Redes Sociales

Máster en Data Science y Big Data

Alejandro Llorente Pinto

2022-04-28

Índice

1 Métricas en redes sociales

Conceptos de la primera sesión

Grafos dirigidos (o no)
Grafos con pesos (o no)
Ejemplos de grafos
Componentes conexas
Tipos de grafos “teóricos”
Representación de grafos

Densidad
Distancias y caminos mínimos
Diámetro
Camino medio
Cliques
Motifs

Reciprocidad

- En una red dirigida, una relación es recíproca si se observa enlaces mutuos entre dos nodos.
- Si quisiéramos medir si una es es muy recíproca o no, ¿cómo lo haríais?

Podemos definir la reciprocidad como

$$r = \frac{L^{<->}}{L}$$

Donde L es el número total de relaciones únicas entre dos nodos, y el numerador es el número de links mutuos.

Sin embargo, cuando observamos redes muy densas, los enlaces mutuos son más probables que aparezcan, por lo que deberíamos cambiar la definición.

Reciprocidad

- Otra forma de verlo es pensar que, por ejemplo, si la red es “muy” recíproca, cada vez que veamos un valor de la matriz de adyacencia distinto de cero, también lo veremos en su recíproco.

$$a_{ij} = 1 \rightarrow \text{(a menudo)} \;\; a_{ji} = 1$$

- Sin embargo, uno podría pensar que también podría haber redes que son todo lo contrario, que tienden a no crear vinculaciones recíprocas, es decir:

$$a_{ij} = 1 \rightarrow \text{(a menudo)} \;\; a_{ji} = 0$$

- En general, podemos cuantificar el grado de variación conjunta dos variables aleatorias mediante la covarianza y su versión normalizada, la correlación lineal:

$$Cov(X,Y) = \frac{\sum (X_i - <X>)(Y_i - <Y>)}{N-1} \qquad \rho = \frac{Cov(X,Y)}{\sigma_X \sigma_Y}$$

Reciprocidad

- Sustituyendo…

$$\rho = \frac{\sum_{i \neq j} (a_{ij} - \bar{a})(a_{ji} - \bar{a})}{\sum_{i \neq j} (a_{ij} - \bar{a})^2}$$

donde

$$\bar{a} = \frac{\sum_{i \neq j} a_{ij}}{N(N-1)}$$

- Esto es simplemente aplicar la definición de correlación lineal al caso de estas variables aleatorias, que deducimos directamente de la matriz de adyacencia. Esto se puede simplificar en la siguiente fórmula que, además, nos permite interpretar de manera clara el concepto de reciprocidad.

$$\rho = \frac{r - \bar{a}}{1 - \bar{a}}$$

- Si este coeficiente es positivo, entonces la red es más recíproca de lo que esperamos aleatoriamente.
- Si es negativo, es menos.
- Pero también puede ser cercano a cero, es decir, redes que no son ni una cosa ni la contraria.

Reciprocidad

Network	ρ	σ_ρ	$\rho_{\min}$
Perfectly reciprocal	1	...	$-\frac{\bar{a}}{1-\bar{a}}$
World Trade Web (53 webs) [10]			
Most correlated (year 2000)	0.952	0.002	($\bar{a} > 0.5$)
Least correlated (year 1948)	0.68	0.01	−0.80
World Wide Web [7]	0.5165	0.0006	−0.0001
Neural networks [13,14]			
Neuron classes	0.44	0.03	−0.04
Neurons	0.41	0.02	−0.03
Email networks [5,6]			
Address books	0.231	0.003	−0.001
Actual messages	0.194	0.002	−0.001
Word networks [15]			
Dictionary terms	0.194	0.005	−0.002
Free associations	0.123	0.001	−0.001
Cellular networks (43 webs) [16]			
Most correlated (*H. influenzae*)	0.052	0.006	−0.001
Least correlated (*A. thaliana*)	0.006	0.004	−0.003
Areciprocal	0	...	$-\frac{\bar{a}}{1-\bar{a}}$
Shareholding networks [17]			
NYSE	−0.0012	0.0001	−0.0012
NASDAQ	−0.0034	0.0002	−0.0034
Food webs [11,12]			
Silwood Park	−0.0159	0.0008	−0.0159
Grassland	−0.018	0.002	−0.018
Ythan Estuary	−0.031	0.005	−0.034
Little Rock Lake	−0.044	0.007	−0.080
Adirondack lakes (22 webs)			
Most correlated (B. Hope)	−0.06	0.02	−0.10
Least correlated (L. Rainbow)	−0.102	0.007	−0.102
St. Marks Seagrass	−0.105	0.008	−0.105
St. Martin Island	−0.13	0.01	−0.13
Perfectly antireciprocal	−1	...	−1

En redes que involucran intercambios, vemos una gran reciprocidad.

En redes neuronales (las de verdad), también vemos que son altamente recíprocas.

En otras redes, como las de email, vemos reciprocidad alta pero no tanto como las anteriores.

Existen redes, como las celulares,que no presentan reciprocidad, tiene que ver con la reversibilidad de procesos químicos.

En redes que tienen que ver con ecosistemas y cadenas tróficas, se exhibe antireciprocidad.

Patterns of Link Reciprocity in Directed Networks

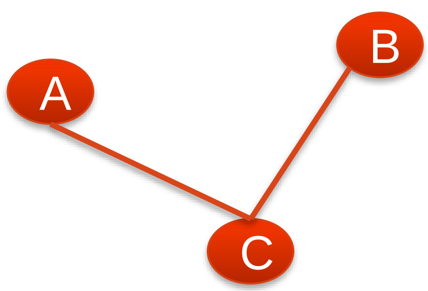
Diego Garlaschelli[1,2] and Maria I. Loffredo[2,3]
[1]*Dipartimento di Fisica, Università di Siena, Via Roma 56, 53100 Siena, Italy*
[2]*INFM UdR Siena, Via Roma 56, 53100 Siena, Italy*
[3]*Dipartimento di Scienze Matematiche ed Informatiche, Università di Siena, Pian dei Mantellini 44, 53100 Siena, Italy*
(Received 23 April 2004; published 20 December 2004)

Global clustering

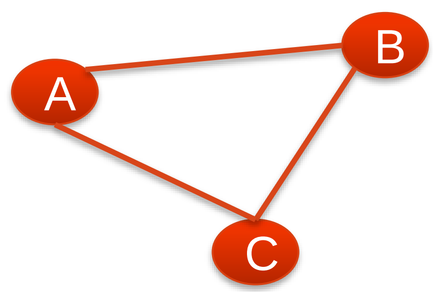
- En el contexto de redes, hablamos de clustering (o transitividad) si se produce que cuando un nodo A está conectado con B y B con C, entonces es probable que se observe un enlace entre A y C.

- Llamamos tripleta a cualquier subgrafo que involucra a tres nodos con, al menos, dos enlaces.
- Llamamos un triángulo (o tripleta) cerrado (a) a una tripleta con tres enlaces.

Tripleta



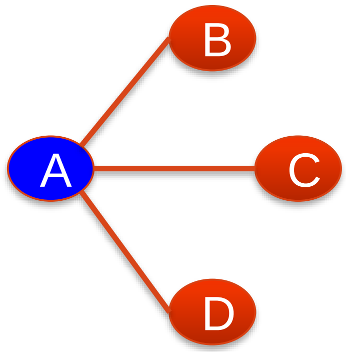
Closed triangle

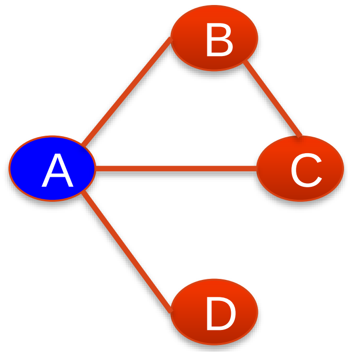


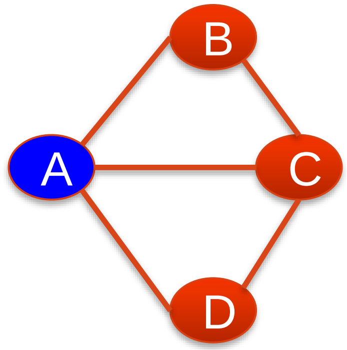
$$C = \frac{3 \times \text{number of triangles}}{\text{number of connected triplets of vertices}} = \frac{\text{number of closed triplets}}{\text{number of connected triplets of vertices}}.$$

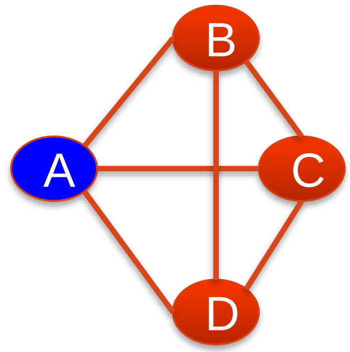
Clustering Local

- Sin embargo, también podemos pensar en el clustering como una propiedad local a un nodo.









Todas las tripletas en las que participa A son abiertas, por tanto, el clustering local es 0.

En este caso, A participa en un triángulo cerrado, por lo que su clustering local es 1/3.

En este otro caso, participa en dos triángulos cerrados, por lo que el clustering local es 2/3.

Finalmente, en este caso, todas las tripletas en las que participa están cerradas, por lo que su clustering local es 1.

Clustering Local

Table 1 Empirical examples of small-world networks

	L_{actual}	L_{random}	C_{actual}	C_{random}
Film actors	3.65	2.99	0.79	0.00027
Power grid	18.7	12.4	0.080	0.005
C. elegans	2.65	2.25	0.28	0.05

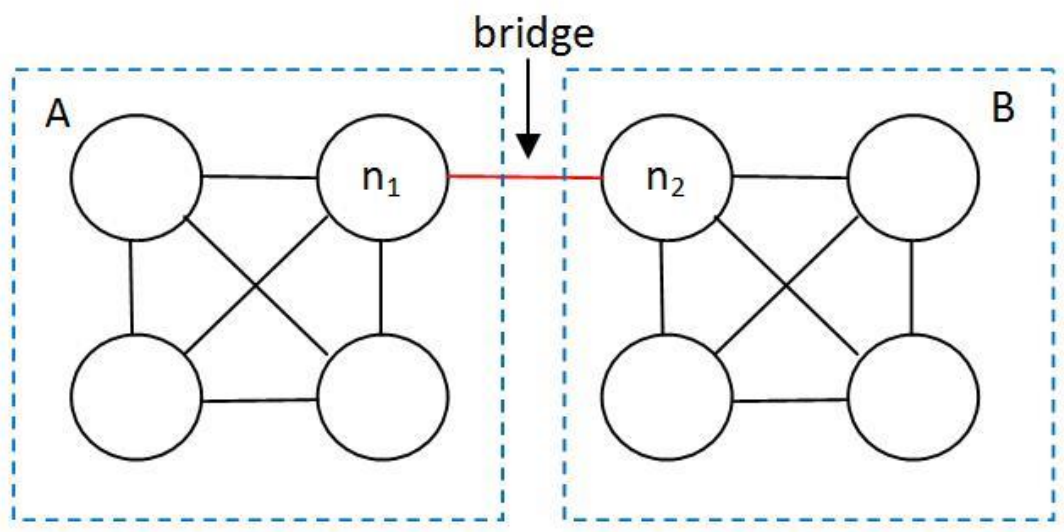
Collective dynamics of 'small-world' networks

