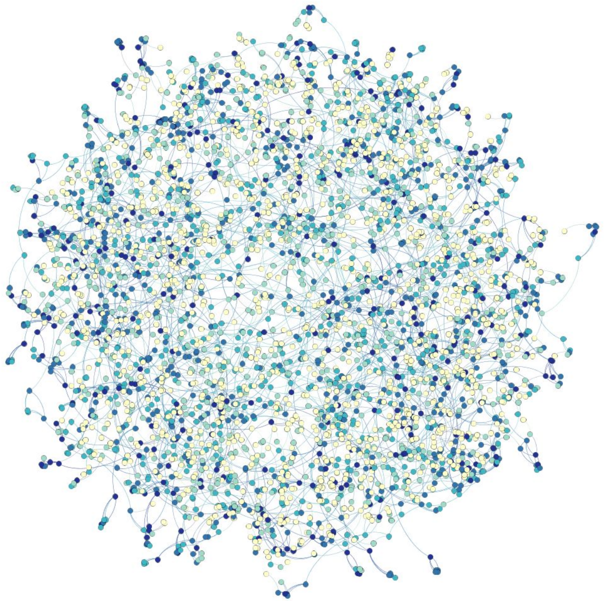
Redes (2): Caracterización de redes complejas.

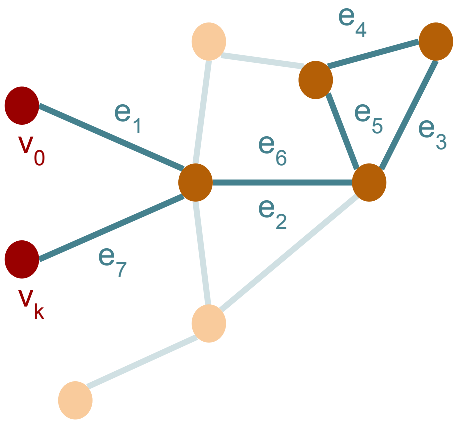
Propiedades

Propiedades cuantitativas para analizar y entender cómo están organizadas las redes.

- Distribución de grado
- Distancias
- Coeficiente de clustering
- Centralidad
- Robustez
- Asortatividad
- Comunidades

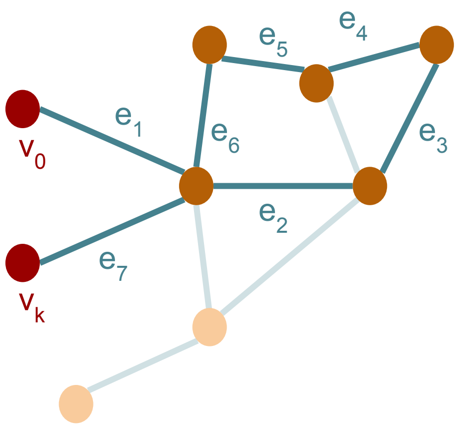


Recorridos y caminos

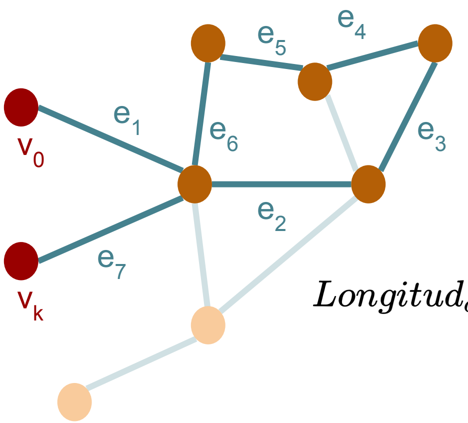


Un **recorrido (walk)** es una secuencia finita o infinita (e_1, e_2, e_3, … , e_{k-1} , e_k) donde los nodos terminales de e_i son a su vez terminales de e_{i-1} y e_{i+1}. Dicho recorrido tiene siempre dos nodos terminales v_0 y v_k.

Un **camino (path)** es un recorrido en el que <u>los enlaces son todos diferentes</u>.



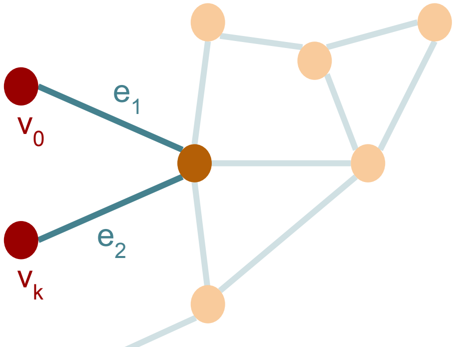
Longitudes y distancias



$$Longitud_{o,k} = 7$$

La **longitud** de un recorrido o camino está dada por el número de enlaces que contienen.

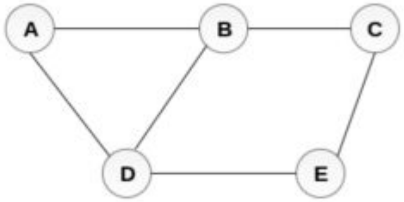
El **camino más corto (shortest path)** entre dos nodos es un camino con longitud mínima. Puede haber más de uno.

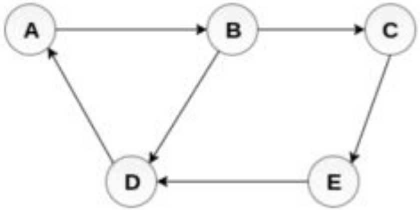


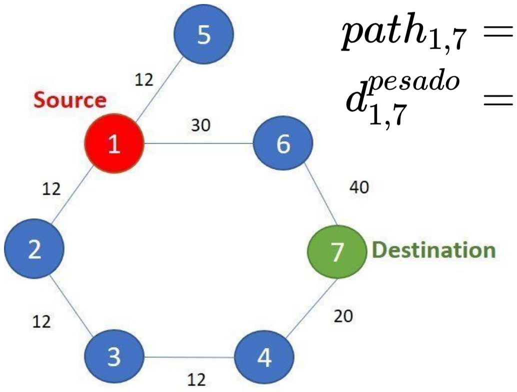
$$Longitud_{o,k} = 2$$

Longitudes y distancias

La **distancia** entre dos nodos es la longitud del camino más corto entre ellos.

<table>
<tr><th>Redes no dirigidas</th><th>Redes dirigidas</th><th>Redes pesadas</th></tr>
<tr>
<td>

$d_{A,E} = d_{E,A} = 2$</td>
<td>

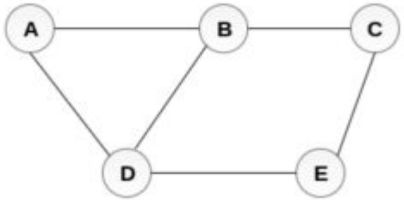
$d_{A,E} = 3$
$d_{E,A} = 2$</td>
<td>

$path_{1,7} = \{1, 2, 3, 4, 7\}$
$d_{1,7}^{pesado} = 56$</td>
</tr>
</table>

Longitudes y distancias

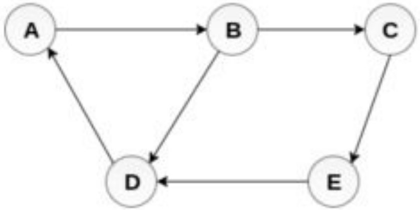
La **distancia** entre dos nodos es la longitud del camino más corto entre ellos.

Redes no dirigidas



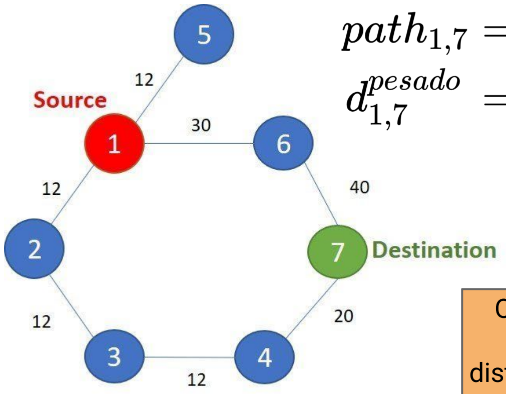
$$d_{A,E} = d_{E,A} = 2$$

Redes dirigidas



$$d_{A,E} = 3$$
$$d_{E,A} = 2$$

Redes pesadas



$$path_{1,7} = \{1, 2, 3, 4, 7\}$$
$$d_{1,7}^{pesado} = 56$$

Ojo! si los pesos son similaridades y no distancias, voy a tratar de maximizar. En este caso haría 1->6->7

Distancias para toda la red

El **diámetro** (*D*) de un un grafo es la máxima distancia entre cualquier par de nodos:

$$D = max_{uv}(\, d_{uv}\,)$$

La **distancia media** o **longitud característica** de un grafo es la distancia promedio entre todos los pares de nodos:

$$d = < d_{uv} >$$

La **eficiencia global** de un grafo es el promedio de la inversa de las distancias entre todos los pares de nodos:

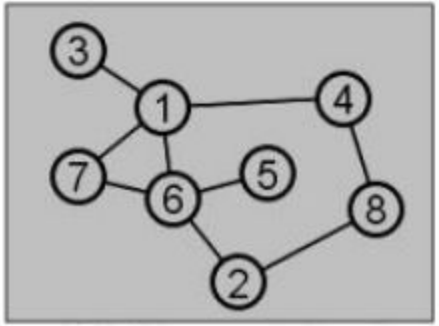
$$eff =< \frac{1}{d_{u,v}} >$$

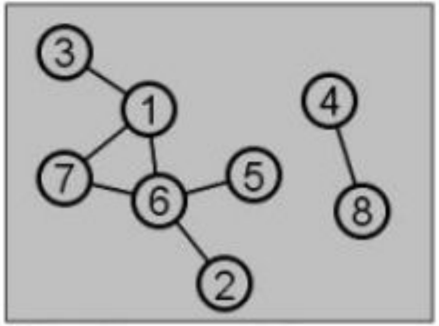
Particularmente útil para pares de nodos desconectados

Conectividad

- Un grafo es **conectado (conexo)** si existe un camino entre cualquier par de nodos.

