



Universidad  
Internacional  
Menéndez Pelayo

Máster Universitario en Investigación en Inteligencia Artificial

Curso 2020-2021

**Recuperación y extracción de información,  
grafos y redes sociales**

# **Análisis y Visualización Básica de una Red Social con Gephi**

8 de enero de 2021

**Laura Rodríguez Navas**  
**DNI: 43630508Z**

**e-mail:** [rodrigueznabas@posgrado.uimp.es](mailto:rodrigueznabas@posgrado.uimp.es)

## La Red

La red *Diseasome*<sup>[1]</sup> seleccionada para realizar esta práctica es una red no dirigida de trastornos y genes de diferentes enfermedades vinculadas por asociaciones conocidas entre trastornos y genes, que nos indican el origen genético común de muchas enfermedades. La forman 526 enfermedades y 903 genes, donde los genes asociados con trastornos similares muestran una mayor probabilidad de interacciones físicas entre sus productos y una mayor similitud de perfiles de expresión para sus transcripciones, lo que respalda la existencia de distintos módulos funcionales específicos de la enfermedad.

El conjunto de datos de *Diseasome* viene como un archivo *.zip*, que se puede descargar en el siguiente enlace: <http://gephi.org/datasets/diseasome.gexf.zip>. Una vez descargado y descomprimido, obtenemos un archivo *.gexf*, que contiene un archivo de grafos. Importamos el archivo de grafos a *Gephi*<sup>[2]</sup> y comenzamos a probar diferentes opciones de visualización.

Después de probar diferentes visualizaciones encojemos el algoritmo de distribución: Fruchterman Reingold (en la ventana *Distribución*). Para evitar que las componentes conexas queden fuera de la vista principal, fijamos el valor del parámetro *Gravedad* a 20. También marcamos las opciones *Disuadir Hubs* y/o *Evitar el solapamiento*. Esto convierte la visualización en un círculo y coloca la red alrededor de la misma área (ver Figura 1).

De aquí pasamos a la detección de comunidades para colorear los clústers de la red. *Gephi* implementa el método *Louvain* disponible en el panel de *Estadísticas*. Damos clic en ejecutar *Modularidad* y veremos como el algoritmo de detección de comunidades nos ha creado un nuevo parámetro de particionamiento (*Modularity Class*). Si seleccionamos este nuevo parámetro observaremos las comunidades encontradas y si finalmente pulsamos *Aplicar* colorearemos los nodos según las comunidades encontradas. Esto hace que la visualización sea más colorida y se vean bien donde se encuentra cada comunidad.

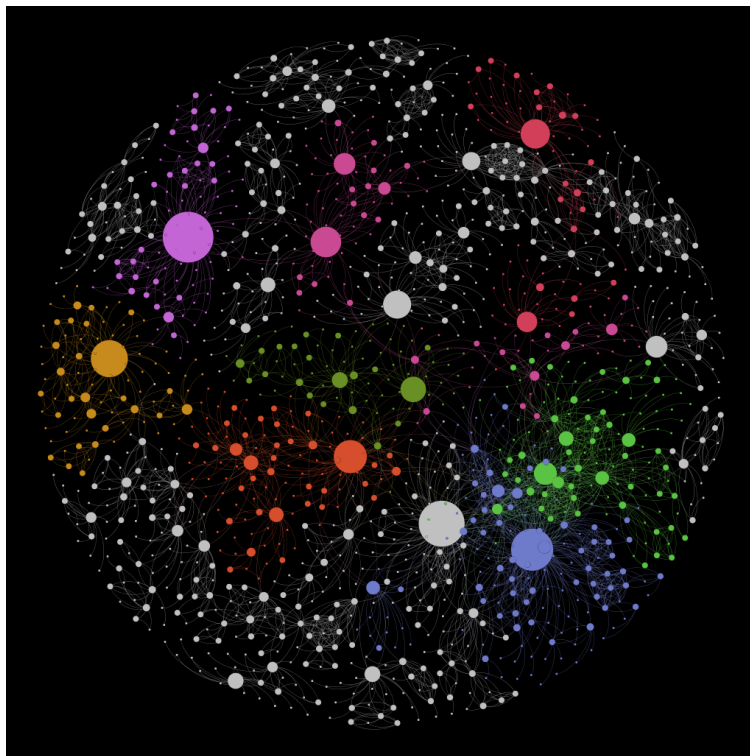
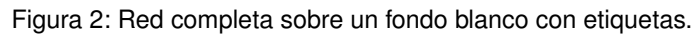


Figura 1: Red completa sobre un fondo negro sin etiquetas.

A continuación, añadiremos las etiquetas a los nodos para ver la red con más a detalle (ver Figura 2).



## Análisis Básico de la Red

Posteriormente, calculamos otra medida global, el grado medio  $\langle k \rangle$ , ejecutando la opción correspondiente en la ventana *Estadísticas*. El valor del grado medio  $\langle k \rangle$  es 5,533. Al realizar el cálculo del grado medio, también obtenemos la distribución de grados de la red completa (ver Figura 3).

3 de 5

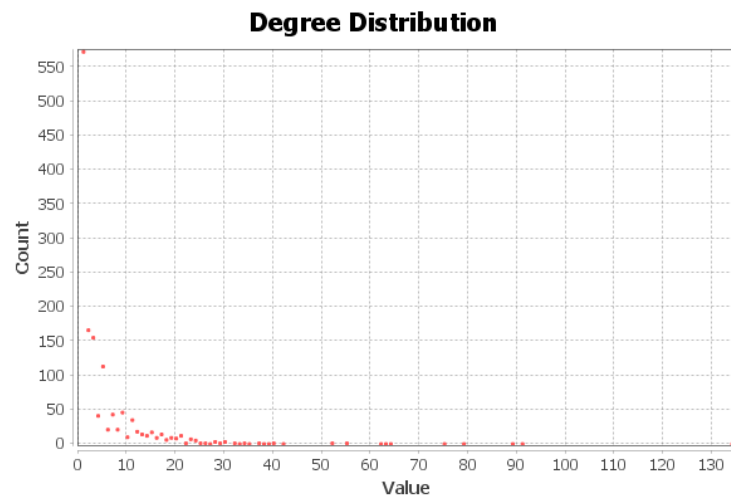


Figura 3: Distribución de grados de la red completa.

## Estudio de la Centralidad de los Actores

## Visualizaciones y Gráficos adicionales

## Bibliografía

- [1] Kwang-Il Goh, Michael E. Cusick, David Valle, Barton Childs, Marc Vidal, and Albert-László Barabási. The human disease network. *Proceedings of the National Academy of Sciences*, 104(21):8685–8690, 2007.
- [2] Mathieu Bastian, Sebastien Heymann, and Mathieu Jacomy. Gephi: An open source software for exploring and manipulating networks. 2009.