Duncan J. Watts* & Steven H. Strogatz

Department of Theoretical and Applied Mechanics, Kimball Hall, Cornell University, Ithaca, New York 14853, USA

Centralidad

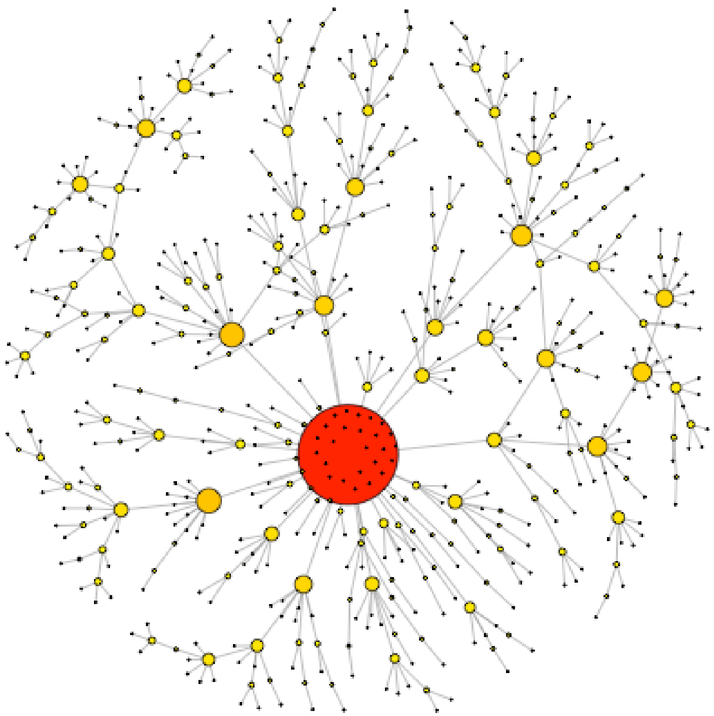
- En muchas ocasiones, es interesante identificar nodos (o enlaces) que están estratégicamente colocados.



- Pueden ser nodos importantes ya que conectan zonas de la red que, de otra forma, estarían desconectadas.
- Son nodos o enlaces clave en la transmisión de información ya que, si no existieran, no se podrían alcanzar estas regiones.
- Son clave porque, sin ser los que más conexiones tienen, llegan a zonas alejadas de la red.

Centralidad por grado

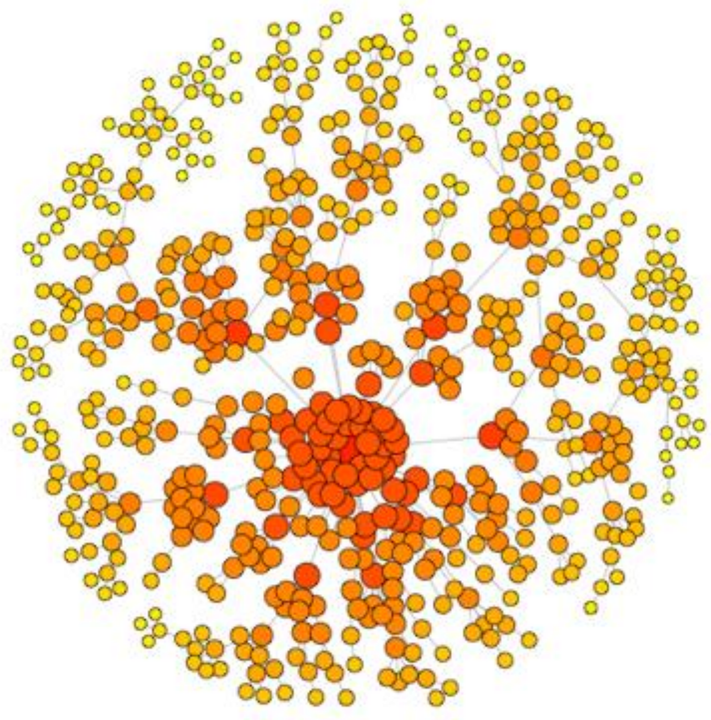
- “El nodo más importante es el que tiene un mayor número de conexiones”



- Da más importancia a los “hubs”, nodos con muchas conexiones.
- Sin embargo, esto no tiene por qué ser verdad en muchos casos:
  - En una red social, ¿difunde a más gente el que tiene un mayor número de conexiones?
  - Si las redes se estructuran por grupos (comunidades), ¿son todos los enlaces de la misma?

Centralidad por cercanía - Closeness

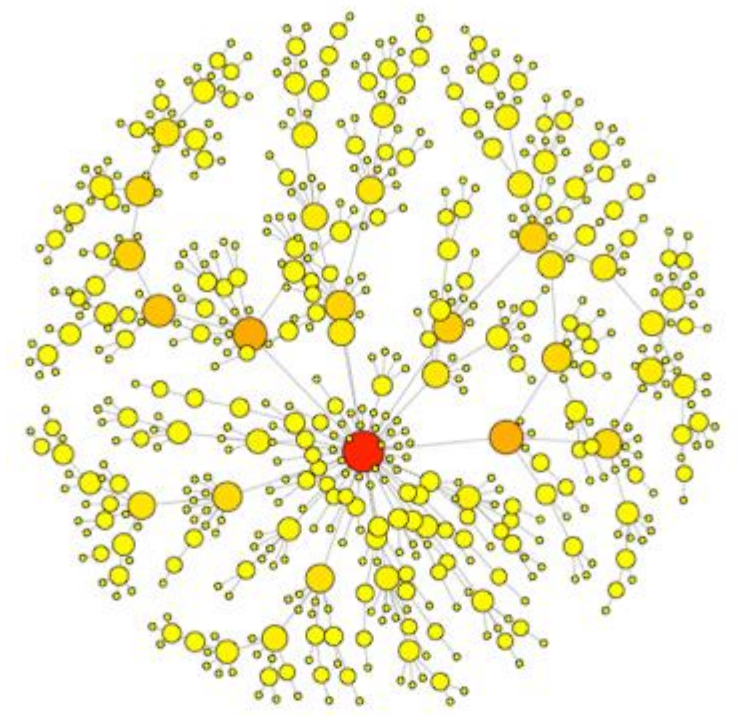
- “Un nodo es más importante si el resto están cerca de él”



- Calculamos la distancia media de cada nodo al resto de nodos.
- Esto tiene mejor pinta ya que nodos puente es cierto que están, de media, más cerca de dos grupos de nodos que están alejados en el grafo.
- Sin embargo, viendo esta imagen, ¿qué pega le pondríais a esta medida para, por ejemplo, detectar “influencers”?

Centralidad por “estar en medio” - Betweenness

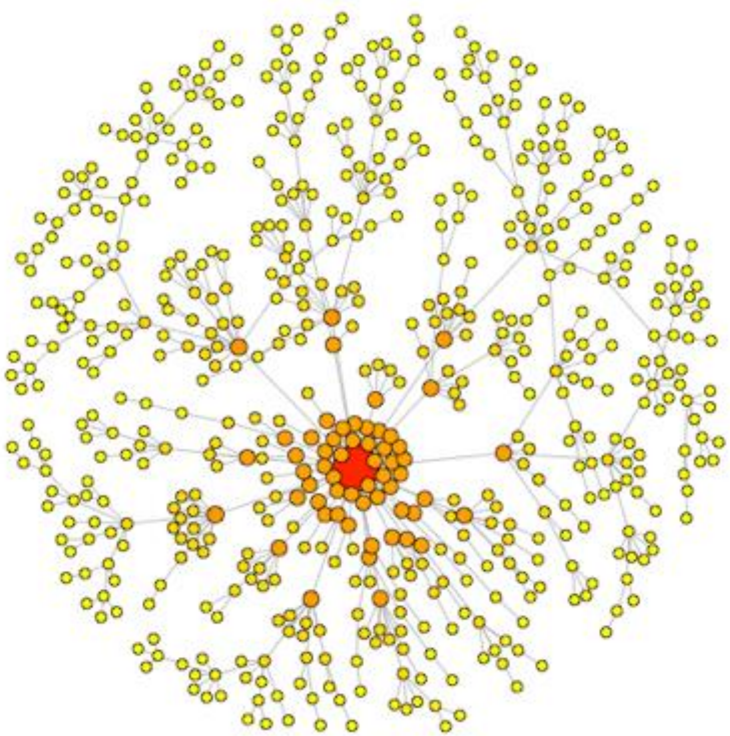
- Refinemos: un nodo diremos que es importante si, obligatoriamente, aparece frecuentemente en los caminos mínimos entre dos nodos.



- Esto es mucho mejor: se identifica un nodo importante pero también otros nodos que acaban generando grandes subgrafos, aunque no todos los nodos estén conectados con ellos.
- Es la métrica más estándar de centralidad.
- ¿Le encontráis alguna pega?

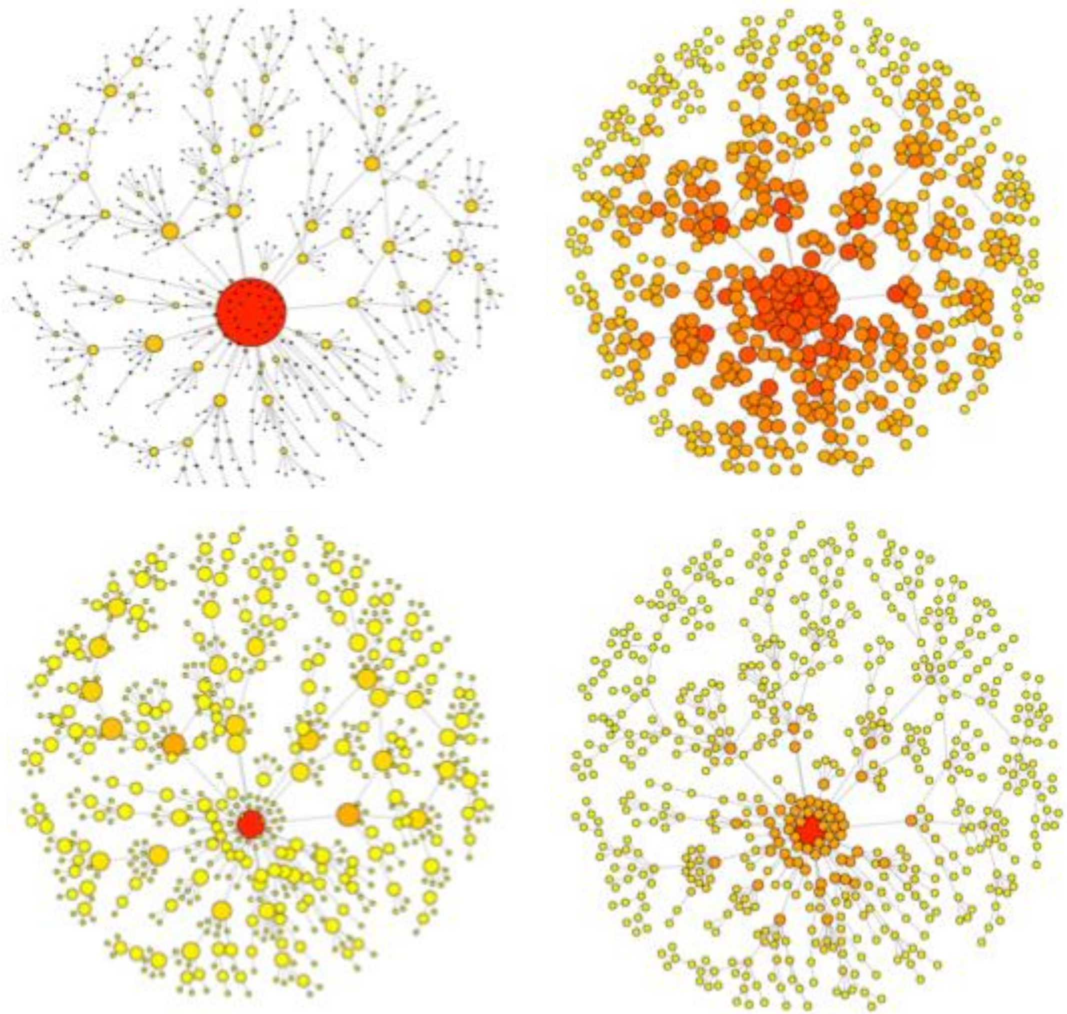
Centralidad por “autovalor” (eigenvector)