- Un grafo G puede estar compuesto de varias **componentes conectadas**.





En este caso, se analiza la mayor componente conectada (“**componente gigante**”)

Coeficiente de clustering

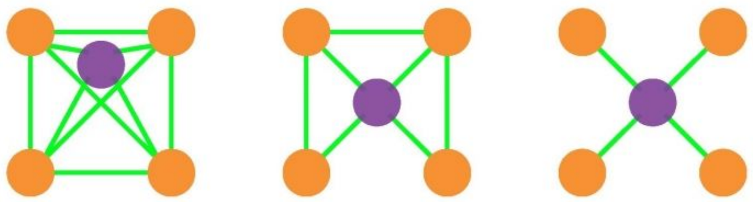
Medida de densidad local

¿En qué medida los primeros vecinos de un nodo *i* son vecinos entre sí?

- C_i es la probabilidad que los primeros vecinos se conecten entre sí.

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

L_i pares de vecinos enlazados



C_i=1 C_i=1/2 C_i=0

- C_i = 0 si los vecinos no se conectan entre sí.
- C_i = 1 si todos los vecinos se conectan entre sí (los vecinos del nodo *i* forman un “grafo completo” o “clique”).

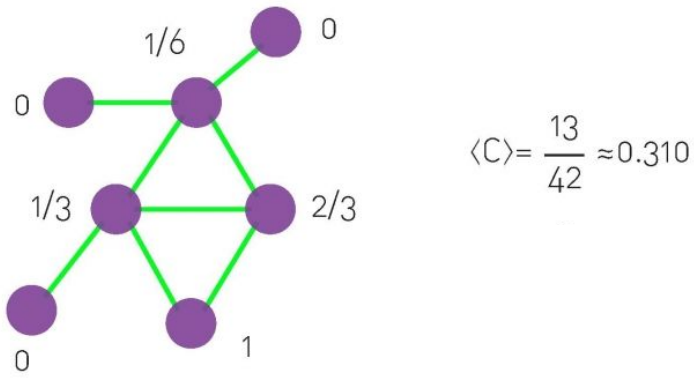
Coeficiente de clustering

Medida de densidad local

¿En qué medida pares de vecinos de un nodo tomado al azar en la red son vecinos entre sí ?

- *Coeficiente de clustering medio*

$$< C >= \frac{1}{N}\sum_{i}^{N} C_i$$



Coeficiente de clustering

Medida de densidad local

¿En qué medida pares de vecinos de un nodo tomado al azar en la red son vecinos entre sí ?

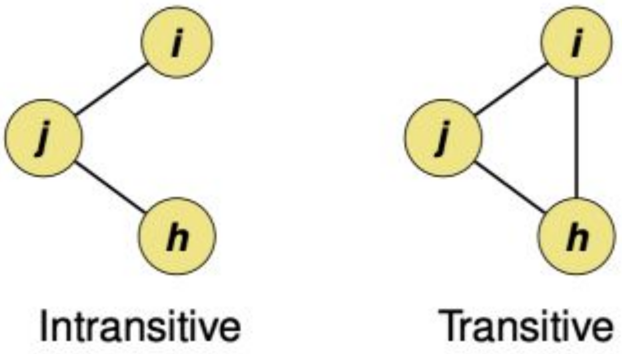
- *Coeficiente de clustering medio*

$$< C >= \frac{1}{N}\sum_{i}^{N} C_i$$

- *Coeficiente de clustering global* o *Transitividad*

Hay 3 tripletes en cada triángulo

$$C_\Delta = \frac{3 \text{ x Número de triángulos}}{Número\, de\, tripletes\, conectados}$$



Coeficiente de clustering

Medida de densidad local

¿En qué medida pares de vecinos de un nodo tomado al azar en la red son vecinos entre sí ?

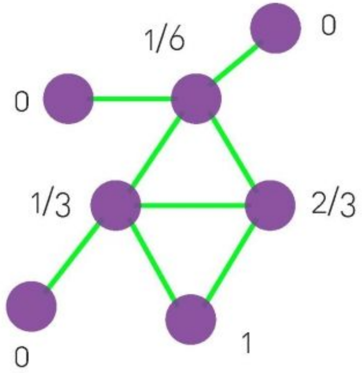
- *Coeficiente de clustering medio*

$$< C >= \frac{1}{N} \sum_{i}^{N} C_i$$

- *Coeficiente de clustering global* o *Transitividad*

Hay 3 tripletes en cada triángulo

$$C_\Delta = \frac{3 \text{ x Número de triángulos}}{\textit{Número de tripletes conectados}}$$



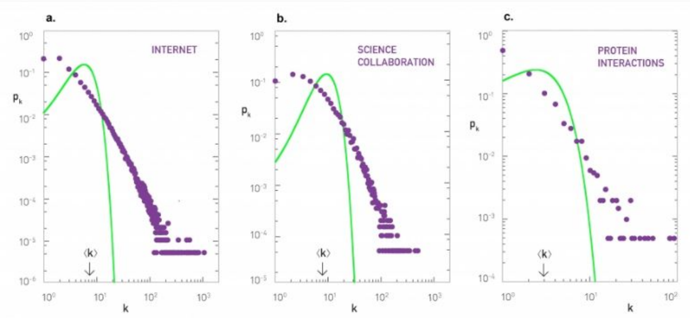
$$\langle C \rangle = \frac{13}{42} \approx 0.310$$

$$C_\Delta = \frac{3}{8} = 0.375$$

Algunos tipos de redes

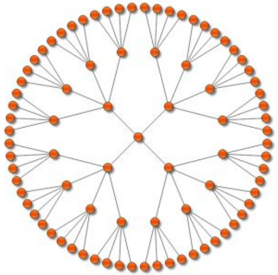
- **Redes Scale-free:**

Distribución de grado tipo power law
Hubs -> distancias pequeñas
(Coef. de clustering varía de acuerdo a otros características de la red, pero tiende a ser bajo)



- **Redes Small World:**

Diámetro y distancia pequeñas

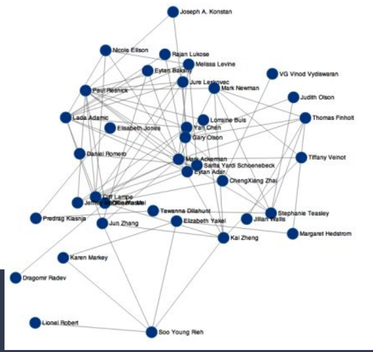


Red de 85 nodos con diámetro (D) = 6

- **Redes Transitivas:**

Coeficiente de clustering alto

Red de coautores



Propiedades de los nodos. Medidas de centralidad

Definiciones

➔ ¿Qué tan relevante es un nodo en la red?

➔ De manera abstracta la **centralidad (*C*)** es una función que asigna un valor numérico a cada nodo de una red.

➔ El nodo *i* es más central que el nodo *j* si ***C(i) > C(j)***.

Propiedades de los nodos. Medidas de centralidad

Definiciones

➔ Las medidas de centralidad sirven para caracterizar la conectividad de los nodos de una red (medidas locales o globales).

➔ Los valores de centralidad son relativos a una red: sólo son comparables dentro de una misma red.

➔ Algunas medidas de centralidad solo se pueden aplicar a redes conexas.

Propiedades de los nodos. Medidas de centralidad

Definiciones

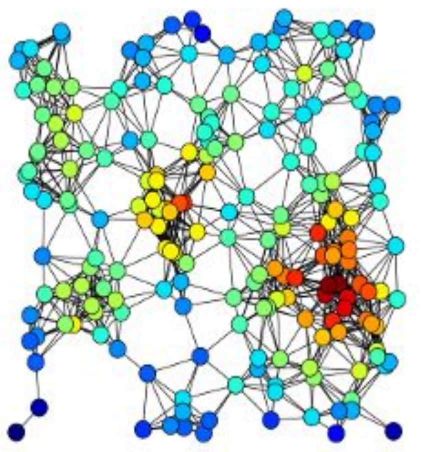
➔ ¿Qué tan relevante es un nodo en la red?

