

ANEXO I. PROPIEDADES DE UN GRAFO

Aunque en las últimas décadas se han propuesto e investigado muchas medidas cuantitativas para caracterizar las redes complejas, hay tres conceptos que juegan un papel clave: la longitud media de los caminos (una medida de integración), el coeficiente de *clustering* (una medida de segregación), y la distribución de grados (una medida de centralidad). Debido a la imposibilidad de manejar la información de las redes de una forma individual para los nodos, era necesario definir algunas medidas globales que permitieran caracterizar las redes con el fin de poder comparar la similitud entre ellas, así como la adecuación de los diversos modelos matemáticos con las redes reales que pretende modelar. Ha de indicarse que, debido a la alta complejidad de las redes reales, no se han encontrado todavía caracterizaciones completas de las mismas, es decir, en la actualidad todavía no disponemos de un conjunto de métricas que caractericen por completo a cada red (algo así como un código genético que permita con absoluta precisión establecer comparaciones entre ellas).

En este anexo se mostrará una descripción de las principales propiedades del grafo comentadas en el párrafo anterior con el objetivo de servir de apoyo al documento principal. Junto a ellas, se encontrarán otras métricas que también pueden resultar de interés en este trabajo. Para ofrecer una descripción más intuitiva, las métricas mostradas están orientadas a un grafo binario no dirigido (BU, de *Binary Undirected Network*), aunque todas ellas tienen su homólogo en grafos ponderados no dirigidos (WU, de *Weighted Undirected Network*). No obstante, todo el código utilizado contiene una explicación más detallada y referencias que contienen la descripción formal sobre la función utilizada, especialmente aquellas de las librerías Brain Connectivity Toolbox (BCT) de Matlab y NetworkX de Python. Por último, se realizará una descripción de los principales tipos de redes tratadas en este trabajo.

1. CENTRALIDAD

En teoría de grafos y análisis de redes, la centralidad en un grafo se refiere a una medida de un vértice en dicho grafo, que determina su importancia relativa dentro de éste. Poder reconocer la centralidad de un nodo puede ayudar a determinar, por ejemplo, el impacto de una persona involucrada en una red social, la relevancia de una habitación en un edificio representado en sintaxis del espacio, la importancia de una carretera en una red urbana, o los componentes esenciales de una red de computadoras.

Por ejemplo, de acuerdo con una medida de centralidad razonable, en un grafo estrella el nodo central debería ocupar un valor máximo de centralidad, mientras que los nodos de las puntas ocuparían un valor de centralidad inferior.

- Grado:

La más simple, y quizás también la más usada, característica individual de un nodo es su grado. El grado, k_i , de un nodo i se define generalmente como el número total de sus conexiones. El promedio de los grados de los nodos de una red se llama grado medio de la red, y se denota por \bar{k} . En grafos ponderados, el peso del enlace es ignorado para calcular esta métrica. En principio, aquellos nodos con mayor grado tendrán un rol más importante en la red ya que intercambian información con un mayor número de nodos. No obstante, esto no siempre es así, dado que existen otros parámetros como el peso de los nodos o el grado/rol de los nodos a los que está conectado que pueden modificar este grado de “importancia”.

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 grado 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 (es decir, la función se anula para todos los posibles grados, menos para el grado que tienen todos sus nodos, donde se alcanza el valor 1). Cualquier aleatoriedad en la red modificará 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} . Debido a este descenso exponencial, la probabilidad de encontrar un nodo con k enlaces se convierte en insignificante para $k \gg \bar{k}$.

En los últimos años, muchos resultados empíricos muestran que en la mayoría de las redes reales de gran escala el grado de distribución se desvía significativamente de la distribución de Poisson. En particular, para muchas redes el grado de distribución se describe mucho mejor por una ley de potencias de la forma $P(k) \sim k^{-\gamma}$. Esta distribución cae de forma más gradual que una exponencial, lo que permite la existencia de algunos nodos de grado muy alto. Debido a que estas leyes de potencia no se concentran alrededor de una media (escala), es por lo que se llaman libres de escala.