- Centralidad por estar cerca de los nodos con mucho grado



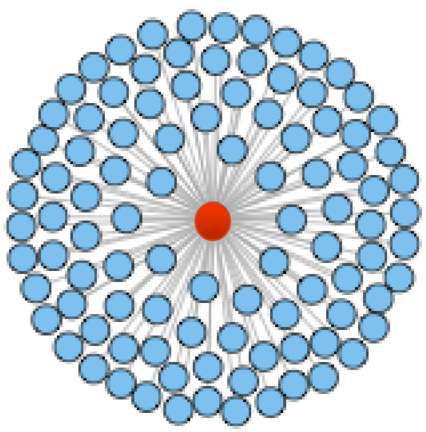
- Este algoritmo, similar al que utilizan los buscadores para generar sus score de importancia como el Page Rank, identifica a los nodos con mucho grado como importantes y a los que están cerca de ellos.

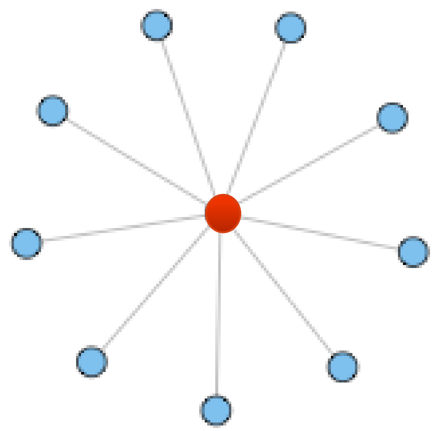
Comparación visual de medidas de centralidad



Diversidad de conexiones (redes sin pesos)

- ¿Cómo podemos medir lo diversas que son las conexiones de un nodo?
- **Primera definición**: el grado. En vuestra opinión, ¿cuál de estos dos nodos tiene una mayor diversidad de conexiones?

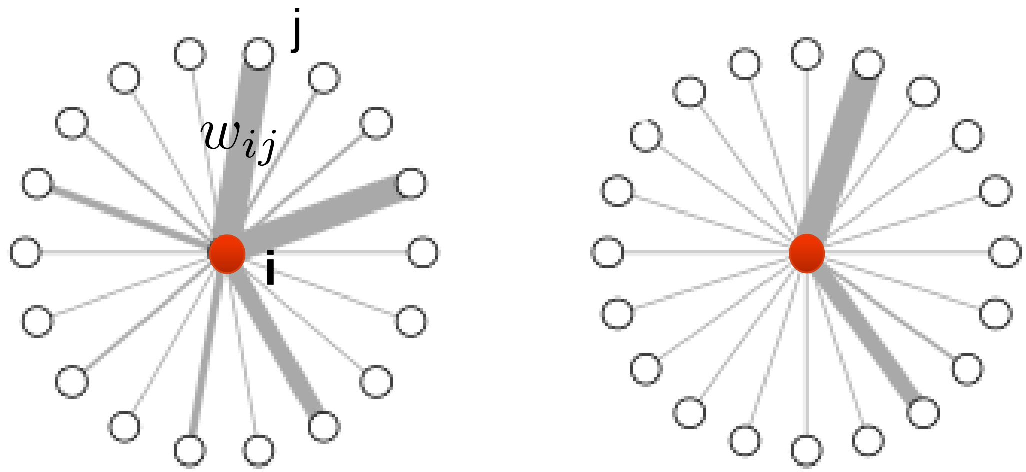




- OK, para redes sin peso vale. ¿Y en redes con pesos? ¿Qué ocurre con nodos con el mismo grado?

Diversidad de conexiones (redes con pesos)

- **Segunda definición**: si el grosor de cada enlace representa su fuerza o intensidad (peso), ¿cuál pensáis que tiene una red más diversa?



Podemos medirlo mediante la entropía de Shannon:

$$S_i = -\sum_{j \in N_i} p_{ij} \log_2(p_{ij}) \quad \text{donde} \quad p_{ij} = \frac{w_{ij}}{\sum_{j \in N_i} w_{ij}}$$

Diversidad de conexiones (redes con pesos)

- Analicemos los dos casos extremos para ver cómo se comporta este valor. $$S_i = -\sum_{j \in N_i} p_{ij} \log_2(p_{ij})$$

- Si toda la fuerza se concentra mayoritariamente en un solo enlace, la probabilidad de dicho enlace es casi uno y, por tanto, el logaritmo se anula, así que la entropía es 0.

- En cambio, si distribuimos uniformemente la intensidad entre todos los enlaces (caso más diverso):

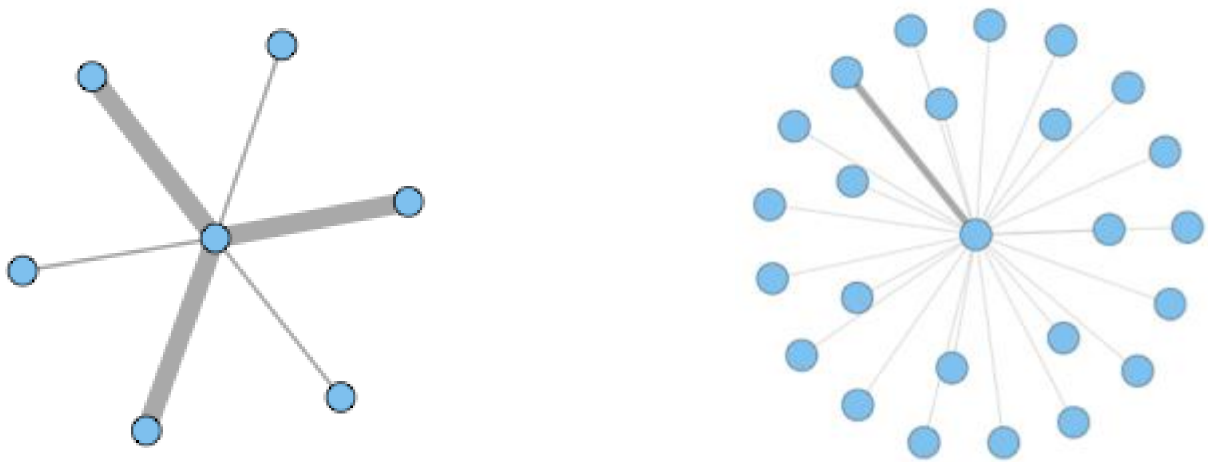
$$p_{ij} = \frac{1}{N_i} \Rightarrow$$
$$S_i = -\frac{1}{N_i} \sum_j \frac{1}{N_i} \log_2 \frac{1}{N_i} = \log_2 N_i$$

- Es decir que el valor máximo de la entropía depende del grado del nodo que estemos analizando.

- Esto es un problema ya que en los grafos no todos los nodos tienen el mismo grado y parece existir algo parecido a un factor de escala a la hora de medir la entropía.

Diversidad de conexiones (redes con pesos)

- **Tercera definición**: Por tanto, si queremos analizar este caso, para ser justos, tendremos que normalizar por el valor que hemos obtenido en la fórmula anterior:



Medimos la entropía normalizada, que es un valor entre 0 y 1, mediante:

$$S_i^{norm} = \frac{S_i}{\log_2 N_i}$$

Resumen de métricas (hasta el momento)

Nodos	Enlaces	Pares de nodos	Comunidades	Grafo
Grado	Peso	Distancia sin pesos		Componentes conexas
Clustering local	Centralidad	Distancia con pesos		Densidad
Centralidad (4)				Camino medio
Diversidad				Diámetro
				Reciprocidad
				Clustering global

Resumen de métricas (hasta el momento)

Nodos	Enlaces	Pares de nodos	Comunidades	Grafo
Grado	Peso	Distancia sin pesos		Componentes conexas
Clustering local	Centralidad	Distancia con pesos		Densidad
Centralidad (4)				Camino medio
Diversidad				Diámetro
				Reciprocidad
				Clustering global

2 Detección de comunidades

Introducción

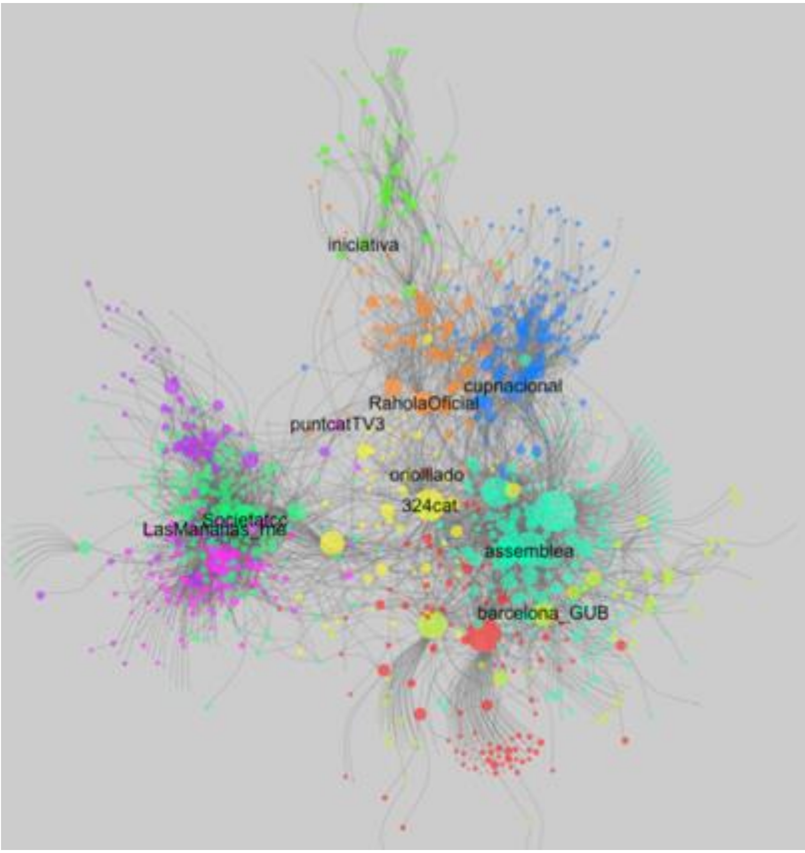
- Una definición informal de comunidad es un **conjunto de nodos** del grafo que **están densamente conectados entre ellos** y menos con el resto de la red.