Noción de **flujo** sobre la red
Noción de **cohesividad** sobre la red

a. Grado
b. Intermediación
c. Cercanía
d. Autovalor

Propiedades de los nodos. Medidas de centralidad

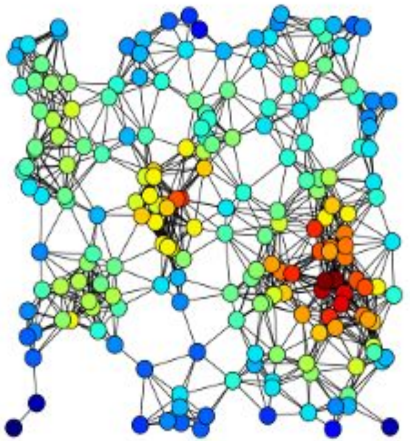
Centralidad de grado



¿Cuántos vecinos tiene el nodo en cuestión?

Propiedades de los nodos. Medidas de centralidad

Centralidad de grado



Número de vecinos:
(grado del nodo)

$$C_i^D = k_i$$

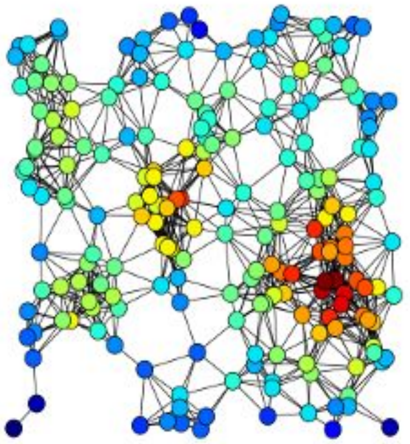
Un “hub” tiene alta centralidad de grado

Puede estar normalizado: $$\overline{C_i^D} = \frac{k_i}{n-1}$$

Para grafos dirigidos se puede considerar en forma separada los grados de entrada y de salida, dando una centralidad de entrada y una de salida.

Propiedades de los nodos. Medidas de centralidad

Centralidad de grado



- Sólo utiliza información local (primeros vecinos)

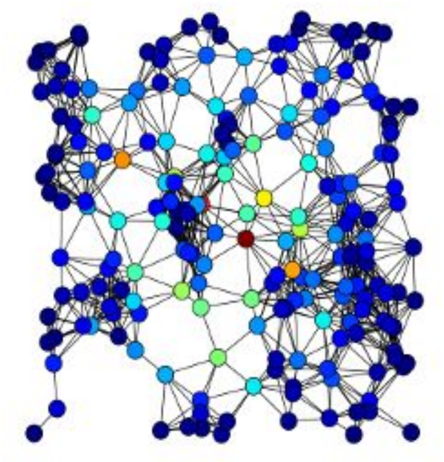
- En términos de flujo:
  - – caracteriza efectos de influencia inmediata (cercanía)
  - – prob de recibir algo que está distribuido aleatoriamente por la red es proporcional al # de contactos.

- En términos de cohesividad:
  - – Hubs proveen atajos entre pares de nodos

- Asume linealidad:
  - -un nodo con el doble de vecinos que otro es dos veces más importante

Propiedades de los nodos. Medidas de centralidad

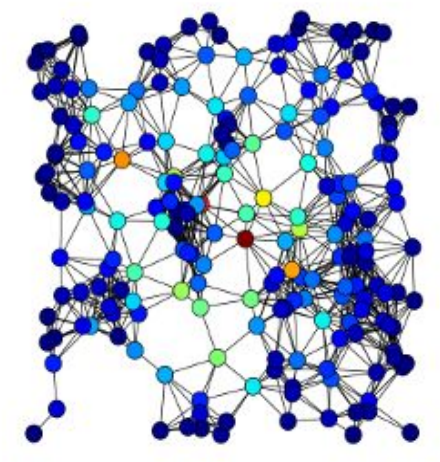
Centralidad de intermediación (betweenness)



¿Cuántos caminos mínimos dependen del nodo en cuestión?

Propiedades de los nodos. Medidas de centralidad

Centralidad de intermediación (betweenness)



Siendo *G=(V,E)* un grafo no dirigido con *i,s,t*: nodos

$$C_{bet}(i) = \sum_{s \neq i} \sum_{t \neq i} \delta_{st}(i)$$

donde:

$$\delta_{st}(i) = \frac{\sigma_{st}(i)}{\sigma_{st}}$$

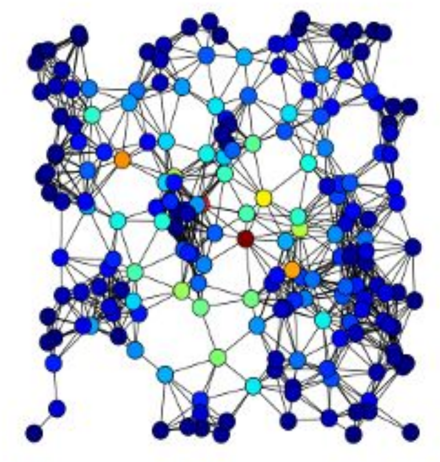
Fracción de caminos mínimos entre *s* y *t* que pasan por *i*

σ_{st} : número de **caminos mínimos** entre *s* y *t*.
$\sigma_{st}(i)$: número de **caminos mínimos** entre *s* y *t* que pasan por *i*

¡Puede haber más de uno!

Propiedades de los nodos. Medidas de centralidad

Centralidad de intermediación (betweenness)



Es una medida global

• En términos de flujo:

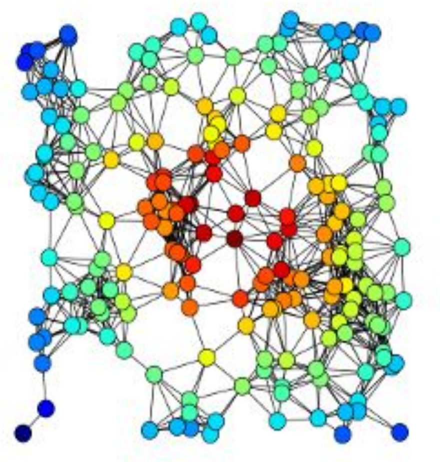
Mide la habilidad de un nodo de monitorear las comunicaciones entre otros nodos

• En términos de cohesividad:

Quitar un nodo de alta intermediación no necesariamente rompe una red, pero hace el flujo más difícil

Propiedades de los nodos. Medidas de centralidad

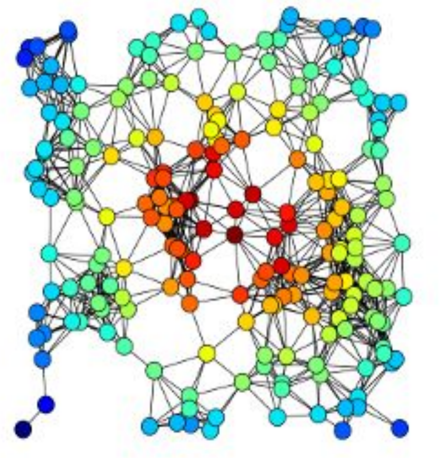
Centralidad de cercanía (*closeness*)



¿Qué tan cerca está el nodo en cuestión del resto de los nodos?

Propiedades de los nodos. Medidas de centralidad

Centralidad de cercanía (*closeness*)



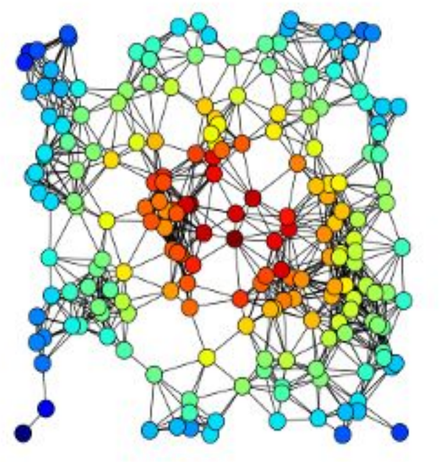
$$C_{\text{cer}}(i) = \frac{1}{N-1}\sum_{i\neq j}\frac{1}{d_{ij}}$$

i,*j*: nodos

Es otra medida basada en la **distancia**.

