Universidad Autónoma Metropolitana

Estimación de la Robustez de Redes de Computadoras sometidas a Errores y Ataques

por

Juan Elier Rosales Rosas
2172040721,

Jorge Isur Balderas Ramírez
2183036000,

Carlos Leonardo Correa Gutierrez
2183036340,

Asesor:
Daniela Aguirre Guerrero

Departamento de ciencias básicas e ingeniería
(Universidad Autónoma Metropolitana)

2021

Índice general

Índice de figuras

Capítulo 1

Marco teórico

1.1. Conceptos básicos de redes

1.1.1. Gráfica

Es una estructura matemática que permite modelar problemas de la vida cotidiana, mediante una representación gráfica formada por un conjunto de nodos que muestra los actores y enlaces que representan los lazos o relaciones entre los actores, además, los enlaces pueden o no tener dirección.

(a) Grafo no dirigido

(b) Grafo dirigido

Figura 1.1: Ejemplos de gráficas

1.1.2. Nodos

Los nodos dentro de una gráfica pueden representar computadoras, proteínas, personas, células, twitts, etc.
Considerando las gráficas de la Fig. 1.1 los nodos están representados por círculos numerados.

1.1.3. Grado

Se refiere a cuantos enlaces tiene un nodo.
Considerando la gráfica de la Fig. 1.1(a) el nodo 2 tiene grado = 3.

1.1.4. Densidad

La densidad de una red es la propiedad que mide la proporción de las relaciones presentes en ella sobre el máximo número de relaciones que pueden existir.

1.1.5. Distribución de grados

La distribución de grados de los nodos en una red viene dado por la función de distribución P(k), que es la probabilidad de que un nodo seleccionado al azar tenga exactamente k enlaces.Una red regular tiene una distribución de grados muy simple, porque todos los nodos tienen el mismo número de enlaces, por lo que su gráfica vendría representada por una distribución de tipo delta.Cualquier aleatoriedad en la red ampliará la forma de este pico, y en el caso límite de una red completamente aleatoria, la distribución de grados sigue una distribución de Poisson, donde la forma de la distribución cae de manera exponencial a medida que nos alejamos del valor máximo, $\bar{k}$.

1.2. Distancia y trayectoria

1.2.1. Trayectoria

- Trayectoria: Se define como una sucesión de arcos distintos que conectan dos nodos.
- Trayectoria no dirigida: Una trayectoria no dirigida del nodo A al nodo B es una sucesión de arcos cuya dirección (si la tienen) puede ser hacia o desde el nodo B.
- Trayectoria dirigida:Una trayectoria dirigida del nodo A al nodo B, es una sucesión de arcos cuya dirección (si la tienen) es hacia el nodo B, de manera que el flujo del nodo A al nodo B, a través de esta trayectoria, es factible

1.2.2. Distancia

Es la cantidad de enlaces que hay entre 2 nodos

1.3. Modelos de redes

1.3.1. Modelo de escala libre

- Modelo de Barabasi-Albert:
  Crecimiento de la red: La red se modela en continuo crecimiento agregando en cada paso un nuevo nodo de grado m, además, tiene enlace preferencial, esto quiere decir que el nuevo nodo se enlaza preferentemente a nodos de alta conectividad.
- Efecto Mateo: El rico se vuelve más rico
  En la evolución de una red de escala libre podemos observar que surgen nodos ricos en enlaces (hubs), los cuales en cada iteración se vuelven más ricos. En contraste, la mayoría de nodos de la red van quedando rezagados en su número de enlaces, es decir, son cada vez más pobres.