- Existen múltiples algoritmos para detectar comunidades, dependiendo de cómo entendamos esta definición:
  - Basados en similaridad estructural
  - Basados en betweenness
  - Mediante detección de patrones como cliques
  - Y, sobre todo, de optimización de la modularidad.

- Cada uno da un resultado diferente, partiendo los grafos del nodo en grupos distintos. Por ello, a cada resultado de una ejecución de este algoritmo lo llamaremos un partición del grafo, es decir, cada posible agrupamiento de los nodos del grafo.

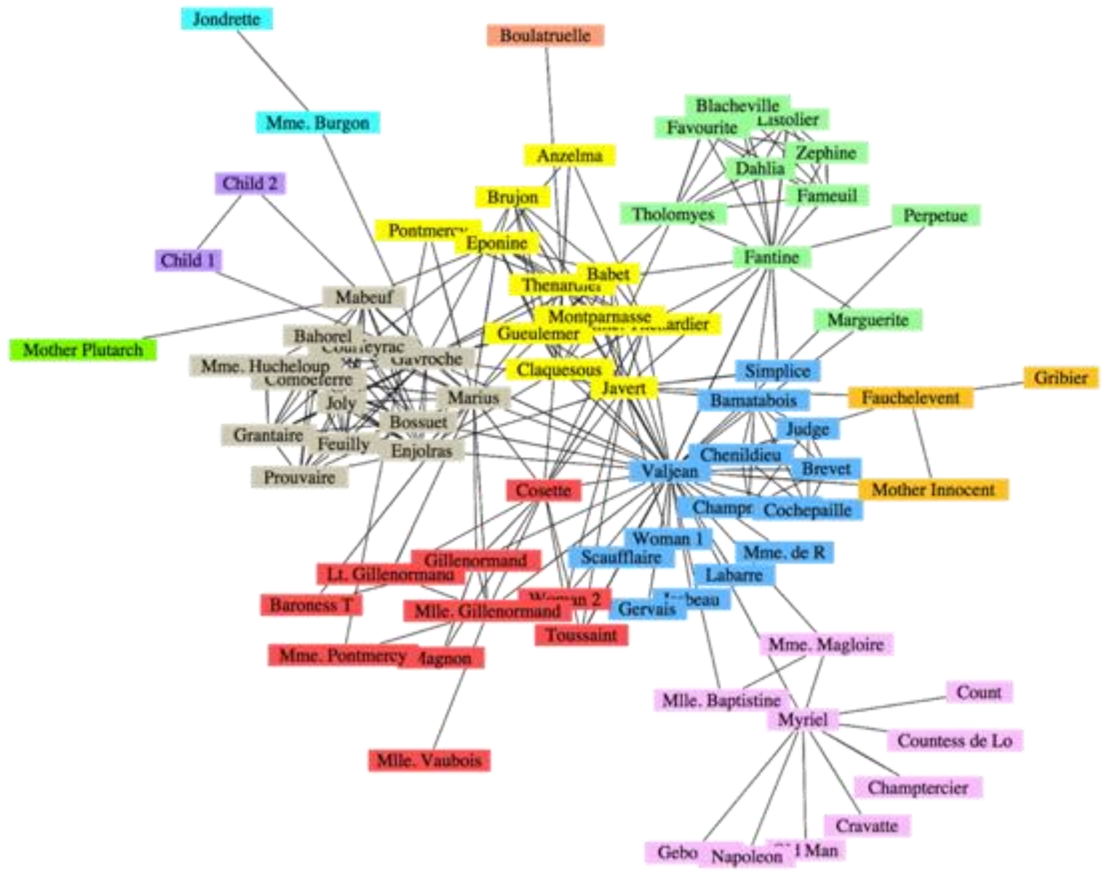
Ejemplos

- **Conversación política en Twitter**



Ejemplos

- **Interacción de los personajes de Los Miserables de Victor Hugo**



Algunas consideraciones

- En general, independiemententemente de la naturaleza del grafo, para detectar comunidades **lo asumiremos como no dirigido** ya que es muy complejo encontrar comunidades en grafos dirigidos (por qué?).
- Existen algoritmos que analizan la pertenencia a comunidades de una forma más “soft”, dando pertenencias de los nodos a varias (es decir, hay **Solapamiento de Comunidades**).
- En general, las comunidades pueden ser muy **heterogéneas**: algunas con pocos nodos, otras muy grandes, etc.
- Detectar comunidades es útil como paso previo a la **detección de influencers** locales.

Densidad inter e intra comunidad

- Dada una comunidad, ¿cómo sabemos si realmente lo es?
- Yéndonos a la definición informal que hemos dado antes, dos primeras medidas interesantes pueden ser la densidad intra e inter comunidad:

$$\delta_{int}(C) = \frac{\text{number of internal edges in C}}{\text{number of all possible edges in C}}$$

$$\delta_{ext}(C) = \frac{\text{number of inter-cluster edges in C}}{\text{number of all possible edges involving C and the rest of the graph}}$$

Algoritmos de clustering de ML

- **Primera opción**: utilizar lo que ya sabemos. ¿Por qué no utilizar un algoritmo como Kmeans, o sus variantes, para este problema?

- Podemos interpretar cada fila de la matriz de adyacencia como un "individuo" que clusterizar, donde si tiene enlaces a otro vecino o no, es cada una de sus características.

- ¿Qué problema le encontráis a esta aproximación?

Similaridad estructural entre nodos

- **Segunda opción**: nos empeñamos en utilizar algo que conocemos, pero creamos medidas de distancia basadas en las propiedades estructurales de los nodos.

1. Similaridad estructural: distancia euclídea sobre las filas de la matriz de adyacencia.

$$d_{ij} = \sqrt{(\sum_{k \neq i,j} (A_{ik} - A_{jk})^2)}$$

2. Solapamiento de vecindades: medir si dos nodos comparten los mismos vecinos.

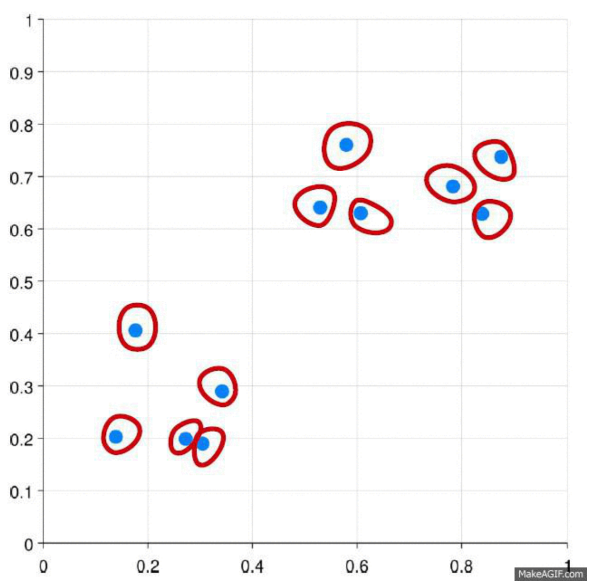
$$w_{ij} = \frac{|N_i \cap N_j|}{|N_i \cup N_j|}$$