Propiedades de los nodos. Medidas de centralidad

Centralidad de cercanía (*closeness*)

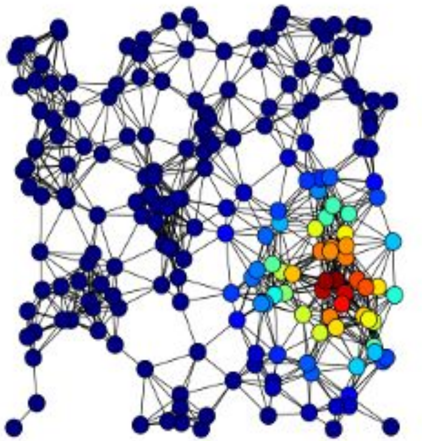


Es una medida global

- En términos de flujo:
  - Nodos que en promedio reciben rápidamente flujo transmitido por la red
  - Son posiciones “estratégicas” en la red

Propiedades de los nodos. Medidas de centralidad

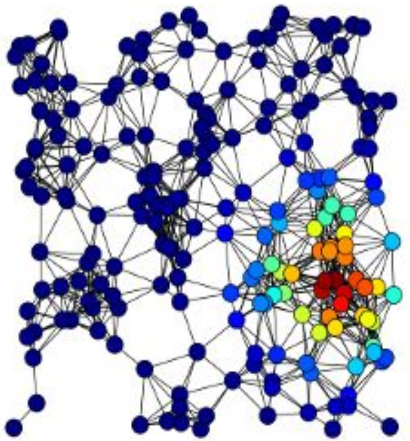
Centralidad de autovector:



¿Qué tan centrales son los primeros vecinos del nodo en cuestión?

Propiedades de los nodos. Medidas de centralidad

Centralidad de autovector:



La importancia del nodo depende de la importancia de sus vecinos (definición recursiva)

$$x_\nu = \frac{1}{\lambda} \sum_{t \in M(\nu)} x_t = \frac{1}{\lambda} \sum_{t \in G} a_{\nu,t} x_t$$

primeros vecinos

todo el grafo

Autovectores:
$A\,x = \lambda\,x$

Siendo $G=(V,E)$ un grafo no dirigido con ν,t: nodos

$M(\nu)$: Conjunto de vecinos del nodo ν

$A = a_{\nu,t}$: Matriz de adyacencia

λ : Máximo autovalor de los autovectores de A

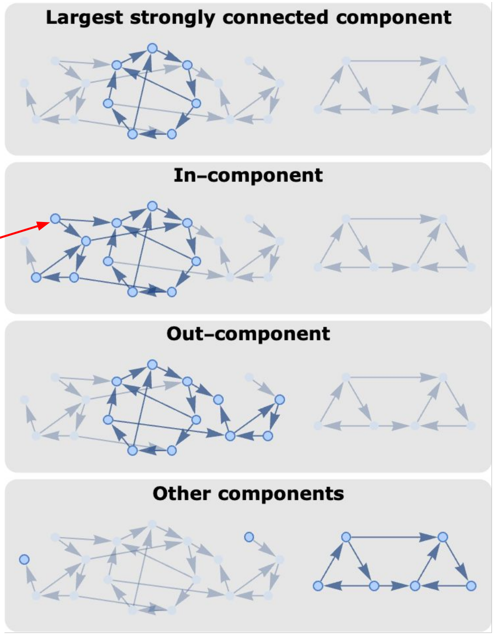
(sólo el autovector con máximo autovalor garantiza que $x_\nu >= 0 \;\; \forall \; \nu$)

Propiedades de los nodos. Medidas de centralidad

Centralidad de autovector:

Problema en redes dirigidas:

- La centralidad de autovector está relacionada a los enlaces entrantes de un nodo
- Solo nodos en la componente strongly-connected o en la Out-component tienen una centralidad de autovector no nula.
- Nodos de la In-component pueden tener centralidad de autovector nula (lo mismo ocurre para nodos que tienen enlaces entrantes provenientes de nodos con centralidad de autovector nula)



Hay un camino para llegar a cualquier nodo del componente

Solución de la **centralidad de Katz**: darle a todos los nodos "gratuitamente" una centralidad de autovector mínima no nula (que podrá "transferir" a otros nodos haciendo enlaces salientes hacia ellos).

Centralidad de Katz:

"Un nodo tendrá alta centralidad si sus vecinos tienen alta centralidad o si recibe muchos enlaces (y ningún nodo tendrá centralidad nula)"

Propiedades de los nodos. Medidas de centralidad

Correcciones a la centralidad de autovector para redes dirigidas:

Centralidad de Katz:

“Un nodo tendrá alta centralidad si sus vecinos son de alta centralidad o si recibe muchos enlaces”

Corolario: “ningún nodo tendrá centralidad nula”

$$x_i = \alpha \sum_j A_{ij} x_j + \beta$$

En la práctica: $\beta = 1$; $0 < \alpha < 1/\lambda$

Propiedades de los nodos. Medidas de centralidad

Correcciones a la centralidad de autovector para redes dirigidas:

Centralidad de Katz:

"Un nodo tendrá alta centralidad si sus vecinos son de alta centralidad o si recibe muchos enlaces"

$$x_i = \alpha \sum_j A_{ij} x_j + \beta$$

En la práctica: $\beta = 1$; $0 < \alpha < 1/\lambda$

Problema 2: un nodo de muy alta centralidad "contagia" un valor alto de centralidad a todos sus vecinos vía k^{out}

Propiedades de los nodos. Medidas de centralidad

Correcciones a la centralidad de autovector para redes dirigidas:

Centralidad de Katz:

“Un nodo tendrá alta centralidad si sus vecinos son de alta centralidad o si recibe muchos enlaces”

$$x_i = \alpha \sum_j A_{ij} x_j + \beta$$

En la práctica: β =1 ; 0 < α < 1/λ

Problema 2: un nodo de muy alta centralidad “contagia” un valor alto de centralidad a todos sus vecinos vía k^{out}

Centralidad de PageRank:

“Un nodo tendrá alta centralidad si ***sus vecinos de alta centralidad lo conectan específicamente*** *o si recibe muchos enlaces”*

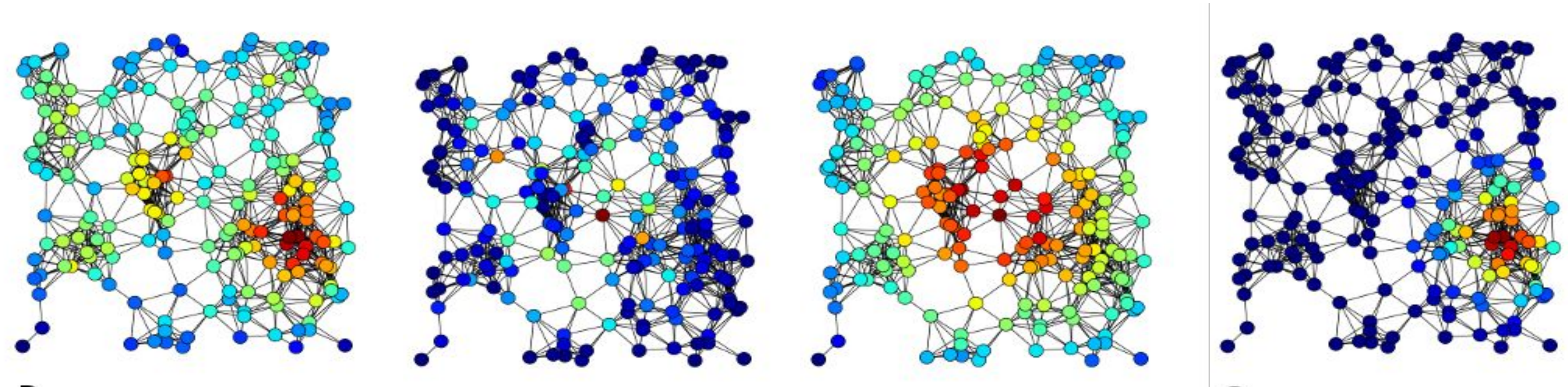
$$x_i = \alpha \sum_j \frac{1}{k_j{}^{out}} A_{ij} x_j + \beta$$

Normaliza por la “promiscuidad” del nodo que lo enlaza

En la práctica: α = 0.85

Propiedades de los nodos. Medidas de centralidad

Centralidad de grado | **Centralidad de intermediación** | **Centralidad de cercanía** | **Centralidad de autovector**

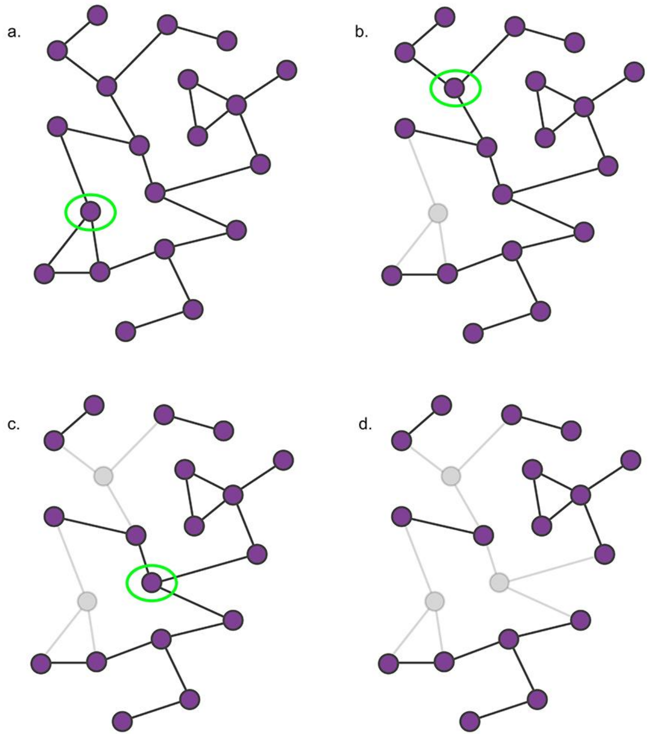


Para cada problema particular siempre hay que preguntarse: ¿Centralidad con respecto a qué?

