

INSTITUTO TECNOLÓGICO Y DE ESTUDIOS SUPERIORES DE OCCIDENTE

Reconocimiento de validez oficial de estudios de nivel superior según acuerdo secretarial 15018,
publicado en el Diario Oficial de la Federación el 29 de noviembre de 1976.

Departamento de Electrónica, Sistemas e Informática
MAESTRÍA EN SISTEMAS COMPUTACIONALES



REPORTE DEL DATASET “EXPLORING THE WEB AND INTERNET DOMAIN – THE INTERNET DATASET”

Presentan: David Alfonso Velasco Sedano,
Francisco Alejandro Hurtado Rodriguez

Tutora: Mildreth Alcaraz Mejía

San Pedro Tlaquepaque, Jalisco. 18 de Octubre de 2017.

TABLA DE CONTENIDO

1.	Descripción del Proyecto.....	2
2.	Objetivo del Proyecto.....	3
3.	Análisis	4
3.1.	Libro.....	4
3.2.	Análisis Adicional.....	5
3.2.1.	Promedio de Grados y Promedio Ponderado de Grados	5
3.2.2.	Diámetro de la Red	6
3.2.3.	Densidad del Grafo	7
3.2.4.	Modularidad del Grafo.....	7
4.	Conclusiones.....	10
5.	Bibliografía.....	11

1. Descripción del Proyecto

El Internet Dataset es un caso de estudio creado por Mark Newman en 2006 que representa la estructura del Internet a nivel de sistemas autónomos. El conjunto de datos es la reconstrucción de una muestra obtenida a partir de un proyecto llamado The Route Views, impulsado por la universidad de Oregon, cuya finalidad es entender el sistema global de ruteo. La conexión de los sistemas autónomos se gestiona por un protocolo estándar llamado Border Gateway Protocol (BGP).

El protocolo BGP es aplicado en redes de gran escala para conectar sistemas autónomos. Proveedores de Internet cambian sus tablas de ruteo a través de este protocolo. Un sistema autónomo puede ser cualquier elemento que envíe información. En este caso de estudio, los sistemas autónomos son rúters.

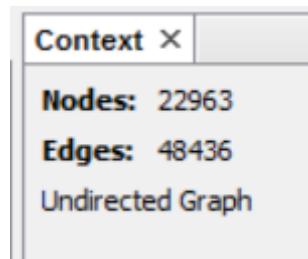
2. Objetivo del Proyecto

El objetivo del proyecto es mostrar y entender la manera en que está formado el sistema de ruteo detrás del Internet en el 2006. Con esto podemos comprender la estabilidad de la red, qué tan pesado es llegar de un rúter a otro e identificar zonas críticas de la red.

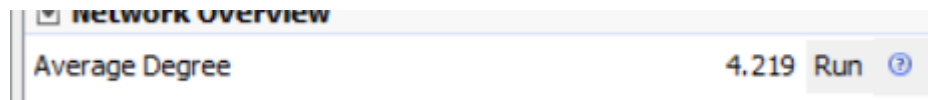
3. Análisis

3.1. Libro

1. Ir al folder del dataset y darle doble click al archivo internet_routers-22july06.GML para abrirlo. Una vez cargado, en el panel de contexto, podemos observar que el grafo contiene casi 23 mil nodos y 49 mil aristas. Haciendo esto un grafo grande:

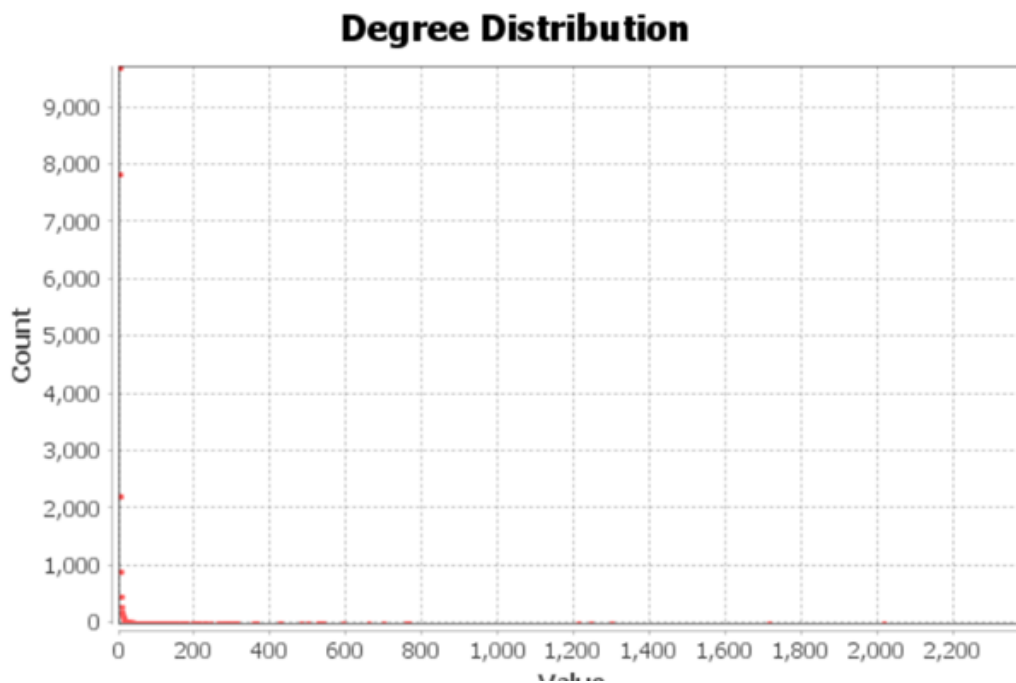


2. En el panel de 'Estadísticas', mandamos a correr la opción de 'Promedio de Grado (Average Degree)'.



Results:

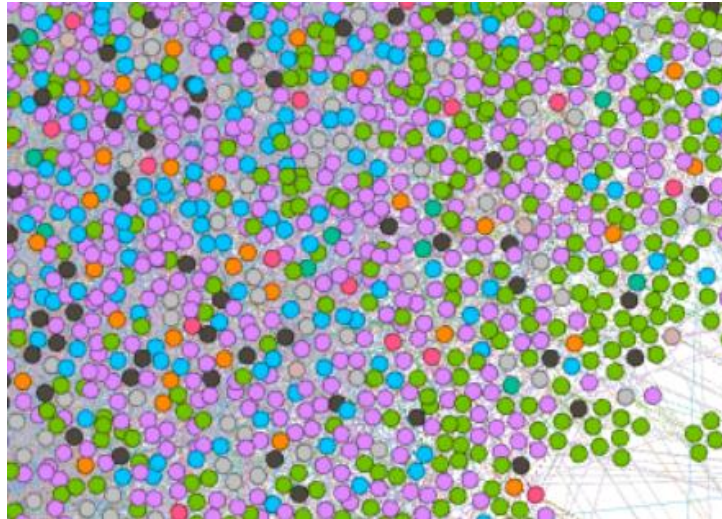
Average Degree: 4.219



3. Una vez completada la ejecución, vamos al panel de 'Partición', damos click en el botón de 'Refrescar'. Esto nos va actualizar el menú con la lista de parámetros de la partición.

4. Seleccionamos del drop-down la opción 'Grado'.

5. Damos 'Aplicar' para ver los nodos con los distintos colores.



6. En el panel de 'Diseño', seleccionamos las opciones habilitadas de 'Force Atlas 2', 'Dissuade Hubs' y 'Prevent Overlap'. Con esto podremos tener una mejor visualización del grafo



3.2. Análisis Adicional

3.2.1. Promedio de Grados y Promedio Ponderado de Grados

Recordemos que el grado promedio se refiere a la relación de cuántas aristas se encuentran en el grafo con respecto a los nodos. En este caso tenemos un total de 4.219.

Average Degree

4.219

Run



Viéndolo con el contexto, significa que por cada sistema aislado dentro de nuestra red debemos esperar ver 4.219 (4) conexiones.

Mientras tanto, el grado promedio ponderado se refiere a cuántas aristas deberíamos esperar en cada vértice. En este escenario tenemos un resultado de 4.219 (4) grado promedio ponderado.