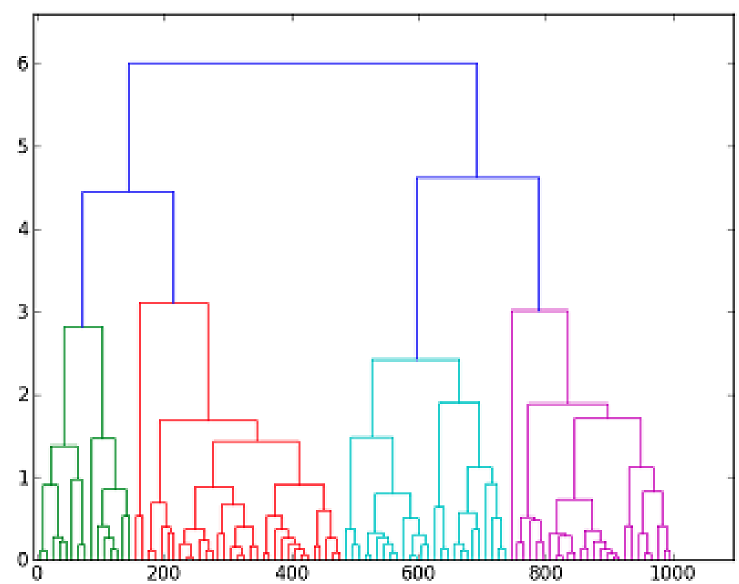
Número de vecinos compartidos

Todos los posibles vecinos entre los dos

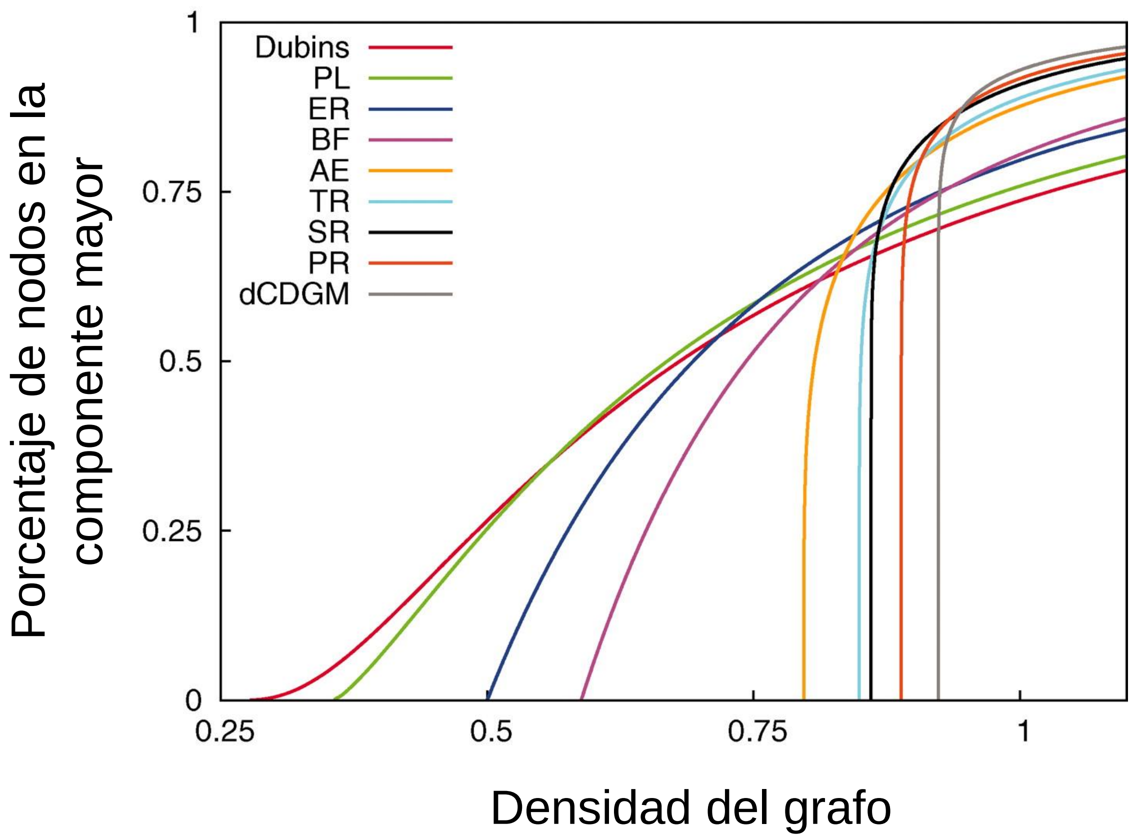
Clustering jerárquico

1. En el primer paso, cada nodo es un cluster.
2. En el siguiente paso, unimos aquellos pares más cercanos.
3. Recalculamos distancias de clusters.
4. Volvemos al paso 2.





Un fenómeno muy raro: la percolación



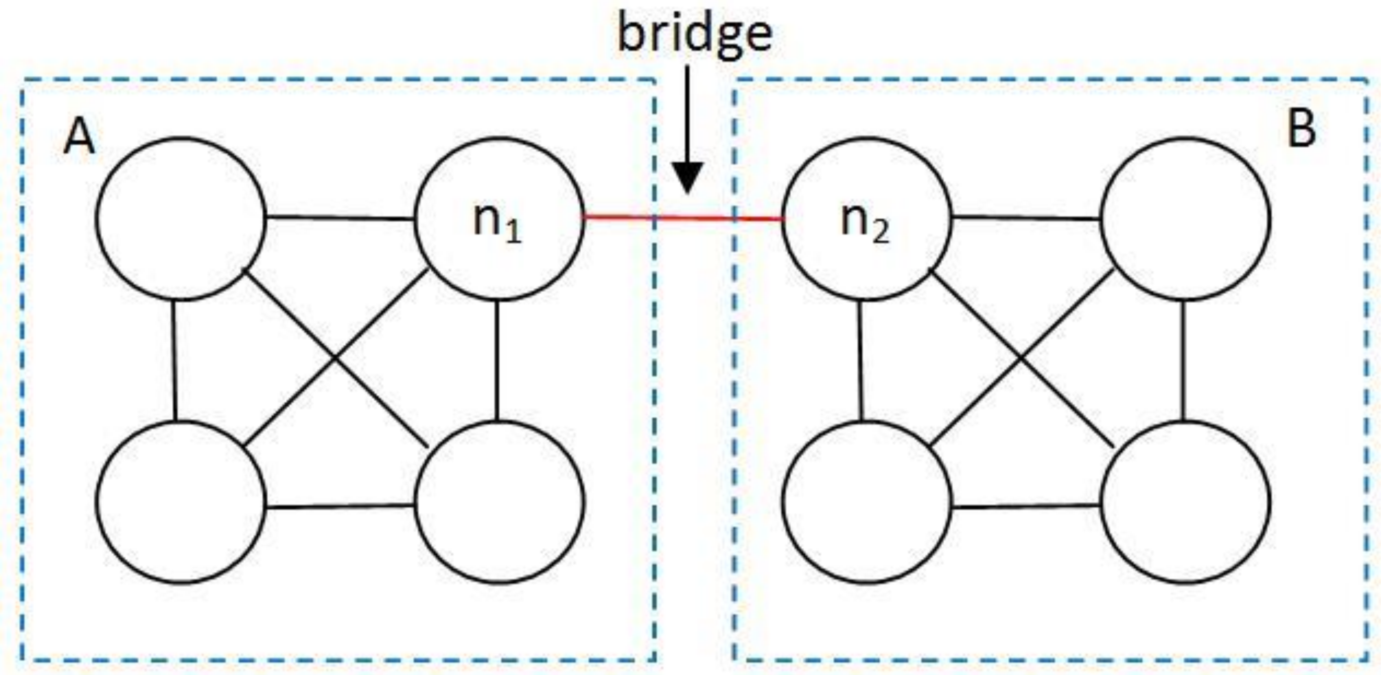
Una tipología de fenómenos que se aprecian en las redes sociales, y que pueden verse en redes aleatorias solamente por efecto de la densidad del grafo, son los **fenómenos de percolación.**

Llamamos así cuando algo se produce de forma abrupta en forma de escalón.

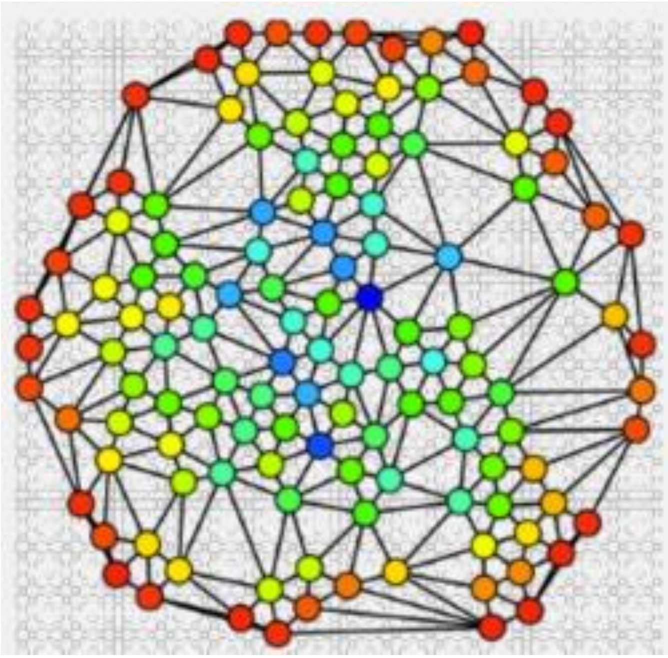
Esto es lo que sucede durante el crecimiento de un grafo cuando, por ejemplo, se van añadiendo enlaces a un conjunto inicial de nodos.

¿Quiere decir esto que, si yo parto de una red formada, y elijo de forma inteligente qué enlaces o nodos quitar, puedo romper la red sin tener que destruir muchos de ellos?

Y volviendo a un ejemplo conocido…



Usando betweenness para comunidades



The Girvan-Newman algorithm

- Calculamos la betweenness para detectar los enlaces con mayor valor.
- Eliminamos los enlaces con mayor centralidad.
- Calculamos el número de componentes conexas. Si sigue habiendo muchos nodos en la componente conexa mayor, iteramos.
- Por percolación, sabemos que, en unos pocos pasos, partiremos la red en varias componentes conexas, lo que previamente eran sus comunidades.
- **Gran problema**: para grafos grandes es inasumible.

Modularidad

$$Q = \frac{1}{2m}\sum_{i,j}(A_{ij} - \frac{k_i k_j}{2m})\delta(c_i, c_j)$$

1 si existe un enlace entre i y j

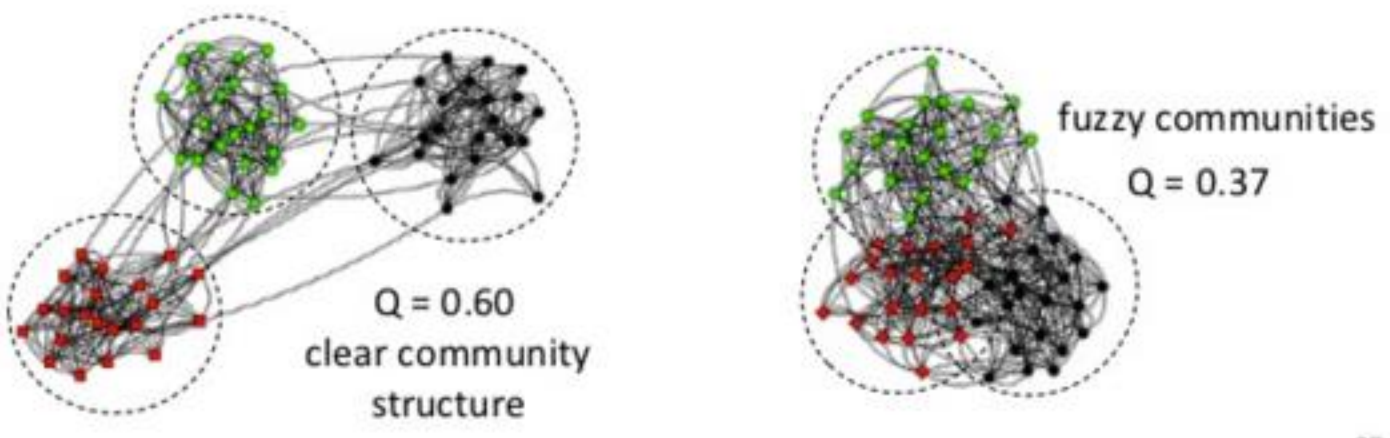
Probabilidad de encontrar un enlace entre i y j en un grafo aleatorio

1 si i y j pertenecen a la misma comunidad; 0 si no.

- Sumamos este valor para todas las posibles parejas del grafo.
- Es una medida de bondad sobre la **partición**.
- Útil para comparar particiones y resultados de algoritmos.

Modularidad

- Se puede comprobar que en un grafo aleatorio, Q=0.
- Típicamente en grafos no aleatorios, la modularidad varía entre 0.3 y 0.7.