Robustez

¿Cuántos nodos puedo borrar de la red sin generar impacto en la integridad de la red?

¿Cómo se ve afectada la **componente gigante**?

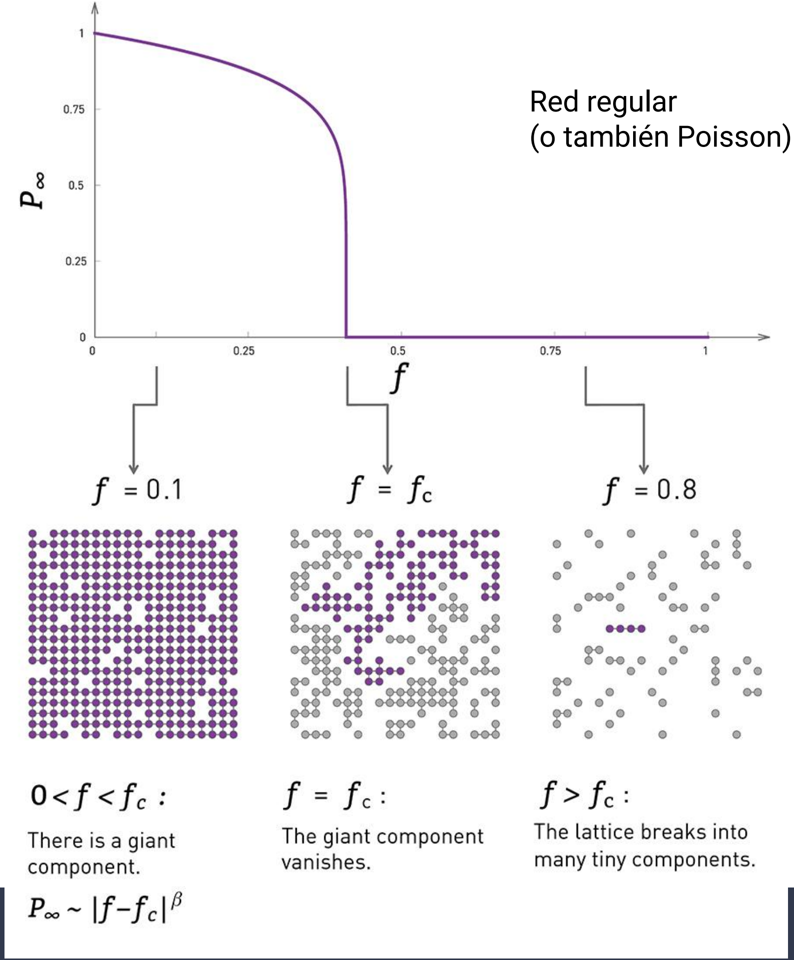


Robustez

Método de Fallas:
Borrar nodos aleatoriamente

f fracción de nodos a borrar

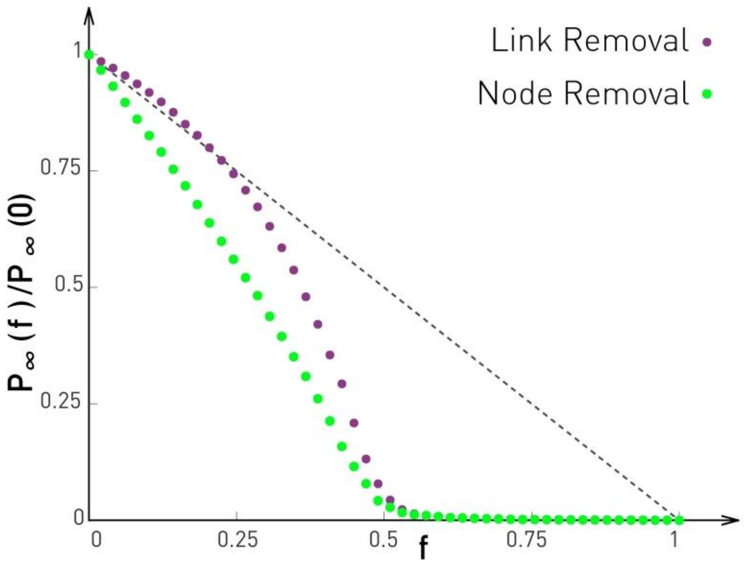
p_∞ componente gigante de la red



Robustez

Método de Fallas: borrando nodos al azar

En una red regular (o Poisson) el f_c de borrado de nodos o enlaces es similar



$$p_\infty(f) \text{ componente gigante de la red para la fracción } f$$

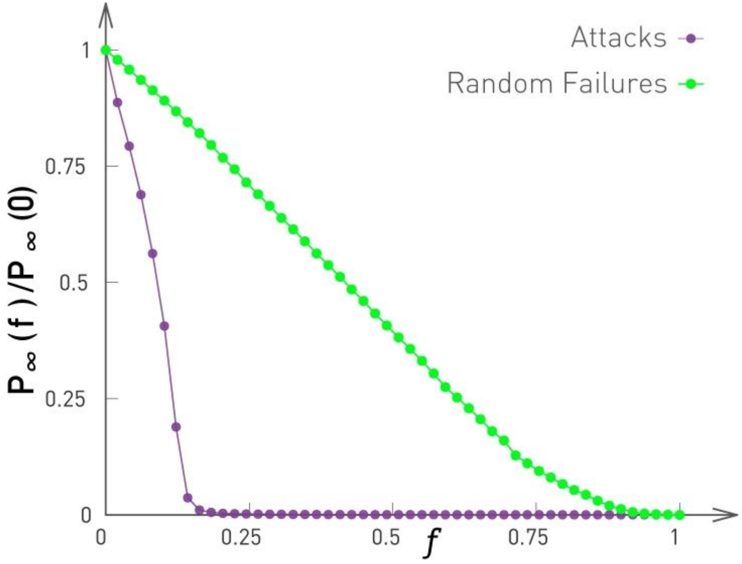
Robustez

Método de Ataques: dirigidos a nodos “importantes”

- Las redes scale-free son mucho más robustas a ataques al azar que las redes Poisson

- Sin embargo, son muy susceptibles a ataques dirigidos a sus nodos de mayor grado (sus “hubs”)

Red Scale Free



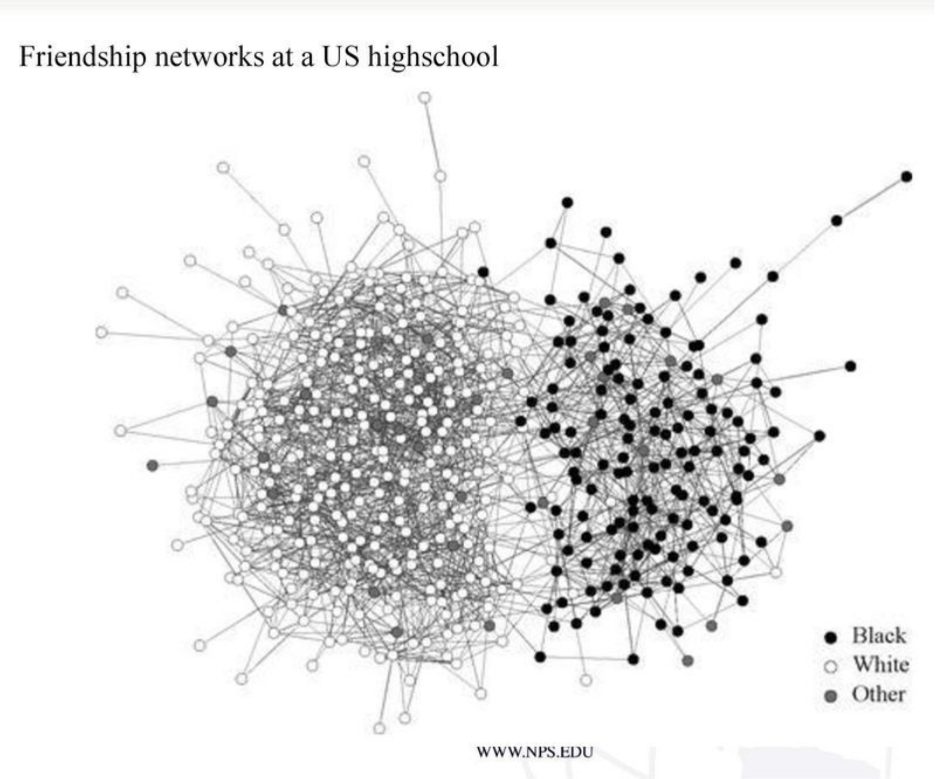
Mezclado selectivo (Assortative mixing)

Red no selectiva (disortativa)

Los nodos de **una clase** se conectan preferencialmente con nodos de **otra clase**.