- **Intermediación o *betweenness*:**

Es la fracción de todos los caminos más cortos en la red que pasan por dicho nodo. Es decir, la intermediación cuantifica la frecuencia o el número de veces que un nodo actúa como “puente” a lo largo del camino más corto entre otros dos nodos. Los nodos con una alta intermediación, suelen jugar un rol crítico en la estructura de la red, y en ocasiones la integración de algunos componentes de la red depende ellos. Los nodos que poseen una posición de intermediarios de alguna manera son también controladores o reguladores del flujo de información. Así, si un nodo con alta intermediación desaparece, el camino característico más corto de la red se vería incrementado, disminuyendo la eficiencia de la red.

- **Autovalor o *eigenvalue*:**

La centralidad de vector propio mide la influencia de un nodo en la red. Los nodos que poseen un valor alto de esta medida de centralidad están conectados a muchos nodos que a su vez están bien conectados, también en este sentido; por lo tanto, son buenos candidatos para difundir información. Los nodos más centrales en este sentido corresponden a centros de grandes grupos cohesivos. Mientras que en el caso de la centralidad de grado, cada nodo *pesa* lo mismo dentro de la red, en este caso la conexión de los nodos se pondera de forma diferente.

El cálculo del *PageRank* de Google, utilizado para medir la relevancia de páginas web en Internet, es una variante de esta medida.

- **Coefficiente de participación:**

Evalúa la diversidad de conexiones intermodulares de los nodos, donde por “diversidad” se entiende a las múltiples posibles combinaciones de enlaces que pueden darse entre dos nodos: nodo poco conectado con nodo muy conectado, nodo que une varias comunidades con nodo que solo está conectado con nodos de su propia comunidad, etc. Por ejemplo, nodos con alto grado de enlaces intra-módulo pero bajo coeficiente de participación, conocidos como *hubs* provinciales, favorecen la segregación modular. Por otro lado, nodos con alto grado de participación suelen favorecer la integración intermodular y son conocidos como *hubs* conectores.

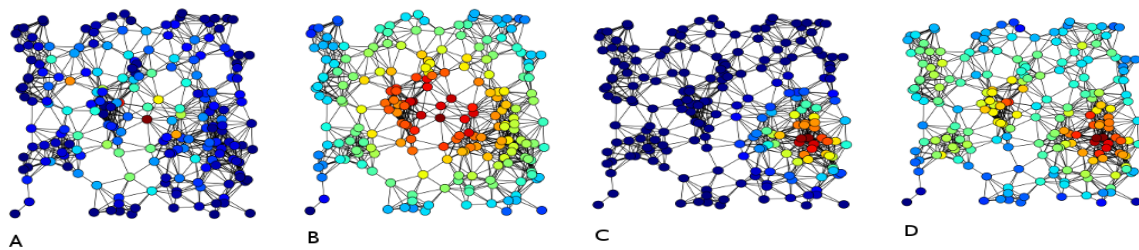


Fig.1. Distintas medidas de centralidad sobre una red (Wikipedia). A) Intermediación, B) cercanía, C) Autovalor, D) Grado.

2. **ESTRUCTURA DE LA RED**

- **Densidad:**

Es la fracción de las conexiones que tiene un nodo entre todas las posibles conexiones que tendría. En WU, el peso de los enlaces es ignorado.

- **Componentes conectadas:**

Se define como subgrafo en el que cualquier par de nodos contenidos en él están conectados por algún camino. Algunas propiedades de la red, como el camino característico más corto, no pueden ser evaluadas si hay componentes no conexas (habría pares de nodos cuyo camino entre ellos sería infinito); una alternativa común es evaluar estas propiedades en la componente más grande, también denominada componente gigante, y asumirlas como representativas de la red.

- **Clustering:**

El coeficiente de *clustering* de un nodo es el número de triángulos alrededor de un nodo y es equivalente a la fracción de vecinos de un nodo que son vecinos entre sí (que forman un triángulo). Es decir, es una forma de cuantificar la interconexión de un nodo con sus vecinos, definiendo así el coeficiente de *clustering* como la proporción media de pares de vecinos de un nodo que también son vecinos entre sí.