Optimización de la Modularidad

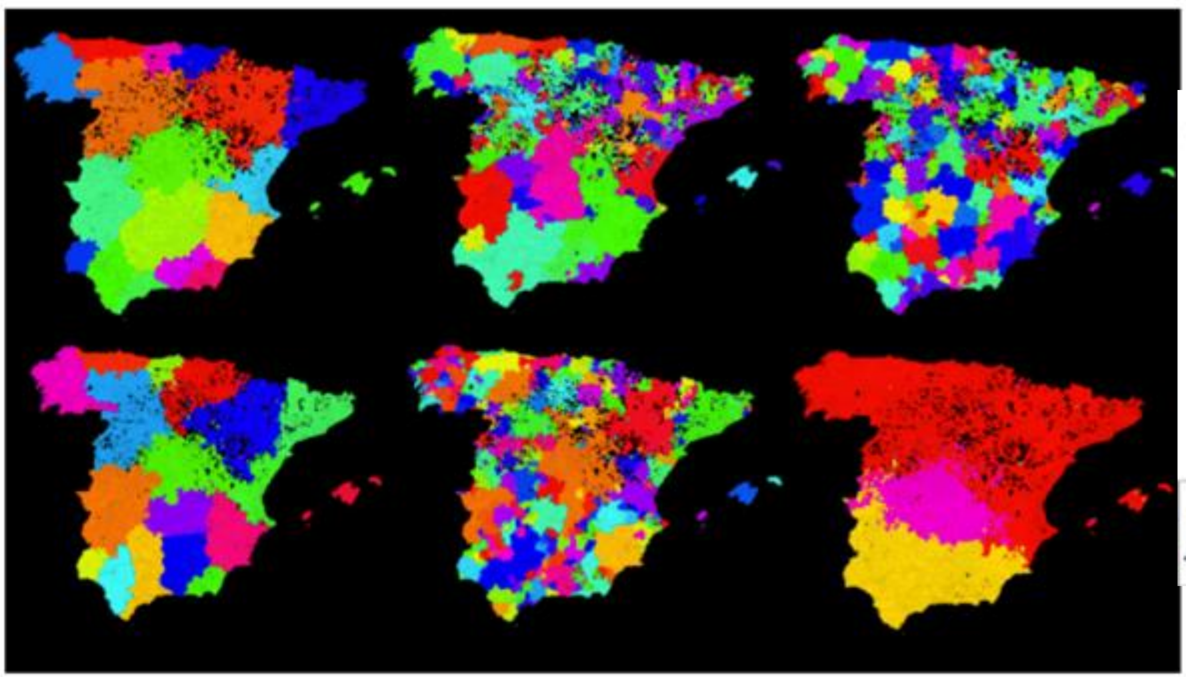
- Calcular betweenness era costoso, ¡pero la modularidad es muy sencillo!
- Busquemos optimizarla ya que probar todas las posibles particiones también es extremadamente costoso.

- Newman (2003) diseñó una heurística para encontrar máximos locales de la modularidad de forma aglomerativa (similar al clustering jerárquico).

    1. Todos los nodos comienzan siendo clusters independientes.
    2. Para cada par de clusters, calculamos su modularidad.
    3. Combinamos aquellos clusters que maximizan la modularidad.
    4. Repetimos desde el paso 2.

Resultados entre algoritmos

- Existen infinidad de algoritmos de optimización de la modularidad (Louvain, Infomap, Label Propagation, Walktrap,…). ¿Con cuál nos quedamos?



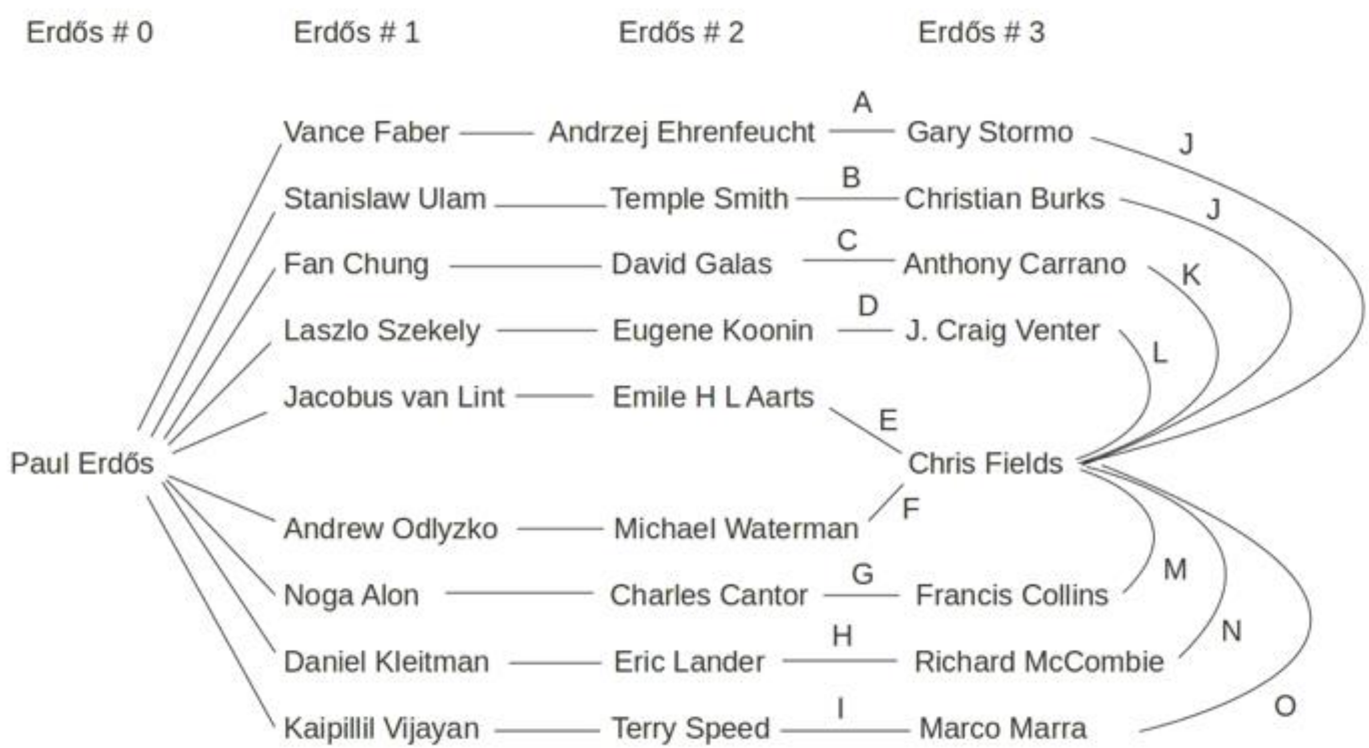
Resultados entre algoritmos

- Un método muy útil para comparar resultados de algoritmos de detección de comunidades o para comparar tu partición con una partición a priorri (por ejemplo, departamentos o provincias) es la **Información Mutua Normalizada**.

- En R, dentro de Igraph, existe una función directamente llamada compare.communities, que devuelve un valor entre 0 y 1 (0 es que no son similares y 1 es que son exactamente la misma partición).

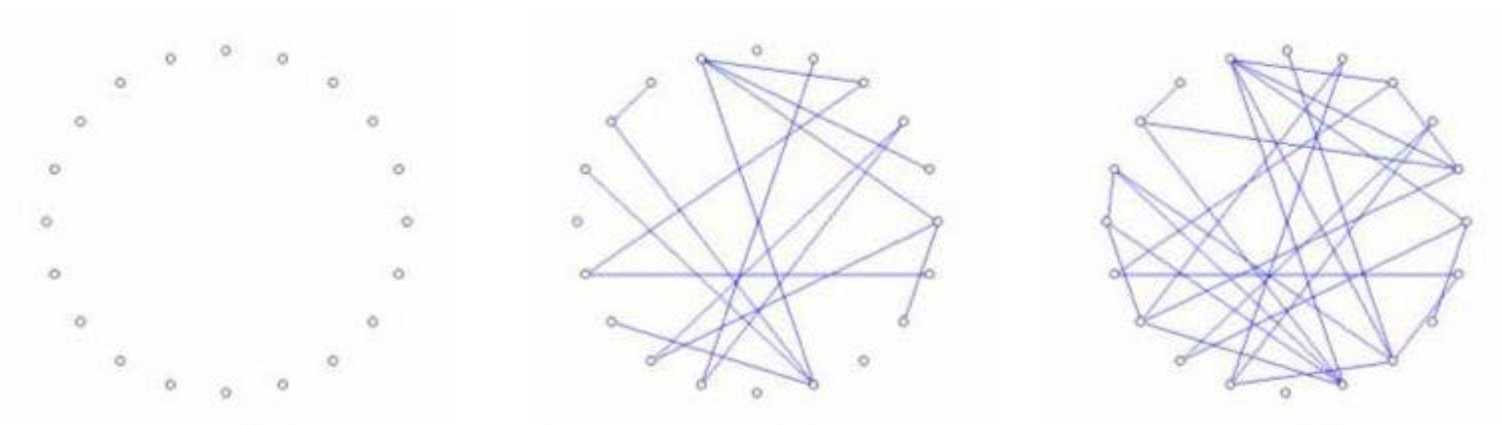
3 Modelos y redes aleatorias

El número de Erdos



El modelo de Erdos-Renyi

- Desarrollado por Erdos & Renyi in 1959.
- El propósito de este modelo es generar redes aleatorias y analizar sus propiedades.
- Se define con dos parámetros: el número de nodos y la probabilidad de conexión entre cualquier pareja de nodos.



$p = 0$ (a) $p = 0.1$ (b) $p = 0.2$ (c)

El modelo de Erdos-Renyi

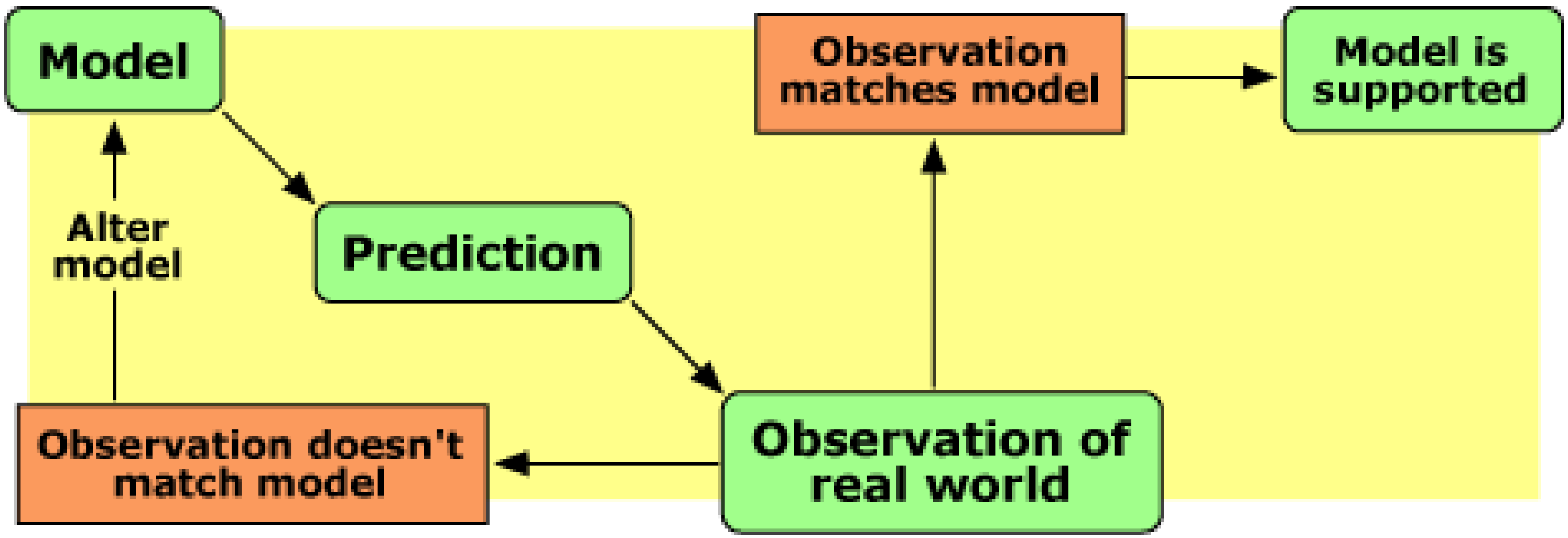
- ¿Por qué analizar este modelo tan simple?
- Algunas conclusiones que sacaron E&R al analizar este tipo de grafos:
  - Si n*p = 1 => el tamaño de la CC mayor es, aproximadamente, $n^{2/3}$
  - If n*p >> 1 => es seguro que el grafo tendrá una componente gigante.