Red selectiva (asortativa)

Los nodos de **una clase** se asocian preferentemente entre ellos.



Mezclado selectivo (Degree Assortative mixing)

Red no selectiva (disortativa)

Los nodos con **alto grado** se conectan preferencialmente con nodos de **bajo grado**.

En vez de clases podemos hablar de grado de nodos.

Red selectiva (asortativa)

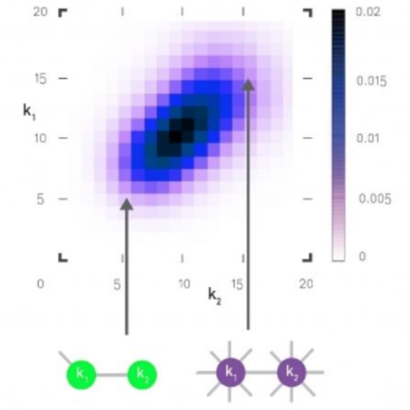
Los nodos de **alto grado** se asocian con otros nodos de alto grado, y los de **bajo grado** se asocian entre sí.

Las **redes sociales** tienden a ser selectivas y las **redes biológicas y físicas** no selectivas.

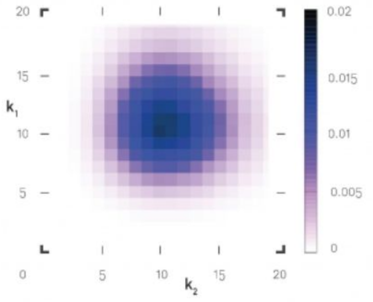
Mezclado selectivo (Degree Assortative mixing)

Correlación de grados

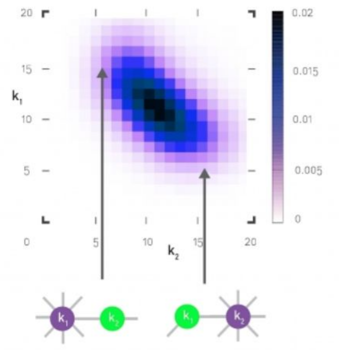
Suele usarse el coeficiente de correlación de Pearson (*r*) entre el grado de nodos adyacentes.



r>0 →Red selectiva



r=0 →Red neutra



r<0 →Red no selectiva

Mezclado selectivo (Assortative mixing)

Correlación de grados

Este coeficiente no es muy robusto cuando se lo aplica a redes muy heterogéneas

Asociación entre grados de vecinos

Evaluar la correlación de grados de un nodo y el grado medio de sus vecinos (***gmv* o k_{nn}**).

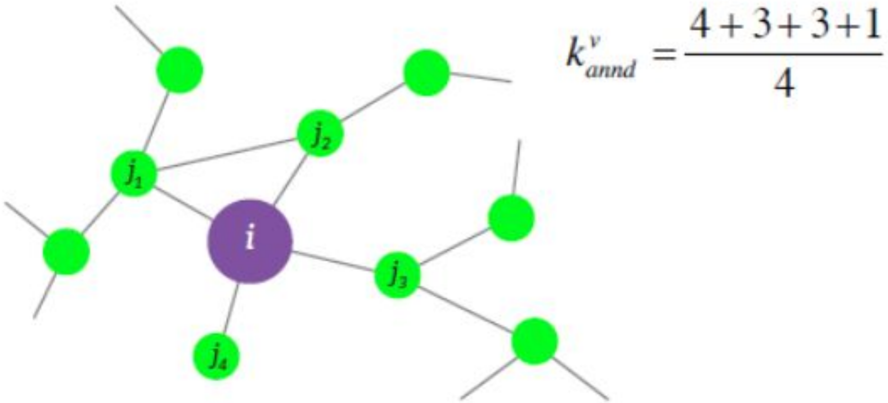
Mezclado selectivo (Assortative mixing)

Asociación entre grados de vecinos

Grado medio de los vecinos (***gmv(i)*** o $\boldsymbol{k_{nn}(i)}$).
Para cada nodo i:

$$gmv = \frac{1}{|N(i)|} \sum_{j \in N(i)} k_j$$

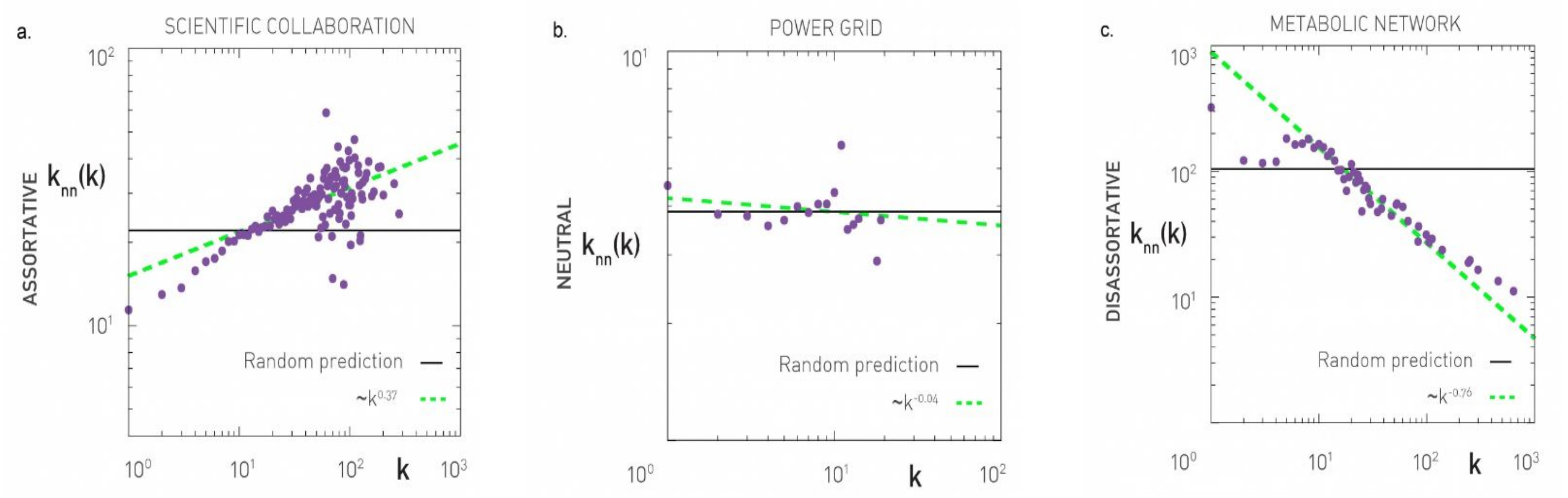
donde $N(i)$ son los vecinos de i, y k_j es el grado de j.



Después se puede graficar, por ejemplo, gmv vs k, o $<gmv>$ vs k.

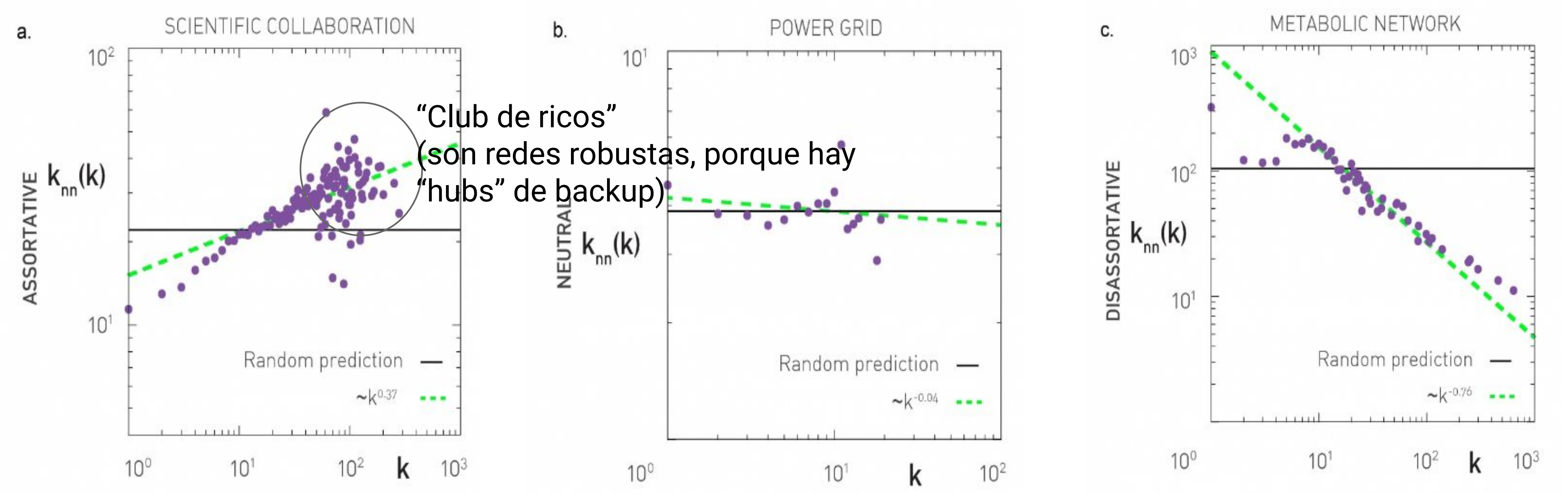
Mezclado selectivo (Assortative mixing)

