distintas, con una modularidad de 0.355. A continuación se presenta una imagen coloreada por comunidad:



Figura 7: Imagen coloreada por comunidad

3.7. Asortatividad

Calcule el coeficiente de correlación de Newman para medir la asortatividad de la red. ¿Le parece asortativa, disasortativa, o ninguna?

El coeficiente de correlación de Newman es -0.18 para esta red. Esto indica que es disasortativa (por ser un valor negativo), sin embargo, no está lo suficientemente cerca de -1 como para ser completamente disasortativa

3.8. Modelo estructural

Genere 10 redes aleatorizadas que compartan la distribución de grados de su red, y para cada una de ellas determine la modularidad (aplicando el mismo algoritmo de Lovaina), el coeficiente de correlación, el coeficiente de clustering local, y la profundidad de k-cores. Compare los valores con el obtenido para su red: ¿pueden explicarse sus valores a partir de la distribución de grados, o parecen ser propiedades específicas de la red? A continuación se presenta una tabla con los valores obtenidos al generar 10 redes aleatorias:

| No. | Modularidad | Correlación | Clustering | k-cores |
|-----|-------------|-------------|------------|---------|
| 1 | 0.2058 | -0.0477 | 0.0665 | 18 |
| 2 | 0.2099 | -0.0412 | 0.0670 | 18 |
| 3 | 0.2054 | -0.0479 | 0.0706 | 18 |
| 4 | 0.2094 | -0.0467 | 0.0723 | 19 |
| 5 | 0.2081 | -0.0528 | 0.0686 | 18 |
| 6 | 0.2036 | -0.0444 | 0.0667 | 18 |
| 7 | 0.2040 | -0.0533 | 0.0694 | 18 |
| 8 | 0.2076 | -0.0547 | 0.0685 | 18 |
| 9 | 0.2066 | -0.0288 | 0.0634 | 19 |
| 10 | 0.2105 | -0.0430 | 0.0690 | 18 |

Cuadro 2: Análisis de redes aleatorizadas

Considerando los valores de la tabla anterior, se puede observar lo siguiente:

- Modularidad: En la red original la modularidad es de 0.355 mientras que en las redes aleatorizadas varían entre 0.2036 y 0.2105. Dado que esta es una variación significativa de la red original, indica que hay características específicas de la red que generan estos valores
- Profundidad de k-cores: En la red original son 20 y en las aleatorizadas varía entre 18 y 19. Estos valores son bastante cercanos entre sí, por lo que sí se podría explicar por la distribución de grados
- Coeficiente de clustering: En la red original el coeficiente es 0.11, y en las redes aleatorias varían entre 0.0634 y 0.0723. Esto sugiere que

el coeficiente de clustering de la red original no puede explicarse simplemente por la distribución de grados.

- Correlación de grado: En la red original se tiene un coeficiente de correlación de -0.18 mientras que en las redes aleatorias varían entre -0.0288 y -0.0547 . La red original tiene una correlación de grado más negativa que las redes aleatorizadas, indicando que la tendencia de nodos de alto grado a conectarse con nodos de bajo grado es más pronunciada en la red original que en las aleatorizadas.