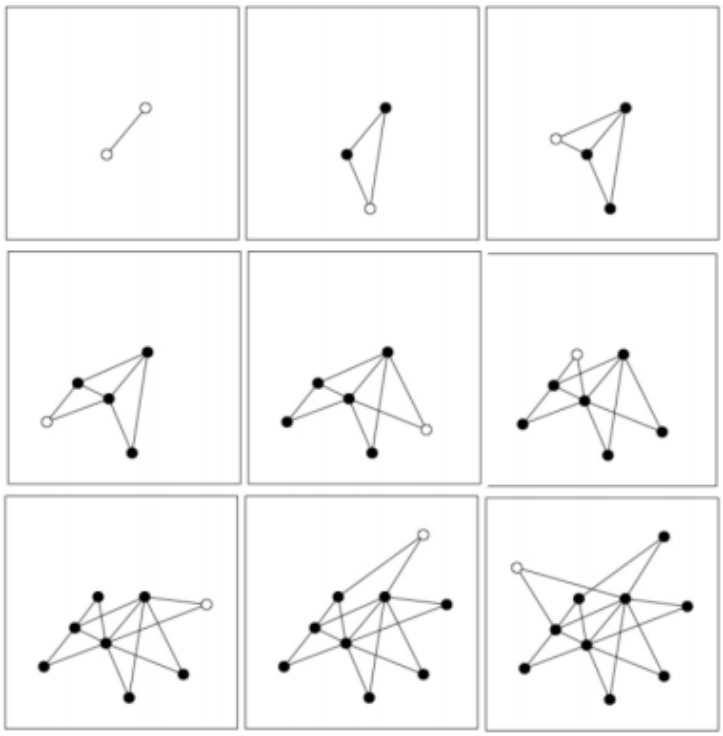

Figura 1.2: Modelo Barabasi-Albert

1.3.2. Modelo aleatorio

- Modelo Erdös-Rényi:
  Es un modelo de grafo $G(N, p)$ no dirigido con N nodos donde cada par de nodos está conectado aleatoriamente con una probabilidad prefijada p .
- Modelo de Wattz-Strogatz o mundos pequeños:
  Este transforma un grafo regular en una red aleatoria al recablear enlaces añadiendo o moviendo los ya existentes.

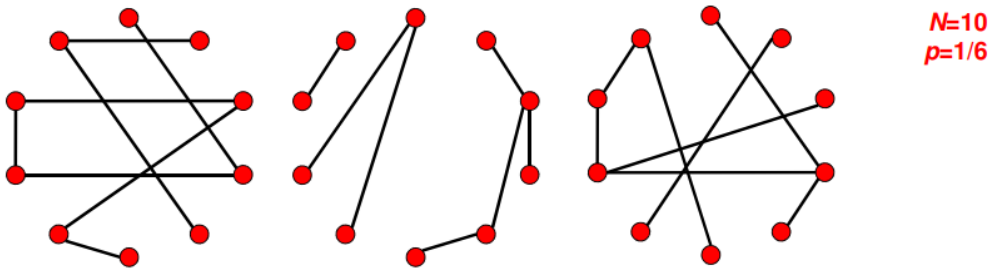

Figura 1.3: Distintas redes aleatorias con mismos parámetros.
Modelo Erdös-Rényi.

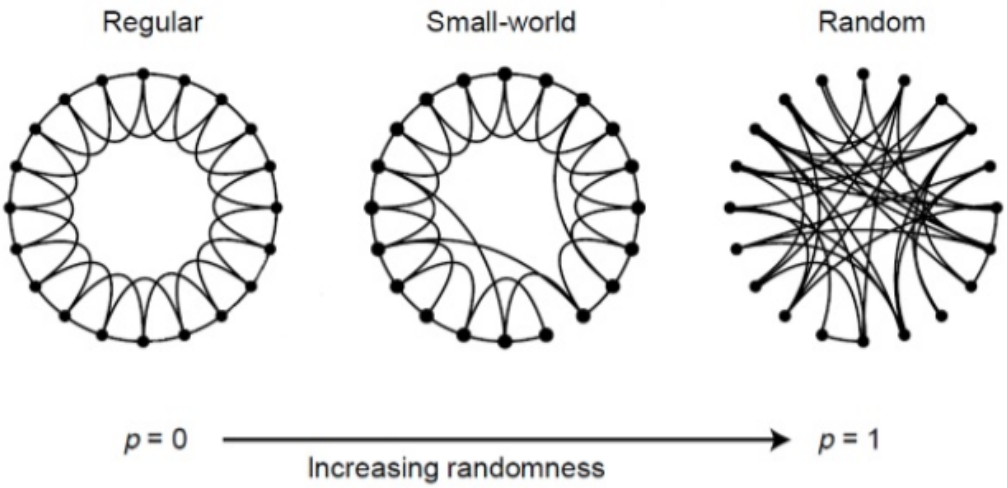

Figura 1.4: Recableado de una red regular hasta llegar a una aleatoria.
Modelo Wattz-Strogatz

1.4. Métricas de centralidad

1.4.1. Centralidad de intermediación/Betweenness Centrality

Los algoritmos de centralidad de intermediación se encargan de calcular la ruta más corta entre cada par de nodos pertenecientes a una gráfica conectada, utilizando el algoritmo de búsqueda de amplitud. Esto es que cada nodo recibe una puntuación en función del número de estas rutas más cortas que pasan por el nodo. Los nodos que se encuentran con mayor frecuencia en estos caminos más cortos tendrán una puntuación más alta. La centralidad de intermediación es utilizada usualmente para encontrar nodos que puedan servir como puente para conectar extremos distantes dentro de un grafo, creando así rutas más eficientes y más cortas.

1.4.2. Centralidad de cercanía/Closeness Centrality

Es una forma de detectar vértices que están dentro de la estructura de un grafo y que tienen la posibilidad de difundir información de forma eficiente.
Este calculo de cercanía de un nodo, también permite conocer su lejanía promedio (distancia inversa) del resto de los nodos, además, los vértices con un puntaje superior de cercanía poseen distancias más cortas que el resto de los nodos que conforman el grafo.