Asociación entre grados de vecinos



Mezclado selectivo (Assortative mixing)

Asociación entre grados de vecinos

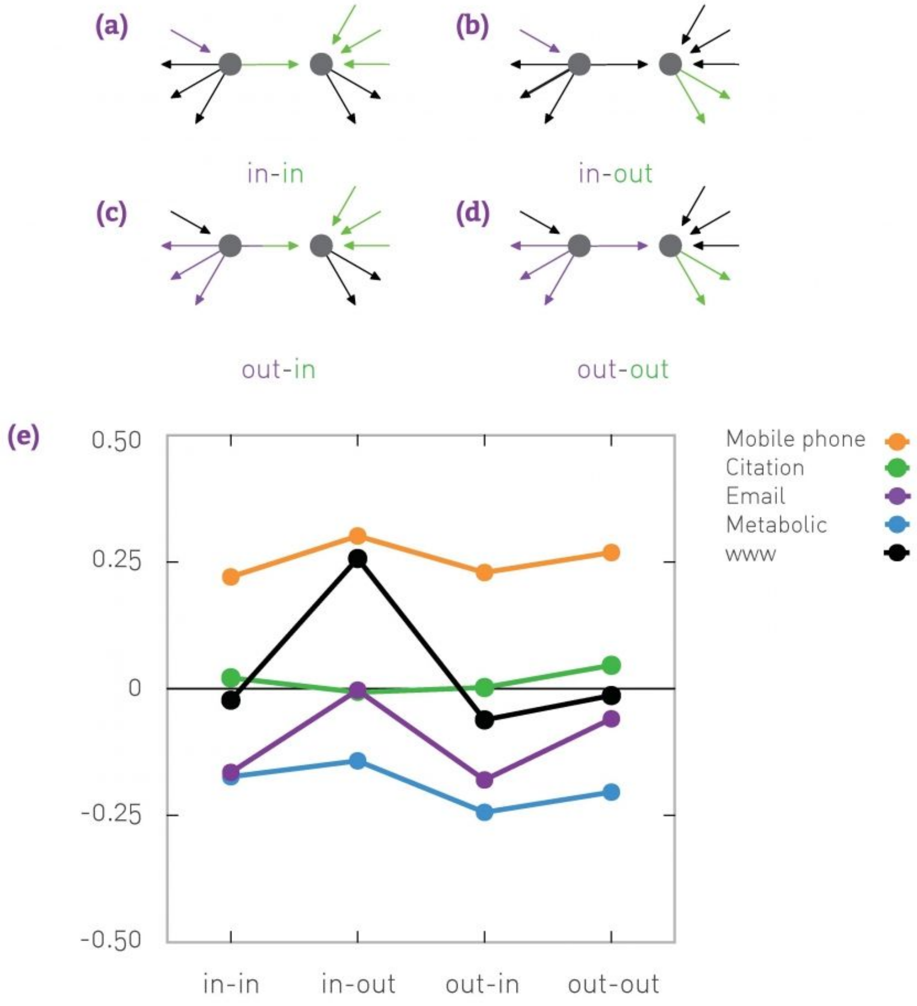


"Club de ricos"
(son redes robustas, porque hay
"hubs" de backup)

Mezclado selectivo

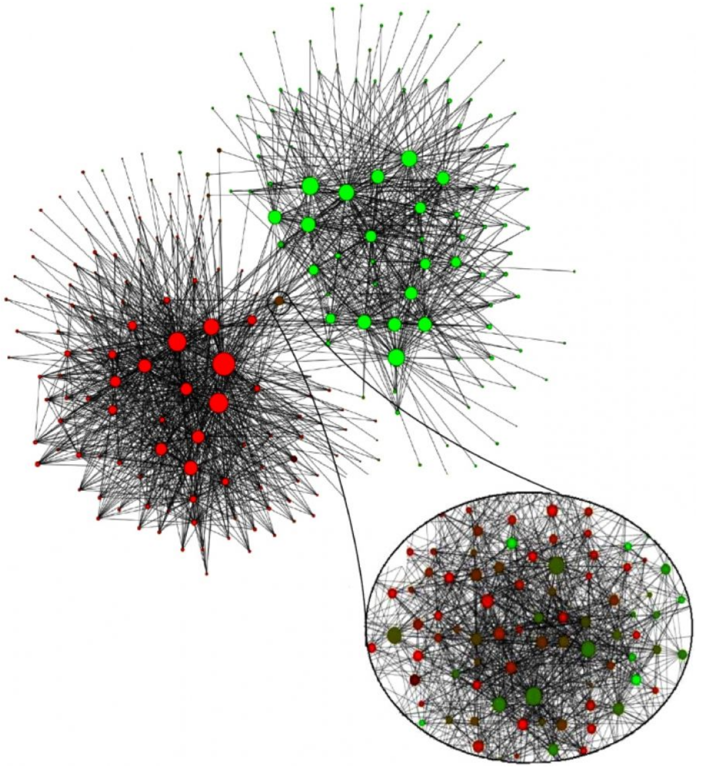
Asociación entre grados de vecinos

Para grafos dirigidos se calcula también la correlación entre los enlaces de entrada y de salida



Comunidades

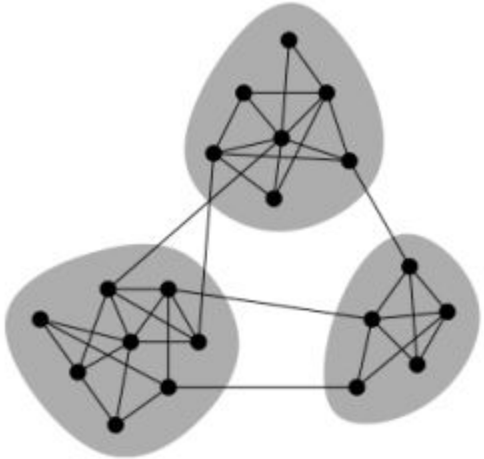
- Clustering jerárquico:
  - Ravasz
  - Girvan/Newman betweenness
- Modularidad
- Método de Louvain



Comunidades

¿Qué es una comunidad?

Cualitativamente: Un subconjunto de nodos que están más conectados entre sí que con el resto de los nodos.

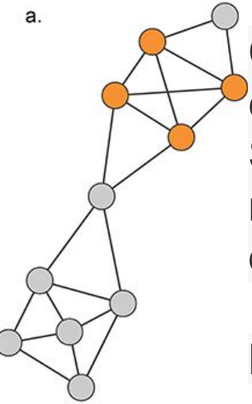


Barabasi cap 9

Comunidades

Definición de comunidades

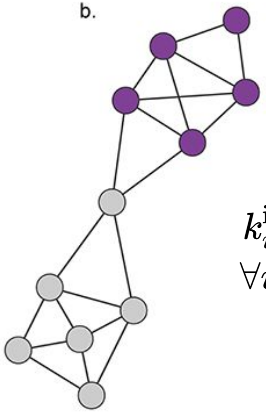
a.



Clique: Subgrafo completo. Subgrafo con la máxima densidad de enlaces.

Muy restrictivo

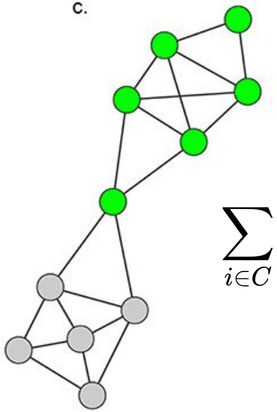
b.



Comunidad fuerte:

$$k_i^{\text{int}}(C) > k_i^{\text{ext}}(C) \quad \forall i \in C$$

c.