Supongamos que un nodo i de la red tiene k_i vecinos (o enlaces). Es evidente que, a lo sumo pueden existir $\frac{k_i(k_i-1)}{2}$ enlaces entre ellos (y esto solo ocurre si todos están conectados entre sí). El coeficiente de clustering, C_i , del nodo i se define entonces como la proporción entre, E_i , el

número de enlaces que de verdad existen entre los vecinos de i , y el máximo número de enlaces posibles en la red, es decir:

$$C_i = \frac{E_i k_i (k_i - 1)}{2}$$

El coeficiente de *clustering* de la red será la media de los coeficientes de *clustering* de todos sus nodos. Obviamente, $C \leq 1$ y solo tomará el valor 1 en el caso de que la red sea completa (es decir, aquella que tiene todas las posibles conexiones entre todos sus nodos). En una red completamente aleatoria de N nodos se puede probar que $C \sim 1/N$, un valor que sería muy pequeño en la mayoría de las redes reales, que tienen un número muy elevado de nodos. Sin embargo, se ha constatado numéricamente que la mayoría de las redes reales tienen tendencia a la agrupación, en el sentido de que sus coeficientes de *clustering* son mucho mayores que $O(1/N)$, a pesar de ser significativamente menores que 1, de donde podemos inferir que la mayoría de las redes complejas reales no son ni completamente aleatorias ni completas.

- **Transitividad:**

Es una medida normalizada del coeficiente de *clustering* que se define como la proporción de triángulos y tripletas en la red. La transitividad también se define como la propiedad de que si A está conectado con B y B está conectado con C , es muy probable que A y C estén conectados.

- **Modularidad y estructura de comunidades:**

La modularidad es la capacidad que tiene un [sistema](#) de ser estudiado, visto o entendido como la unión de varias partes que interactúan entre sí y que trabajan solidariamente para alcanzar un objetivo común, realizando cada una de ellas una tarea necesaria para la consecución de dicho objetivo. Así, en teoría de grafos, la modularidad permite cuantificar el grado en el que una red puede ser subdividida en grupos de nodos. Formalmente, la modularidad es la fracción de los enlaces que caen dentro de grupos dados menos el valor esperado que dicha fracción hubiese tomado si los enlaces se hubiesen distribuido al azar. Es decir, la modularidad refleja la concentración de los nodos dentro de los módulos en comparación con una distribución al azar de los enlaces entre todos los nodos, independientemente de los módulos.

A estos módulos o particiones que se generan se les denomina comunidades y describen la estructura funcional de la red. Se define como estructura de comunidades óptima aquella subdivisión de la red que, sin solapamiento entre los grupos de nodos (suponiendo que se utiliza el algoritmo de Louvain), maximiza el número (BU) o fuerza (WU) de conexiones intra-modular y minimiza el número (BU) o fuerza (WU) de las conexiones entre módulos; esto es, que maximiza la modularidad.

Se ha demostrado que en redes reales, como el cerebro, existe un fuerte comportamiento modular y jerárquico, en el que la estructura de la red es subdividida en comunidades (y estos a su vez vuelven a ser divididos en subcomunidades) con tareas diferentes asociadas a cada una de ellas, de modo que puede interpretarse como una medida de segregación de la red.

3. DISTANCIAS

- Camino:

Los caminos son las posibles secuencias de enlaces que unen dos nodos, de forma que no pasen por el mismo nodo más de una vez. A través de esta propiedad se pueden generar la matriz de alcance, que determina si dos nodos están conectados (por un camino), y la matriz de distancia, que contiene la longitud del camino más corto entre dos pares de nodos. El camino característico de una red se define como la media de las distancias entre todos los pares de nodos, es decir, la separación típica entre pares de nodos.

Desde una perspectiva histórica, a medida que se tenían más datos de redes complejas reales, se constató que la longitud promedio de los caminos de la mayoría de ellas era relativamente pequeña, incluso en los casos en que estos tipos de redes tuvieran muchos menos enlaces de los posibles que se podrían dar (a menor cantidad de conexiones en el mundo, parece claro que haya que hacer recorridos más largos para poder llegar de un nodo a otro). Esta característica es lo que se llamó efecto de mundo pequeño, y de ahí el nombre de redes de mundo pequeño que intentan modelarlo.