1.4.3. PageRank

Es un algoritmo creado por Google utilizado para asignar un valor a un sitio web y posicionarlo en la misma. El pageRank de un nodo indica que tan probable es que el usuario llegue a el.
Para calcularlo, Google tiene en cuenta varios factores como:

- Visitas totales de la página
- Valor de los contenidos
- Frecuencia de actualización
- Cantidad y valor de los enlaces recibidos

El pageRank de una página puede subir si enlaza a páginas con un pageRank alto.

Capítulo 2

Robustez de redes sometidas a errores y ataques

2.1. Ataques dirigidos por métricas de centralidad

Primero hay que definir el concepto de un error en redes: Un error en un grafo es la eliminación de un nodo con sus respectivos enlaces adyacentes, podemos tener errores secuenciales los cuales son errores que ocurren sucesivamente uno tras otro; los errores simultáneos que son errores que se efectúan al mismo tiempo eliminando más de un nodo a la vez; y los errores en cascada que ocurren cuando la eliminación de un nodo afecta a otro, replicando este mismo comportamiento varias veces más.
Ahora el concepto de ataque hace uso del concepto de error, pues un ataque es la eliminación de un nodo (concepto de error) tomando en cuenta cual es el nodo más importante, para determinar cual es el nodo más importante debemos poder medir algunos atributos de este, a estas métricas se les llama: Métricas de centralidad.
Una vez definido esto podemos hablar de los ataques dirigidos por métricas de centralidad, en pocas palabras son ataques que están programados para eliminar el nodo más importante según alguna métrica de centralidad cómo:

- Intermediación/Betweenness Centrality
- Centralidad de cercanía/Closeness Centrality
- Page Rank
- Grado nodal

2.2. Errores aleatorios

Retomando el concepto de errores aleatorios nos referimos a errores que son ejecutados sin alguna métrica, sencillamente se elimina un nodo al azar de la red y se vuelve a evaluar los datos de esta para ver que tanto desconecto la red. Los errores aleatorios suelen ser eficaces para las redes aleatorias pues recordando su concepto son redes donde generalmente los nodos se encuentran en el mismo rango respecto a su grado nodal. Mientras que este tipo de errores no son muy eficaces para redes de escala libre pues en estas la mayor parte de los nodos tienen un grado nodal bajo y muy pocos nodos tienden a tener un grado nodal alto, así que sin una métrica para determinar cuál sería el nodo óptimo para atacar es poco probable que logre causar una desconexión muy alta.

2.3. Métricas de robustez

Ahora debemos definir que es la robustez en una red, con esto nos referimos a que tan resistente es a los ataques la red, el principal indicador de esto es que tan rápido se degrada la red, es decir, si tenemos una alta robustez tendremos una red resistente a ataques, que se degradará lentamente y requerirá de más ataques para desconectarse. En caso opuesto, si tenemos una red con una robustez baja es una red que, con una menor cantidad de ataques se comenzará a fragmentar y desconectar rápidamente. Tenemos varias métricas para medir esto pero nosotros nos enfocaremos en dos indicadores principales para medir la robustez:

- Cantidad de componentes conectados: Los componentes conectados se refieren a las redes que están desconectadas del componente gigante pero siguen formando parte del grafo. Cuando hablamos de la cantidad de componentes conectados como métrica de robustez nos referimos a medir su incremento, cuando incrementa esta métrica nos está diciendo que la red se esta partiendo, en este caso, el crecimiento del número de componentes conectados es un indicador negativo para la robustez de la red, pues si crece muy rápidamente podemos tener un importante indicador de que es susceptible a ataques.
- Proporción del componente gigante: Esta es otra métrica para medir la robustez de la red, el componente gigante se refiere al sub grafo más grande de la red, si tenemos una red completa tendríamos una proporción del componente gigante igual a '1' esto nos da a entender que el componente gigante es toda la red, pues no hay sub grafos. Entonces, si la red sufre ataques la proporción del componente gigante se irá reduciendo, esto es un indicador negativo de la robustez, pues si el componente gigante se reduce muy rápidamente nos da a entender que grafos muy grandes se fragmentan pronto, esto quiere decir que nuestra red es muy vulnerable.

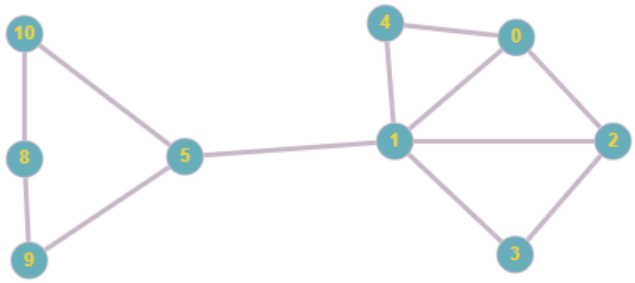

Figura 2.1: Red antes de sufrir un ataque.
Num. De componentes conectados=1
Prop. De componente gigante=1

En la figura anterior tenemos una red con 9 nodos, cada nodo tiene un número como identificador. Para la siguiente figura se realizará un ataque aleatorio que en este caso resultó afectar al nodo número 5, eliminándolo de la red.