Adicionalmente, si $p > \frac{\log n}{n}$ entonces el grafo se muy probable que sea conexo.

Precisamente, este es el fundamento teórico detrás de la percolación: surge directamente como un fenómeno necesario de la densidad de un grafo.

Relación entre modelos y realidad

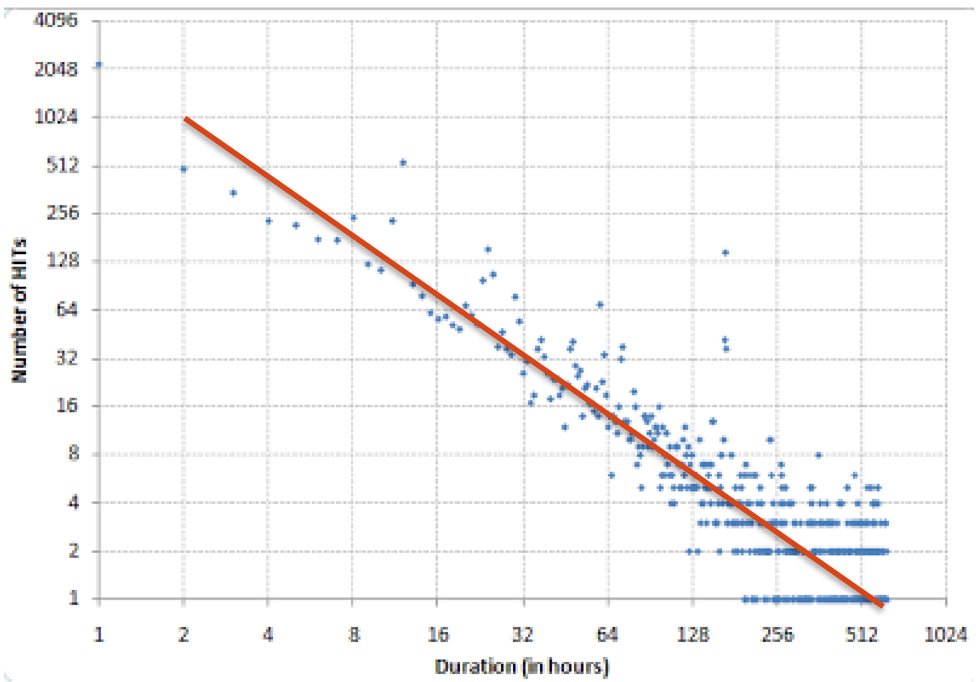
En ocasiones, uno empieza aquí…



Y a veces, uno empieza aquí

Distribución Power Law

- Es una generalización de la conocida regla de Pareto 80-20.
- Es una distribución que no solamente aparece en redes sociales, sino en muchísimos fenómenos económicos, de análisis de tiempos,…



Decimos que una variable aleatoria sigue una distribución PL si su función de densidad es de la forma:

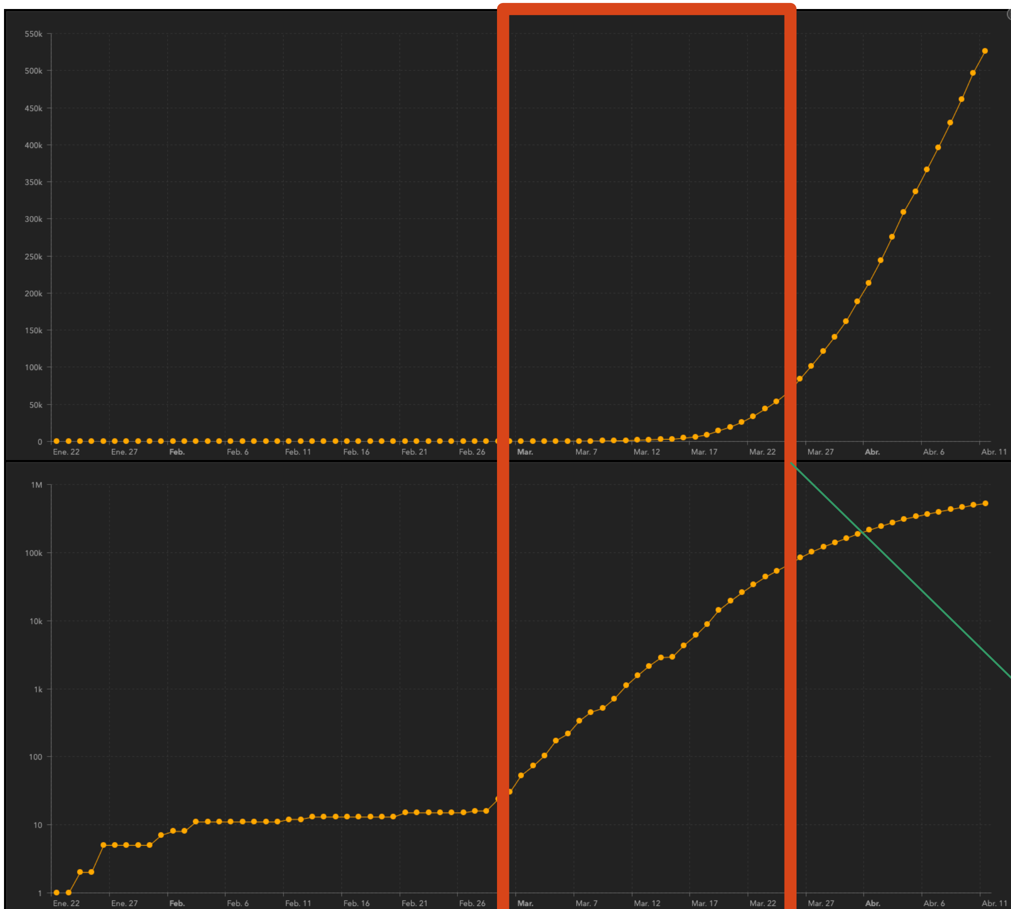
$$p(x) \sim x^{\alpha}$$

Si nos fijamos, tomando logaritmos en ambos lados:

$$\log p(x) \sim \alpha \log(x)$$

Por lo que esta distribución, en escala logarítmica XY, debe aparecer como una recta.

Está muy de moda pintar en escala logarítima



Hemos dicho que, para que aparezca una recta, en una PL pintamos en escala logarítimica tanto eje Y como eje X.

Sin embargo, en los gráficos que podemos ver relacionados con el COVID19, solo se pone la Y en logarítmica.

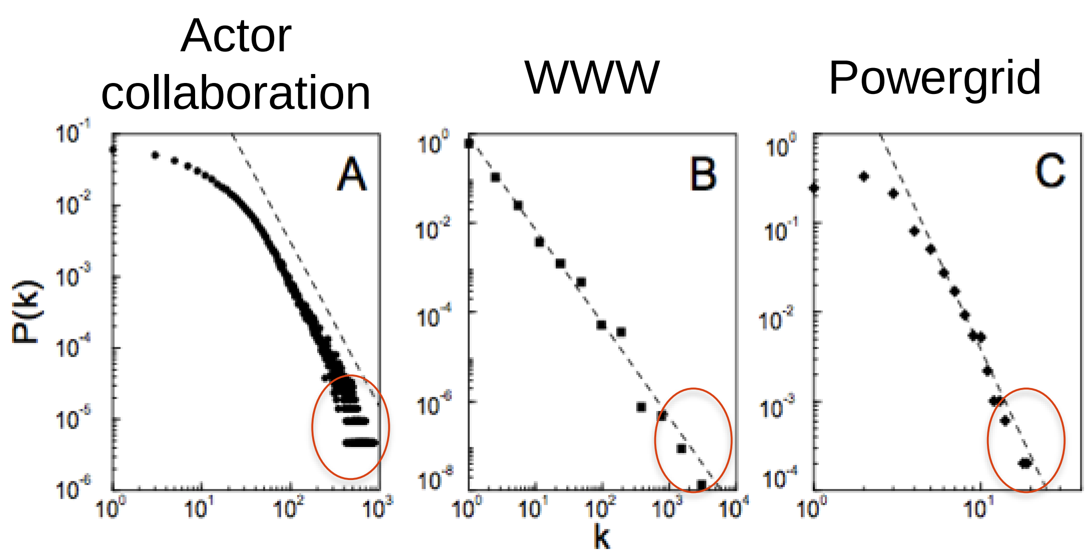
$$I(t) \propto A \exp(\beta t)$$

$$\log I(t) \propto \log A + \beta t$$

Período exponencial de la propagación (el más peligroso).

Redes libres de escala

- Barabasi y Albert estudiaron la distribución del grado en diferentes redes, observando que, en muchos contextos, esta distribución sigue una Power Law.



- Muy pocos nodos con un grado muy alto (los “hubs de la red”).
- Ocurre (casi) siempre: redes sociales, de colaboración, de actores, el world wide web,…
- Tiene características de redundancia: los hubs están conectados con hubs más pequeños.

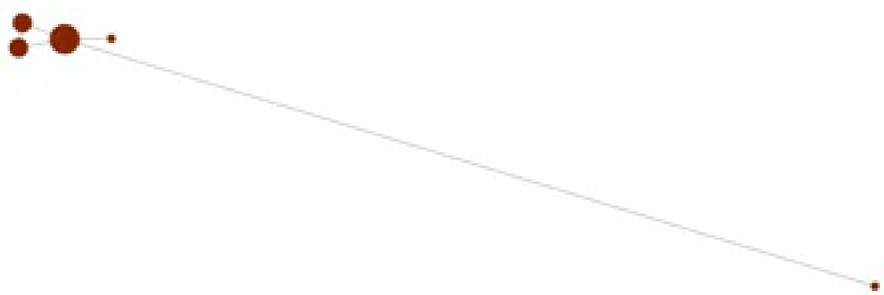
El modelo de Barabasi - Albert

- ¿Por qué surge esta estructura en las redes del mundo real? En un grafo de E&R, la distribución es una binomial, muy distinta de una PL. ¿Hay algún mecanismo que fuerce a esta distribución?

- Pensaron que la evolución de una red no es igualitaria. Cuando un nuevo nodo se incorpora a una red y comienza a crear enlaces es más probable que lo haga con nodos que ya tenían un grado muy alto. Este es el mecanimso de Preferential Attachment.