- Eficiencia:

La eficiencia global es inversamente proporcional al camino característico de la red. La eficiencia local es la eficiencia global calculada en la vecindad del nodo y está relacionada con el coeficiente de clustering.

- *Mean first passage time:*

Es el número de pasos o enlaces estimado necesarios para ir de uno nodo a otro de forma aleatoria.

- Difusión:

Es una métrica inversamente proporcional al *mean first passage time*.

4. SIMILITUD

- Asortatividad:

El coeficiente de asortatividad es una correlación entre el grado de todos los nodos en dos extremos opuestos de un enlace. Un valor positivo de este coeficiente indica que los nodos tienden a vincularse a otros nodos con el mismo grado o un grado similar.

- *Matching Index*

Este índice calcula la cantidad de solapamiento en los patrones de conexión entre dos nodos i, j . Los autoenlaces y los enlaces $i-j$ son ignorados. El *matching index* es una métrica de simetría similar a una correlación o a un producto escalar.

5. MODELOS DE REDES COMPLEJAS

Medir algunas propiedades básicas de una red compleja, como las que hemos definido en el apartado anterior, es el primer paso para comprender su estructura. El siguiente paso es, entonces, desarrollar un modelo matemático de la red que proporcione una topología con propiedades estadísticas similares a las reales, obteniendo una plataforma en la que sea posible aplicar diversos métodos matemáticos para analizar comportamientos generales de redes similares (modelos nulos).

- Redes regulares:

Intuitivamente, una red completa tiene la menor longitud de camino promedio ($L=1$ ya que todos los nodos están conectados) y el mayor coeficiente de *clustering* ($C=1$). Aunque el modelo de red completa captura las propiedades de mundo pequeño y alto *clustering* de muchas redes reales, es fácil darse cuenta de sus limitaciones: una red completa con N nodos tiene $N(N-1)/2$ enlaces, mientras que la mayoría de las redes reales de gran escala parecen tener un número de enlaces de orden $O(N)$ en lugar de $O(N^2)$.

De forma menos extrema, podemos considerar las redes regulares de grado k , que son aquellas en las que todos los nodos tienen el mismo grado, que tienen una cantidad lineal de enlaces si $k \ll N$, un alto coeficiente de *clustering*, y una longitud promedio alta.

- Redes completamente aleatorias:

En el extremo opuesto del espectro de una red totalmente regular están las redes completamente aleatorias, estudiadas por Erdős y Rényi (ER) a finales de la década de los 50. Dado un gran número de nodos ($N \gg 1$) y una cierta probabilidad p de enlace entre dos nodos (la misma siempre), un grafo aleatorio de ER tendría $pN(N-1)/2$ enlaces.

Por ejemplo, una propiedad de este tipo es la de la conectividad global, que se puede medir por medio del tamaño de la componente conexa más grande (la componente gigante): ER mostraron que si la probabilidad p es superior a un determinado umbral, prácticamente todo el grafo está conectado (se puede llegar de un nodo a otro por medio de los enlaces), y por debajo de dicho umbral la conectividad era sensiblemente inferior (casi desconectado), en contra de la intuición que parece conducir a la conclusión de que a medida que se incrementa la probabilidad de conexión, p , aumenta proporcionalmente la conectividad global de la red.

El grado medio de una red aleatoria es $k^- = p(n-1) \sim pN$, y la longitud del camino promedio es $L \sim \ln(N)/k^-$, por lo que, debido a que $\ln(N)$ crece lentamente en función de N , la longitud del camino promedio se mantiene bastante bajo incluso en una red con muchos nodos. Este aumento logarítmico de la longitud del camino promedio con el tamaño de la red es un efecto típico de mundo pequeño. Por otro lado, en una red aleatoria el coeficiente de *clustering* es $C = p = k^-/N \ll 1$, lo que significa que si tiene muchos nodos apenas mostrará *clusterización*. De hecho, para un valor de N muy grande, las redes aleatorias ER son redes casi homogéneas, donde la conectividad sigue aproximadamente una distribución de Poisson.