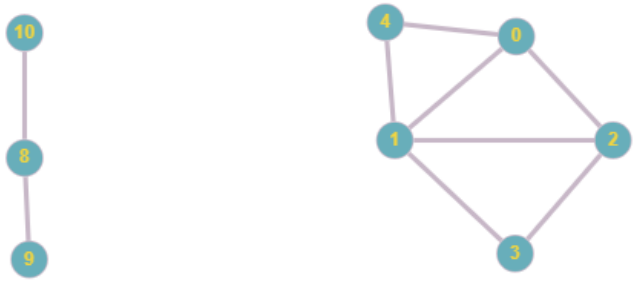

Figura 2.2: Red después de sufrir un ataque.
Num. De componentes conectados=2
Prop. De componente gigante=0.56

El resultado es la división de la red en dos partes, esto dicho de otra manera hizo crecer el número de componentes conectados de 1 a 2.
También esto afectó a la proporción del componente gigante, pasando de 1 a 0.56, (esto está determinado por la proporción: número de nodos en el componente gigante dividido por el número total de nodos).

Capítulo 3

Casos de estudio

Las redes complejas se vuelven importantes e interesantes cuando están integradas por una gran cantidad de nodos y aristas. Existe un gran número de fenómenos naturales, artificiales y sociales, que pueden ser representados y comprendidos por medio de la estructura de redes complejas, por ejemplo, la estructura social puede representarse como una red"de interacciones entre familiares y amigos, a su vez, el cerebro humano puede considerarse como una red compleja al tomar en cuenta su formación basada en neuronas, de más de 300 tipos que se interconectan a través de más de 7,000 conexiones.

3.1. Redes de computadora como redes de escala libre

Barábasi y Bonabeau en 2003 encontraron que muchas redes podían explicarse con el modelo de escala libre, por ejemplo:

- La WWW, con las páginas como nodos y los links como aristas.
  La Web es la colección mundial de páginas de texto, fotografías digitales, archivos de música, videos y animaciones a las que puede acceder a través de Internet. Lo que hace que la Web sea tan especial es la forma en que toda esta información está conectada. Los bloques de construcción básicos de la Web son páginas de texto, páginas Web. Una colección de páginas web en la misma computadora se denomina sitio web. Cada página web tiene frases destacadas llamadas enlaces (o enlaces de hipertexto) por todas partes. Al hacer clic en uno de estos, accederá a otra página de este sitio web ó a otro sitio web completamente nuevo.

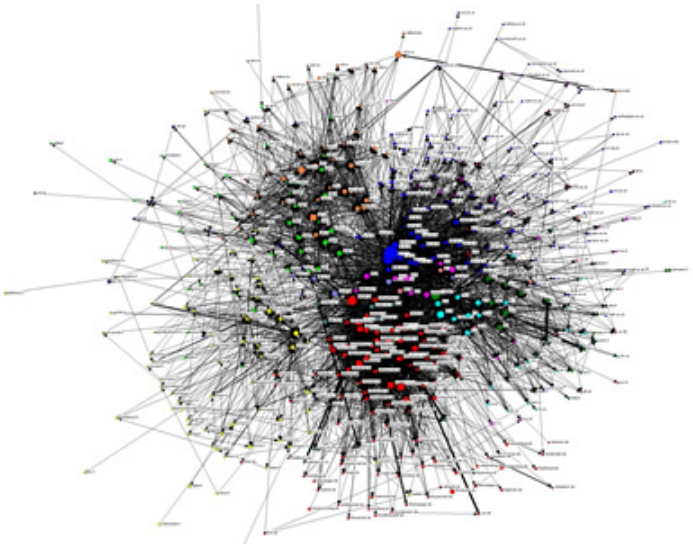

Figura 3.1: Red aproximada de la WWW.

3.2. Redes de computadora como redes aleatorias

Un centro de datos en la nube consta de miles y miles de servidores que brindan servicios a sus clientes. Las redes de centros de datos más populares se basan en topologías regulares y el objetivo principal de todas las soluciones topológicas habituales es ofrecer un gran ancho de banda. El auge del cómputo en la nube ofrecido por los centros de datos distribuidos a gran escala se han convertido en la fuerza que impulsa el crecimiento continuo. La topología de red de interconexión en los centros de datos es la principal característica para garantizar la escalabilidad y la implementación de políticas óptimas de gestión de recursos, ya que gracias a ellas, se puede conocer que tan seguro ante ataques o errores puede ser un centro de datos. Es comúnmente utilizada la topología de red aleatoria en los centros de datos, los cuales están distribuidos a lo largo del mundo en posiciones estratégicas para poder brindar el mejor servicio.

Figura 3.2: Centro de datos de Google.