- Probaron matemáticamente que este tipo de evolución de red produce grafos que exhiben estas características en cuanto a su distribución de grados (surgen redes libres de escala).

El modelo de Barabasi - Albert



El experimento de Milgram

- Experimento llevado a cabo por un importante psicólogo social durante el siglo XX.

- Encontró que el número de pasos necesarios entre cualquier pareja aleatoria de personas está, de media, entre 5.5 y 6.

- Pequeño detalle: solamente el 22% de las cartas llegaron a su destino. El resto, se pararon en algún punto de la cadena.

Redes de mundo pequeño

- La definición formal de una red de mundo pequeño (small-world network) dice que la distancia media entre dos nodos crece proporocionalmente al logaritmo del número de nodos:

$$L \propto \log N$$

- Desde un punto de vista interpretativo: el hecho de añadir nodos no incrementa, al mismo ritmo, el camino medio entre nodos.

- Además de este hecho, se impone que las redes de mundo pequeño también deben tener un clustering alto.

Table 1 Empirical examples of small-world networks

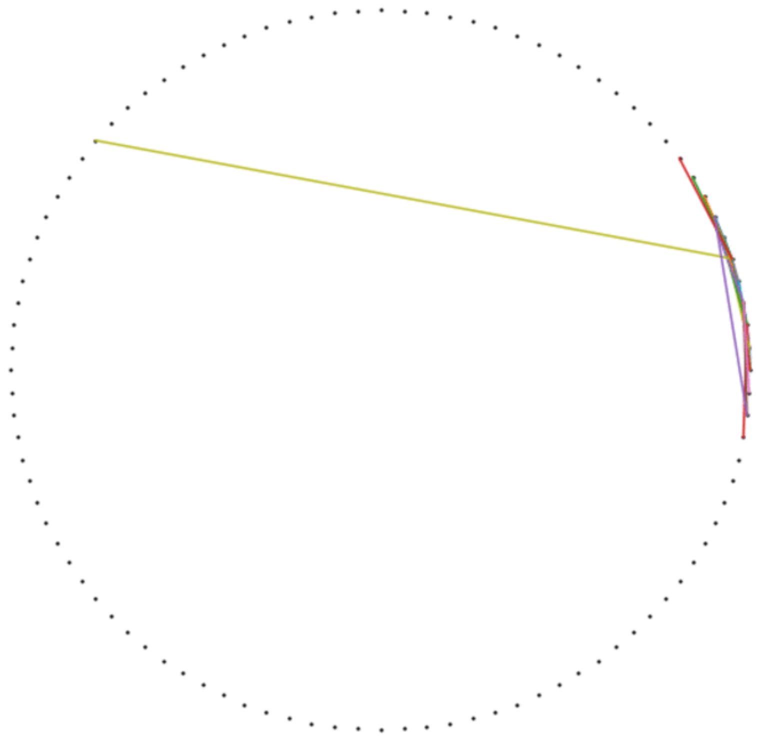
	L_{actual}	L_{random}	C_{actual}	C_{random}
Film actors	3.65	2.99	0.79	0.00027
Power grid	18.7	12.4	0.080	0.005
C. elegans	2.65	2.25	0.28	0.05

El modelo de Watts & Strogatz

- El modelo de WS se basa en tres parámetros:
  - El número de nodos N
  - El grado medio K.
  - La probabilidad de reconexión p.

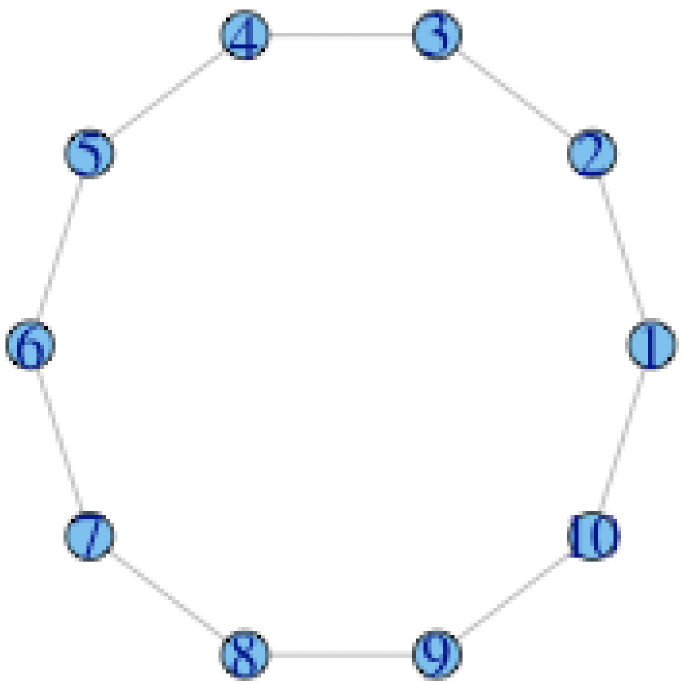
- Inicialmente, conectamos cada nodo con K vecinos de forma que todos tienen grado K pero, alguno de ellos, decidimos reconectarlo aleatoriamente a otro nodo, con probabilidad p.

El modelo de Watts & Strogatz



El modelo de Watts & Strogatz

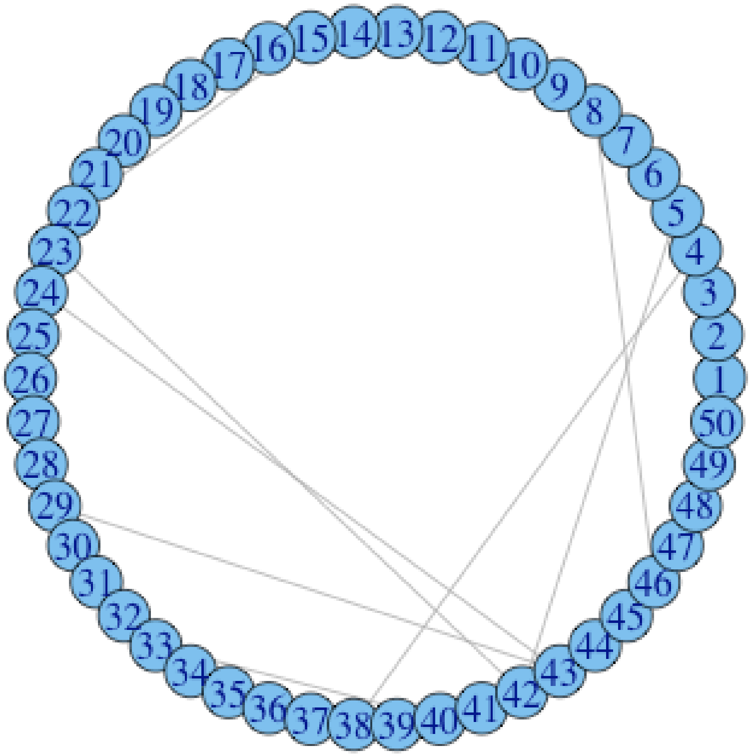
- **Primer ejemplo**: N=10, K=2, p=0.



- Establecemos p=0, lo que hace que no se acorten caminos ni se cierren triángulos.
- Uno podría pensar que es de mundo pequeño ya que log(10) =2.3 y el camino medio aquí es L=2.78.
- Sin embargo, si hacemos grande esta estructura con N=100 obtenemos que L=25.25, muy alejado de log(100)=4.6.

El modelo de Watts & Strogatz

- **Segundo ejemplo**: N=50, K=2, p=0.1.



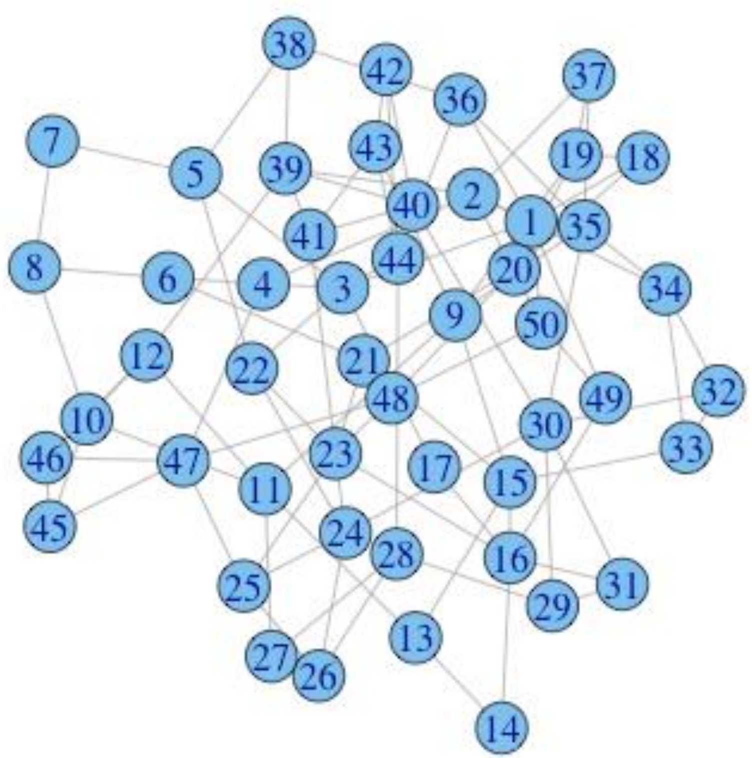
- Añadimos una pequeña probabilidad de reconexión, 0.1.

- De nuevo, encontramos que el clustering es 0, no hemos podido cerrar tripletas.

- Sin embargo, el camino medio es L=8.99, aún lejos de log(50)=3.59 pero ha bajado de 12.8 que habríamos tenido con p=0.

El modelo de Watts & Strogatz

- **Third example**: N=50, K=4, p=0.2.



- Aumentamos tanto el grado medio como la probabilidad de reconexión como el grado.
- Obtenemos un camino medio L=2.96!!
- Y lo que es más interesante, ahora el coeficiente de clustering es 0.26!!
- Es una red de mundo pequeño ☺

Medida de mundo pequeño

$$\sigma = \frac{C/C_{rand}}{L/L_{rand}}$$

Podemos medir si una red es de mundo pequeño con este KPI: si este número es mayor que 1, diremos que la red tiene la propiedad de mundo pequeño.

¿Cómo se relacionan las redes de mundo pequeño con las redes libres de escala?

Se puede demostrar que las redes libres de escala tienen un camino medio aceptable con la definición de mundo pequeño. De hecho, crece como

$$L \sim \log \log N$$

Sin embargo, no presentan clustering, así que no podemos decir que sean de mundo pequeño.

4 Ejercicio práctico

Ejercico práctico

1. Cargar un grafo desde fichero.
2. Realizar el análisis de la primera sesión para afianzar conocimientos.
3. Detectar comunidades y averiguar qué algoritmo da mayor modularidad.
4. Detectar el nodo más central de cada comunidad según cada medida.
5. Visualizar las comunidades y su nodo más central.