- **Redes de mundo pequeño:**

En matemática y física una **red de mundo pequeño** es un tipo de grafo para el que la mayoría de los nodos no son vecinos entre sí, y sin embargo la mayoría de los nodos pueden ser alcanzados desde cualquier nodo origen a través de un número relativamente corto de saltos entre ellos. Los modelos de mundo pequeño, tienen sus raíces en las redes sociales no virtuales, donde la mayoría de la gente es amiga de sus vecinos físicos inmediatos, por ejemplo los vecinos en la misma calle o compañeros en la misma oficina. Por otro lado, muchas personas tienen un par de amigos que están lejos, en la distancia, como amigos de otros países, que están representados por las conexiones de largo alcance.

La idea central de la generación de redes de pequeño mundo está fundamentado en dos propiedades:

- El fenómeno del mundo pequeño - es decir que cualesquiera dos nodos de la red se comunican por un camino de nodos intermedios relativamente pequeño (pequeño número de nodos). Comprobando que la distancia máxima entre dos nodos crece logarítmicamente con el número de nodos en la red.
- Poseen valores altos de coeficiente de agrupamiento (*clustering coefficient*) - este valor fue empleado por Watts y Strogatz. Este valor viene a indicar que si dos vértices o nodos se conectan mediante la intervención de otros nodos, existe una gran probabilidad de que estén conectados directamente entre sí

Como se ha comentado, las redes regulares tienen el efecto de *clusterización*, pero no el de mundo pequeño en general. Por otro lado, los grafos aleatorios muestran el efecto de mundo pequeño, pero no el de *clusterización*. Por tanto, no debe sorprender ver que estos modelos no reproducen algunas características importantes de muchas redes reales. Después de todo, la mayoría de las redes del mundo real no son ni enteramente regulares ni totalmente aleatorias.

Con el fin de describir la transición de una red regular en una aleatoria y disponer de un modelo que mezcle las propiedades deseadas que tienen ambos modelos, Watts y Strogatz (WS) introdujeron un interesante modelo de red de mundo pequeño. Las redes del modelo WS pueden ser generadas siguiendo el siguiente algoritmo:

1. Se comienza por una red regular de N nodos.
2. Se selecciona cada enlace al azar con probabilidad p y se reconecta uno de sus extremos con cualquiera de los otros nodos de la red.

Este proceso introduce $pNK/2$ enlaces de largo alcance que conectan nodos que de otra forma estarían en diferentes vecindarios. Tanto el comportamiento del coeficiente de *clustering* $C(p)$ como de la longitud del camino promedio, $L(p)$, pueden considerarse como una función de la probabilidad de reconexión, p .

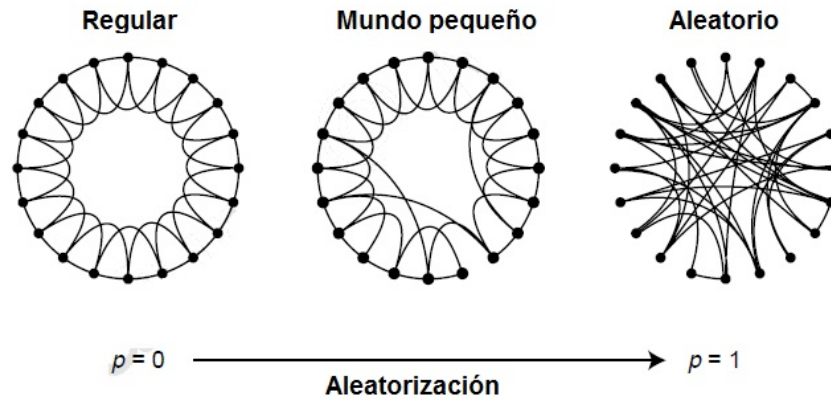


Fig.2. Algunos tipos de redes. En la red regular la estructura y el grado de los nodos es homogéneo. En las redes de mundo pequeño, los nodos pueden tener conexiones con nodos lejanos además de con sus vecinos. En redes aleatorias, existe una probabilidad p que puede enlazar cualquier par de nodos de la red

Ejemplos de redes de pequeño mundo se han encontrado en numerosas redes presentes en la naturaleza, una de ellas es la que define las proteínas que interaccionan en el metabolismo de las bacterias o redes cerebrales como la del gusano *C. Elegans*.