3.3. Robustez de redes sometidas a errores

En este ejercicio se someterá a las redes ante una serie de errores aleatorios para poder medir la robustez de cada modelo de red, y ver que red es más resistente a los errores aleatorios.

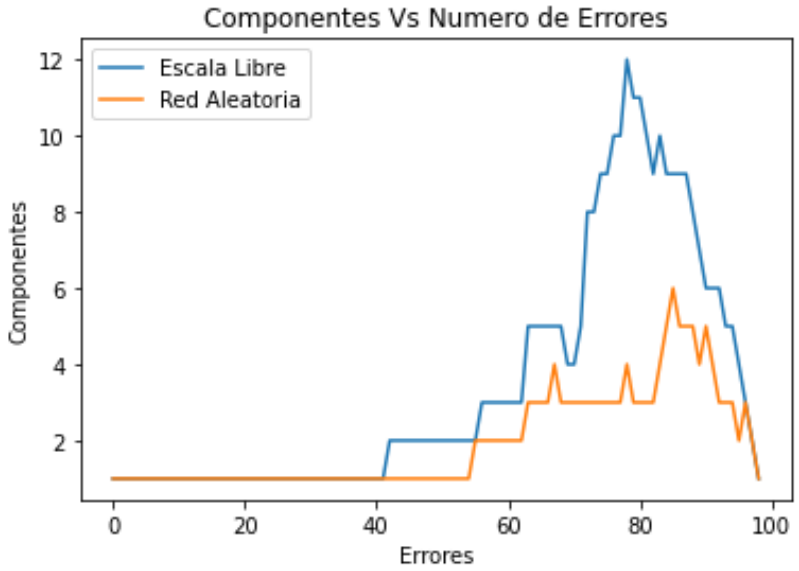

Figura 3.3: Componentes conectados en ambas redes.

Como podemos ver, la red que se fragmenta de manera más rápida es la de escala libre, la cual se fragmenta en los 41 errores, mientras que la red aleatoria se comienza a fragmentar en los 54 errores. Por lo tanto, la red más robusta ante este tipo de ataques es la red aleatoria.

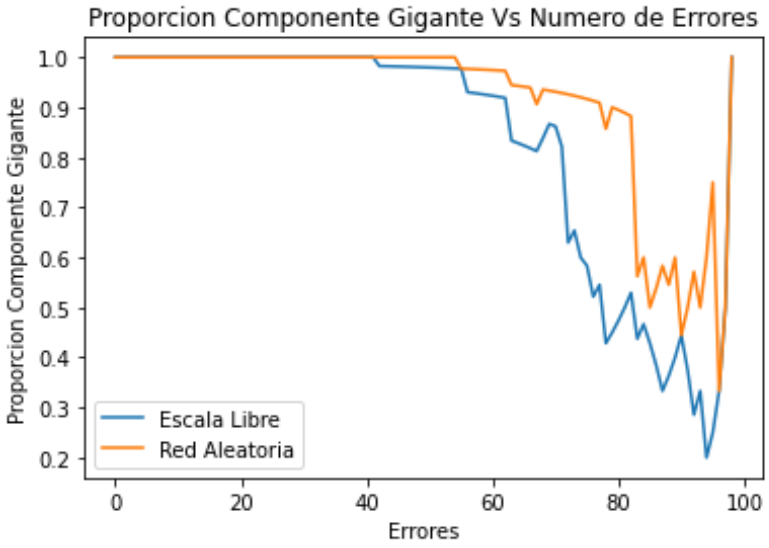

Figura 3.4: Proporción del componente gigante en ambas redes.

La anterior gráfica nos confirma lo anterior dicho, ya que podemos apreciar como el componente gigante comienza a disminuir justo en los 41 errores en la red de escala libre, y continua con esa tendencia hasta los 93 errores.

3.4. Robustez ante un ataque por intermediación

Para este ejercicio se probó la robustez de una red de escala libre modelo Barabasi-Albert y una red aleatoria modelo Erdös-Rényi ante un ataque por intermediación.

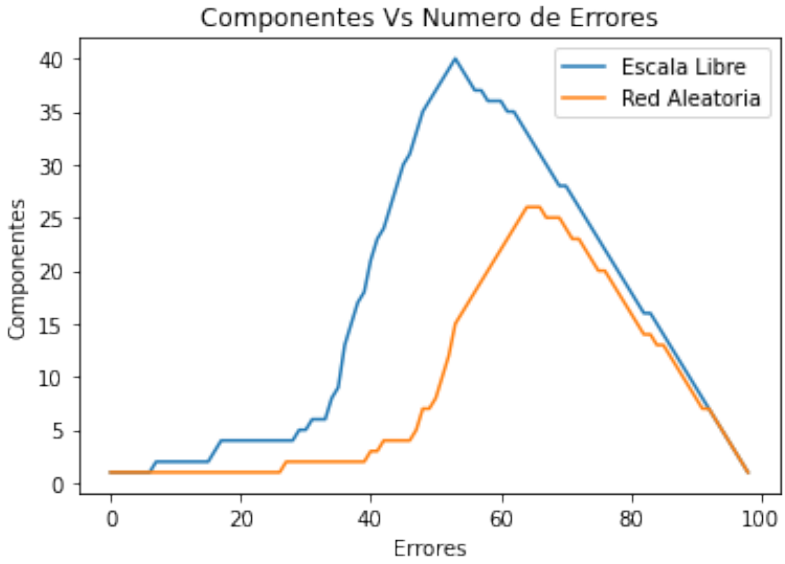

Figura 3.5: Componentes conectados en ambas redes ante ataques por intermediación.

