#### 1

# Tarea 1 - Gephi

# Claudia Hazard Valdés

21 de Septiembre de 2019.

# 1. Introducción

Se ha seleccionado la red "The human disease network" [1]. Esta corresponde a una red donde los nodos son genes humanos o enfermedades, buscando el origen genético o relación de genes con distintas enfermedades.

Colon cancer is

Figura 1: Imagen de la red "Diseasome" donde los nodos de color rosado corresponden a genes y los verdes a enfermedades.

Esta red es no dirigida, consta de 1419 nodos y 3926 aristas. No cuenta con pesos en las aristas, ademas de ser una red conexa. Cuenta con etiquetas respecto al nombre del gen o enfermedad, un tipo el cual es gen o enfermedad y una clase la cual corresponde a una agrupación

según el tipo de enfermedad (que afecta cierta parte del organismo) o gen en otro caso.

## 2. ANÁLISIS

# 2.1. Diámetro de la red, distancia promedio y distribución de grados.

Se calcula el diámetro de la red, este corresponde a 15, junto con una distancia promedio de 6.782. La distribución de grados se grafica a continuación:

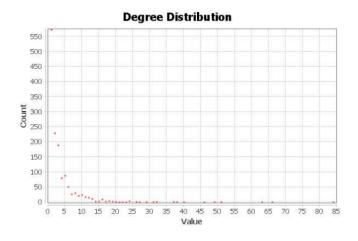


Figura 2: Gráfico de distribución de grados.

Se observa que la distribución se asemeja a la ley de potencia, por lo que se coincide con la mayoría de las redes reales. Esto hace sentido pues existe una gran cantidad de genes y estos solo se conectan a unas pocas enfermedades, además existen ciertas enfermedades relacionadas con muchas otras y muchos genes como el cáncer de colon.

## 2.2. Coeficiente de clustering local

Se calcula el coeficiente de clustering, correspondiente a 0.819, con una cantidad de 2651 triángulos y generando el siguiente gráfico.

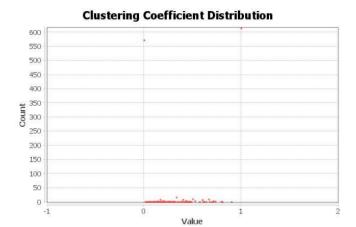


Figura 3: Gráfico de distribución del coeficiente de clustering.

Mostrando un parecido a una distribución normal bastante baja, con 2 outliers que tienen valores inmensamente mas altos de clustering en comparación a los otros. Por lo que cumple con ser libre de escala, algunos nodos tienen muchísimas conexiones pero en su mayoría el grado es bastante bajo.

### 2.3. Modularidad

Se aplica el algoritmo de Lovaina de detección de comunidades, obteniendo 23 comunidades y una modularidad de 0.855.

Se observa que la comunidad mas grande es de aproximadamente 145 nodos mientras que la mas pequeña corresponde a 11 nodos.

Se observan comunidades definidas al aplicar la partición con colores, se muestra primero una vista general de la red con los labels originales de cada enfermedad o gen.

Se observa que para la clase de enfermedades tipo cáncer la mayoría de estas quedaron conectadas en 3 grupos cercanos, mientras que solo 3 nodos quedaron alejados de estas.

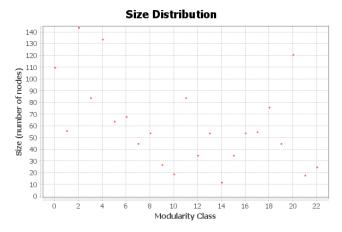


Figura 4: Tamaño para cada una de las comunidades generadas.

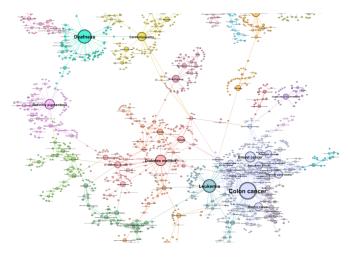


Figura 5: Red coloreada según las comunidades generadas, junto con el nombre de cada enfermedad o gen.

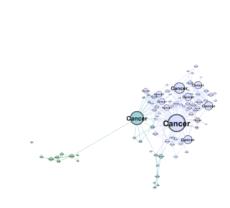


Figura 7: Nodos del tipo cáncer.

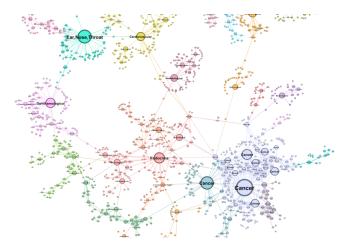


Figura 6: Red coloreada según las comunidades generadas, junto con el nombre de la clase correspondiente.

Además, los tipos de cáncer relacionados a los huesos se encuentran en un mismo grupo junto con enfermedades a los huesos y genes de este tipo.

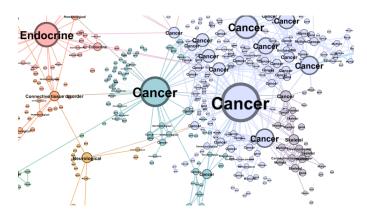


Figura 8: Conexiones en la red, de enfermedades y genes a enfermedades tipo cáncer.

Finalmente, se puede observar que la diabetes se encuentra en el mismo grupo (color) que la obesidad, lo cual tiene sentido con la realidad pues estas enfermedades muchas veces están relacionadas.

# 3. Conclusión

La red "Disasome", cumple con las características de una red compleja, es conexa, cumple con ser una red libre de escala. Es posible apreciar ciertas enfermedades con una gran cantidad de conexiones con otras u otras

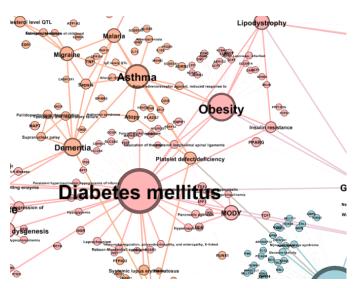


Figura 9: Nodo de Diabetes y Obesidad en la red Disasome.

de cierto tipo especifico por lo que es una relación interesante a investigar. El programa gephi no tiene problemas en procesar y observar una red de 1419 nodos y 3926 aristas.

#### REFERENCIAS

[1] K.-I. Goh, M. E. Cusick, D. Valle, B. Childs, M. Vidal, and A.-L. Barabási, "The human disease network," vol. 104, no. 21, pp. 8685–8690, 2007.