Avg. Weighted Degree

4.219

Run

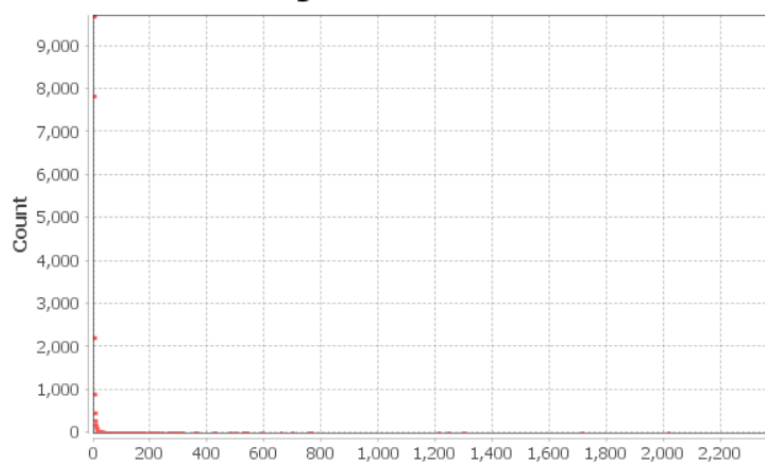


Weighted Degree Report

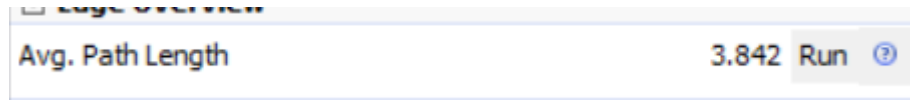
Results:

Average Weighted Degree: 4.219

Degree Distribution



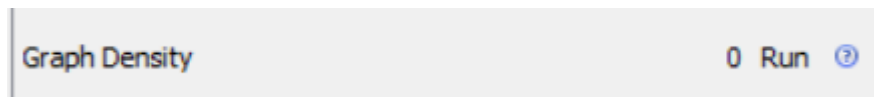
Significando que el camino más largo (yendo por las distancias cortas) simplemente tendremos que recorrer 11 sistemas autónomos para llegar a nuestro destino. Expandiendo esto, calculamos el promedio de distancia por nodo para saber en general cuántos saltos realiza cada nodo para llegar a otro. Para este conjunto de datos vemos que, a partir de cualquier nodo, podemos esperar un promedio de casi 4 saltos para llegar a cualquier sistema autónomo.



Avg. Path Length	3.842	Run	?
------------------	-------	-----	---

3.2.3. Densidad del Grafo

Ahora, para establecer la robustez de la conexión de la red en el 2006, corremos el análisis de densidad del grafo. Esto nos tira un valor de 0



Graph Density	0	Run	?
---------------	---	-----	---

Si hacemos el cálculo por nuestra cuenta utilizando la siguiente fórmula, $GD = (2E / (V * (V-1)))$, obtenemos el valor de:

0.00018372

Esto significa que la red del 2006 no es robusta. Fácilmente podemos perder comunicación con bloques enteros de sistemas autónomos.

3.2.4. Modularidad del Grafo

La modularidad nos permite obtener módulos o comunidades que pueden representar una población, o en este caso, un conjunto de sistemas autónomos. Con una resolución de 1:

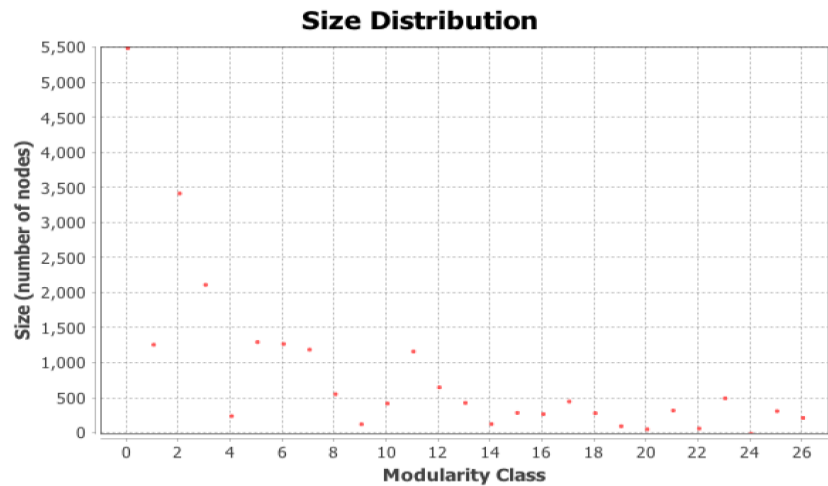
Con una resolución de 1:

Parameters:

Randomize: On
Use edge weights: On
Resolution: 1.0

Results:

Modularity: 0.647
Modularity with resolution: 0.647
Number of Communities: 27



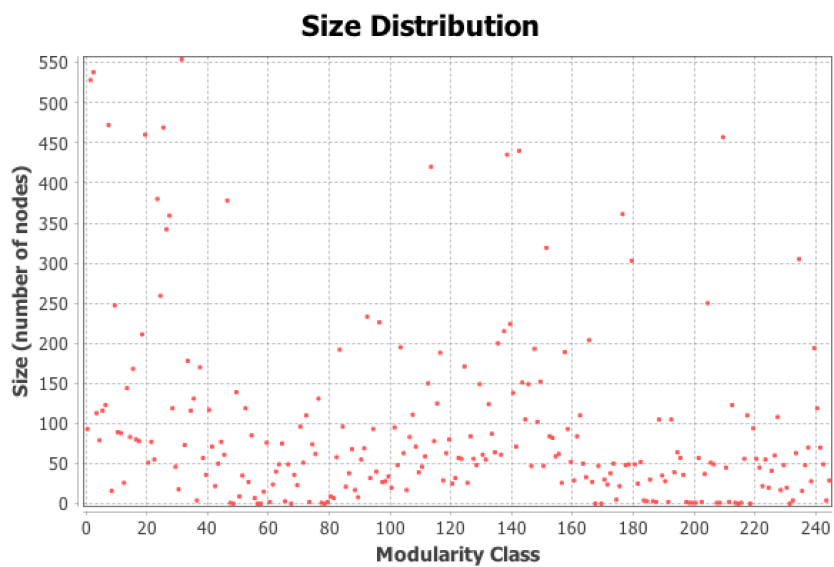
Con una resolución de 0.1:

Parameters:

Randomize: On
Use edge weights: On
Resolution: 0.1

Results:

Modularity: 0.569
Modularity with resolution: 0.047
Number of Communities: 245



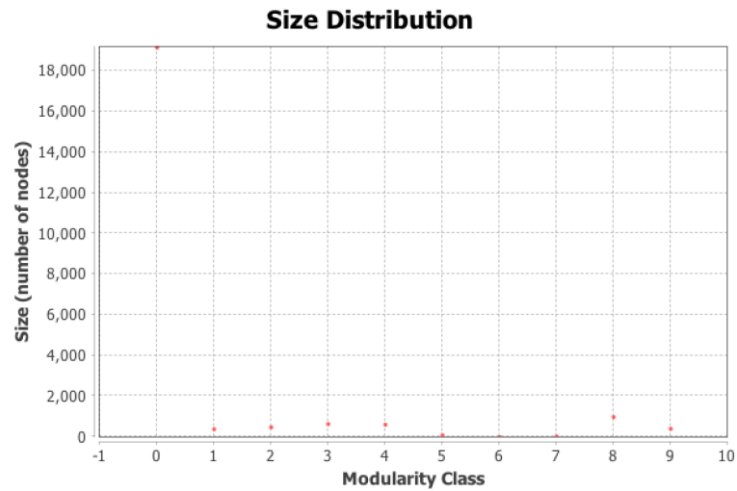
Con una resolución de 5:

Parameters:

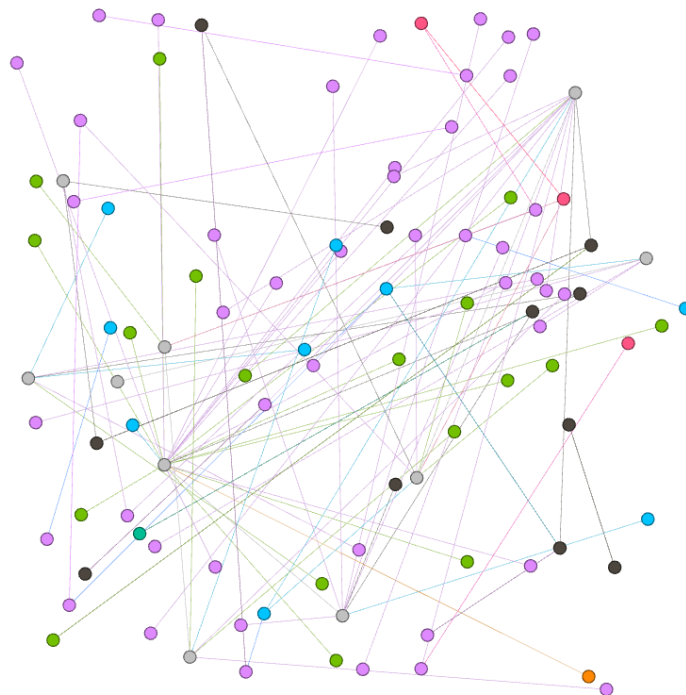
Randomize: On
Use edge weights: On
Resolution: 5.0

Results:

Modularity: 0.208
Modularity with resolution: 4.151
Number of Communities: 10



Las conexiones entre nodos de distintos módulos son altas ya que no existe una modularidad alta. Si se combina con “Betweenness Centrality” podríamos identificar cuáles son los rúters más vulnerables en cada comunidad. De igual modo, si analizamos una comunidad y obtenemos el grado de los nodos, podríamos ver cuales son los rúters más importantes para dicha comunidad. Esto significa que cualquier sub-conjunto es altamente dependiente de los otros sub-conjuntos. Esto se puede comprobar a través del siguiente filtro:



4. Conclusiones

Una vez terminado el análisis, podemos inferir que la red que gestionaba los rúters durante el 2006 es altamente inestable y cada sistema autónomo es fuertemente dependiente de otros sistemas que se encuentran en otras comunidades. Esto muestra condiciones críticas, ya que, si la red sufre de un ataque o falla, puede dejar completamente inhabilitadas secciones dependientes enteras. Asimismo, algunos nodos tienen grados muy altos y también están estrechamente conectados a otros enrutadores de alto grado en el sistema. Luego hay enrutadores con grados en el rango medio y están conectados principalmente a los enrutadores de alto grado o centrales del sistema cuya función principal es actuar como puentes desde los enrutadores escasamente conectados en el sistema hasta el núcleo de la red. Adicionalmente, hay muchos enrutadores que no están muy bien conectados y se sientan en la periferia del sistema. Con la utilización de filtros, podemos extraer las zonas más vulnerables de la red.

Por igual, el salto entre nodos no es muy grande (siendo sus saltos no mayores a 11) y teniendo recorridos promedios de casi 4 saltos. Asegurando que no es muy costoso el intercambio de mensajes a largas distancias.

Finalmente, este análisis puede ser utilizado por administradores de redes dentro de una institución, por proveedores redes móviles, así como por proveedores de Internet.

5. Bibliografía

Khokhar, D. (2015). *Gephi Cookbook*. Birmingham, United Kingdom: Packt Publishing.

Newman, M. E. (24 de Mayo de 2016). *Modularity and community structure in networks*. (N. A. America, Ed.) Obtenido de Proceedings of the National Academy of Sciences of the United States of America: <http://www.pnas.org/content/103/23/8577.full>

Wikipedia. (2 de Noviembre de 2016). *Wikipedia*. Recuperado el 15 de Octubre de 2017, de Wikipedia: https://en.wikipedia.org/wiki/Dense_graph

Wikipedia. (15 de Octubre de 2017). *Wikipedia*. Recuperado el 16 de Octubre de 2017, de Wikipedia: https://en.wikipedia.org/wiki/Border_Gateway_Protocol