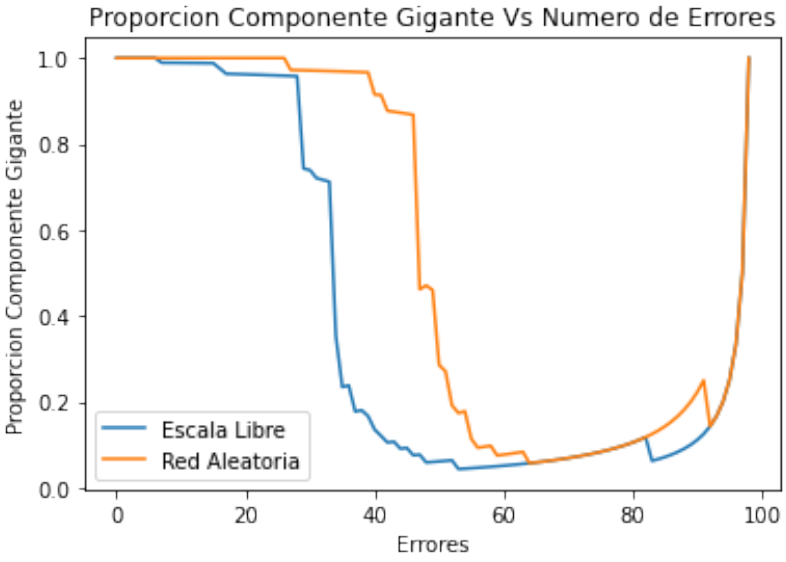

Figura 3.6: Proporción del componente gigante en ambas redes.

En la Fig. 3.5 observamos que la red más vulnerable a este tipo de ataques es la red de escala libre pues la red comienza a dividirse en los primeros 5 ataques alcanzando su maximo de componentes conectados en 47 ataques y a partir de ahí comienza a tener desconexiones, la red aleatoria en cambio comienza a dividirse con 23 ataques. Con la Fig.3.6 se comprueba que la red de escala libre es más vulnerable a este tipo de ataques pues el componente gigante comienza a disminuir con un menor número de errores.

3.5. Robustez ante un ataque por cercanía

Durante este ejercicio se someterá a ambas redes (escala libre, aleatoria) a un ataque dirigido, basado en su valor de cercanía por nodo, y de esta manera medir que tan robuzta es cada modelo de red ante este tipo de ataques. Como podemos ver, en este tipo

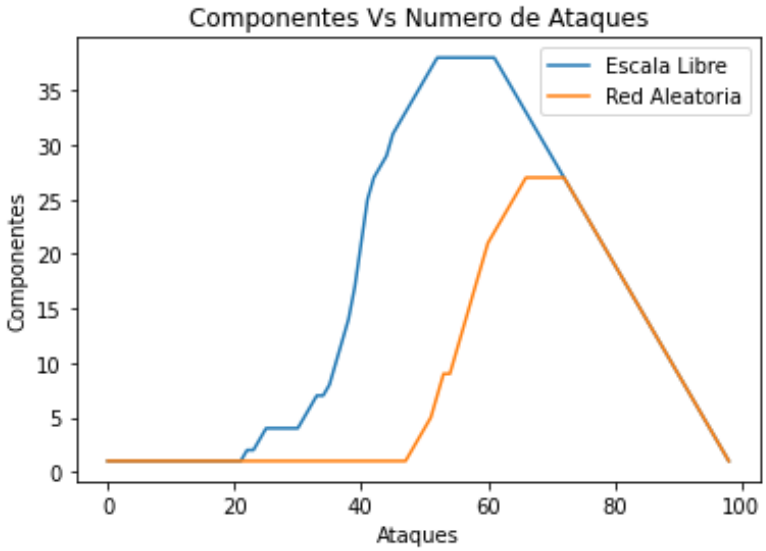

Figura 3.7: Componentes conectados en ambas redes ante ataques por cercanía.

de ataques la red más vulnerable es la de escala libre, la cual tiene un punto de ruptura en los 21 ataques, mientras que la red aleatoria tiene su punto de ruptura en 47 ataques, lo cual la hace más segura ante este tipo de ataques dirigidos.

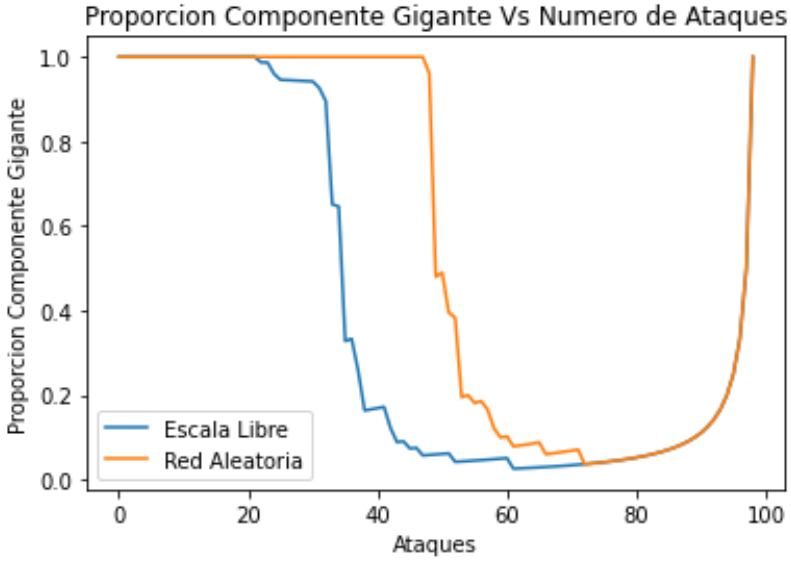

Figura 3.8: Proporción del componente gigante en ambas redes ante ataques por cercanía.

Lo anterior nos permite tener una visión más clara acerca del proceso de fragmentación de la red a medida que los ataques avanzan, y nos deja comprobar lo dicho anteriormente, en la cual la red más insegura ante este tipo de ataques es la red de escala libre.

3.6. Robustez ante un ataque por pageRank

Durante este ejercicio se someterá a ambas redes (escala libre, aleatoria) a un ataque dirigido por la métrica de PageRank, para así medir su robustez ante ataques según esta métrica.
Podemos ver el mapa de calor de la red de escala libre y de escala aleatoria, recordemos que entre más oscuro el color del nodo, más valioso o importante es para la red:

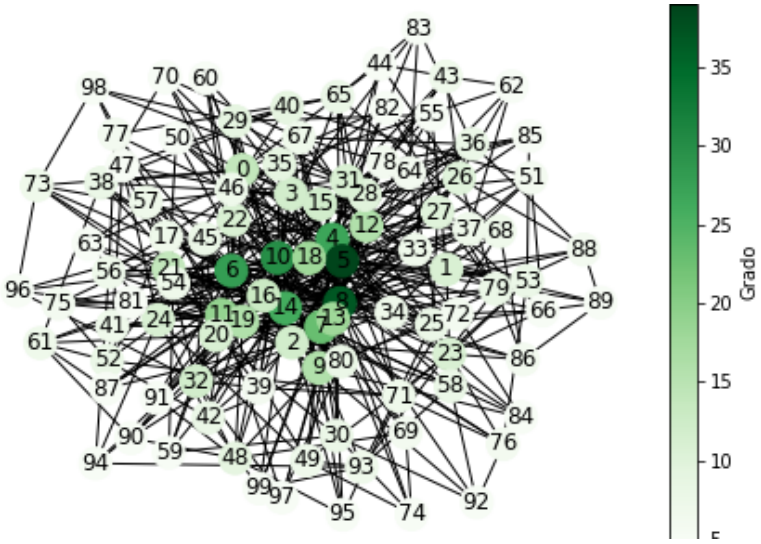

Figura 3.9: Mapa de calor de la red de escala libre.

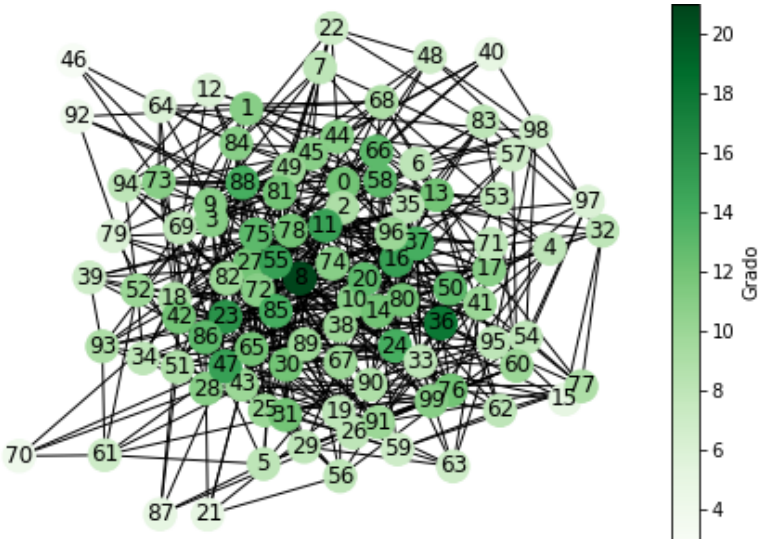

Figura 3.10: Mapa de calor de la red aleatoria.

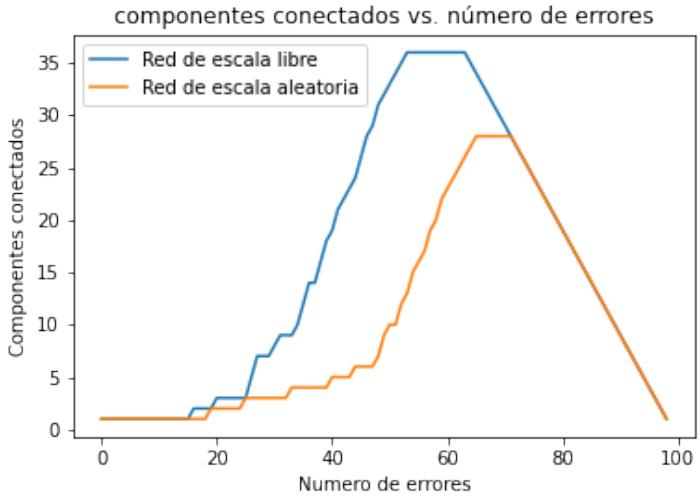

Figura 3.11: Componentes conectados en ambas redes ante ataques por PageRank.

En esta gráfica tenemos componentes conectados contra número de errores, podemos observar que tenemos la gráfica de la red de escala libre (azul) y la red aleatoria (anaranjado), la cantidad de componentes conectados es más alta en la red de escala libre, esto nos da un primer criterio para decir que este tipo de red resulta más susceptible a los ataques por PageRank que la red aleatoria.

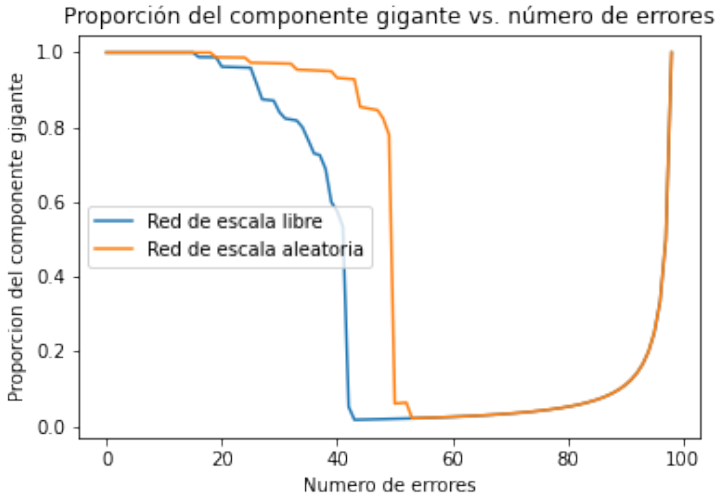

Figura 3.12: Proporción del componente gigante en ambas redes ante ataques por PageRank.

En esta gráfica vemos la proporción del componente gigante contra el número de errores. De igual forma, la red de escala libre (Color azul) parece ser más vulnerable a ataques según la métrica de PageRank, pues el componente gigante disminuye con menos ataques, lo que quiere decir que la red necesitó de menos ataques para comenzar a desconectarse, a diferencia de la red aleatoria (Color anaranjado).

Capítulo 4

Conclusiones

4.1. Conclusiones

Actualmente en el mundo moderno, vivimos una era en que la información puede ser tan delicada como para dejar en bancarrota a alguna compañía, o incluso, provocar un conflicto global, los datos hoy en día son tan delicados que la seguridad de los mismos se prioriza día a día, es por eso que actualmente antes de implementar cualquier tipo de red de computadoras, se tiene que realizar un análisis de robustez de la red propuesta, para comprobar que riesgos tiene y así, minimizar la posibilidad de daño/pérdida de información durante cualquier desastre, entre los cuales pueden ser de origen aleatorio tanto humano. En todos los casos evaluados la proporción del componente gigante disminuyo antes y más abruptamente en la red de escala libre, ocurre algo similar con la cantidad de componentes conectados, en todos los casos la cantidad de componentes conectados incrementó antes (menos ataques) en la red de escala libre, esto nos da a entender que la red de escala libre es más susceptible a ataques, pues se divide con más facilidad.
La red más robusta ante errores es claramente la red aleatoria, la cual es capaz de resistir más errores que una red de escala libre, por lo cual es muy poco vulnerable una red aleatoria ante los errores, gracias a esto, es la topología más comúnmente utilizada en los centros de datos para aumentar la seguridad de los datos y la infraestructura misma.
Lo mismo sucede en caso de ataques dirigidos por intermediación o PageRank, donde la red aleatoria comprobó ser la más robusta de todo el análisis.