Comunidad débil

$$\sum_{i \in C} k_i^{\text{int}}(C) > \sum_{i \in C} k_i^{\text{ext}}(C)$$

(ojo, el verde es un ejemplo de comunidad, podría seguir agregando nodos)

Barabasi cap 9

Comunidades

Dado el grafo $G(n, m)$

La densidad de la red está dada por: $\rho = \dfrac{m}{n(n-1)/2}$

Tiene una comunidad formada por n_s nodos y m_s enlaces

Densidad interna

$$\rho_{\text{int}} = \frac{m_s}{n_s(n_s - 1)/2}$$

y densidad externa

$$\rho_{\text{ext}} = \frac{m_{ext}}{n_s(n - n_s)}$$

Comunidad debe cumplir que

$$\left\{ \begin{array}{r} \rho_{\text{int}} > \rho \\ \rho > \rho_{ext} \end{array} \right\}$$

Es una definición de comunidad débil, ya que estoy viendo densidades y no nodo a nodo

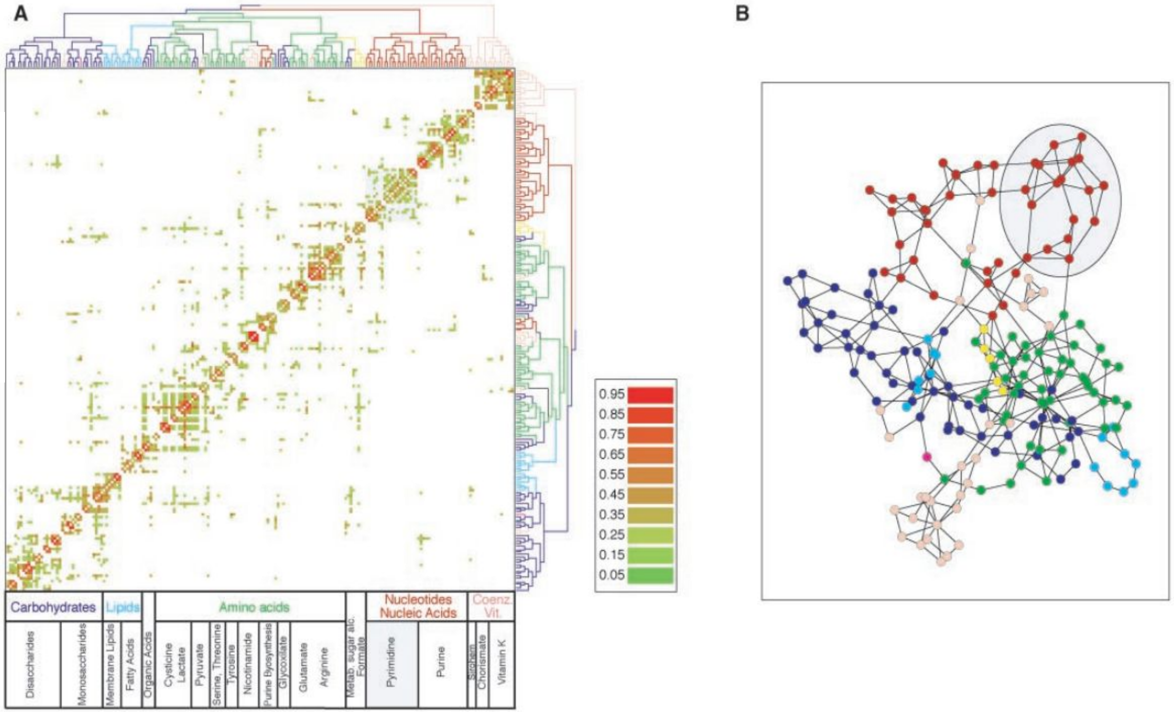
Detección de Comunidades

Agrupamiento jerárquico (Ravasz)

Hierarchical Organization of Modularity in Metabolic Networks

E. Ravasz,[1] A. L. Somera,[2] D. A. Mongru,[2] Z. N. Oltvai,[2*] A.-L. Barabási[1*]

Identificar la topología modular del metabolismo de la E. Coli



Science 2002

Detección de Comunidades

Agrupamiento jerárquico (Ravasz)

1. Se construye una matriz de similaridad de vértices a partir de la matriz de adyacencia

2. Estrategia aglomerativa: se adosan sucesivamente nodos y comunidades de alta similaridad

3. Agrupamiento jerárquico

4. El dendrograma describe el orden en el que nodos fueron agregados a comunidades. Es posible definir la partición en comunidades buscada a partir del dendrograma

Single linkage
Complete linkage
Average linkage

Similaridad en redes

Equivalencia estructural

Similaridad superposición topológica (topological overlap)

¿Cuán similares son los vecinos de un par de nodos?

Compara el número de vecinos en común con los esperados por azar.

- TO_{ij}=1 si i y j tienen un link entre ellos y los mismos vecinos.
- TO_{ij} =0 si i y j no tienen vecinos en común y tampoco link entre ellos.

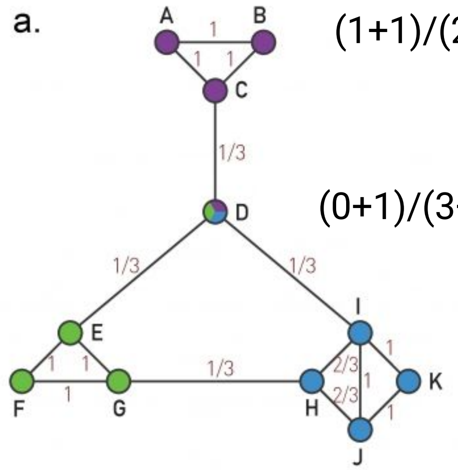
$$TO_{ij} = \frac{n_{ij} + \Theta(A_{ij})}{min(k_i, k_j) + 1 - \Theta(A_{ij})}$$

η_{ij} número de vecinos en común

$$\Theta(x) = \begin{cases} 0 \; si \; x \leq 0 \\ 1 \; si \; x > 0 \end{cases}$$

binarización

Barabasi cap 9



(1+1)/(2+1-1)=1

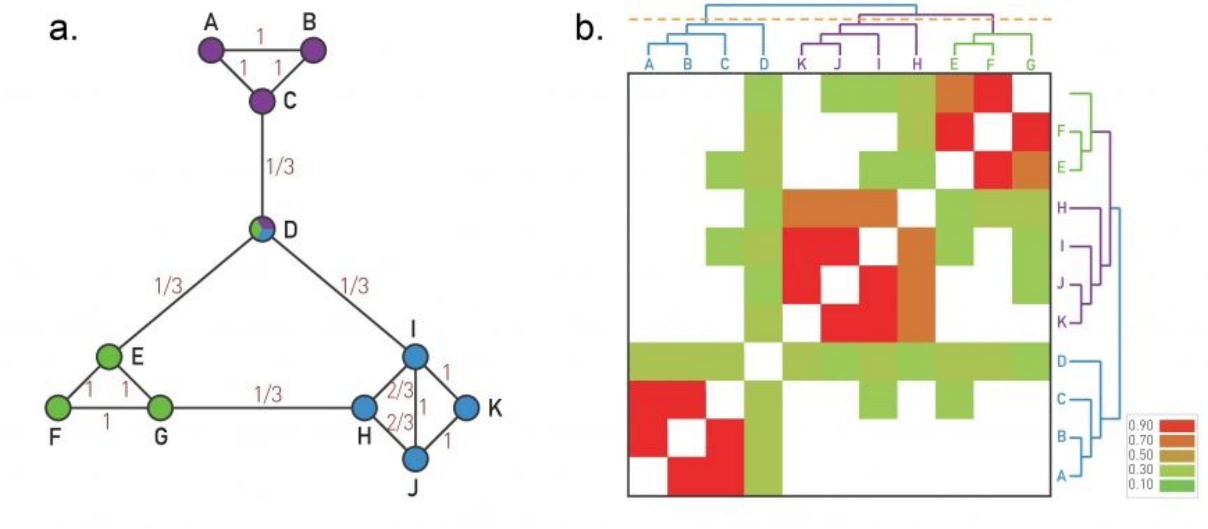
(0+1)/(3+1-1)=1/3

Detección de Comunidades

$$TO_{ij} = \frac{n_{ij} + \Theta(A_{ij})}{min(k_i, k_j) + 1 - \Theta(A_{ij})}$$

Agrupamiento jerárquico aglomerativo (Ravasz)

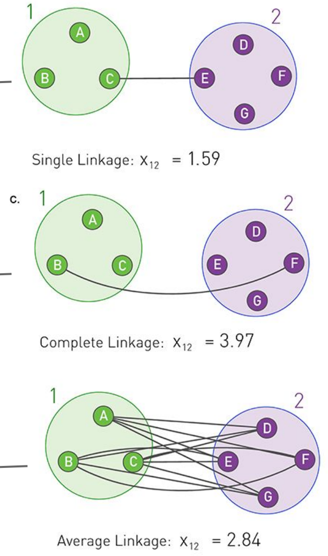
1. Define la matriz de similaridad (Topological overlap).
2. Decide similaridad de grupo



$\min(TO_{ij})$

$max(TO_{ij})$

$mean(TO_{ij})$

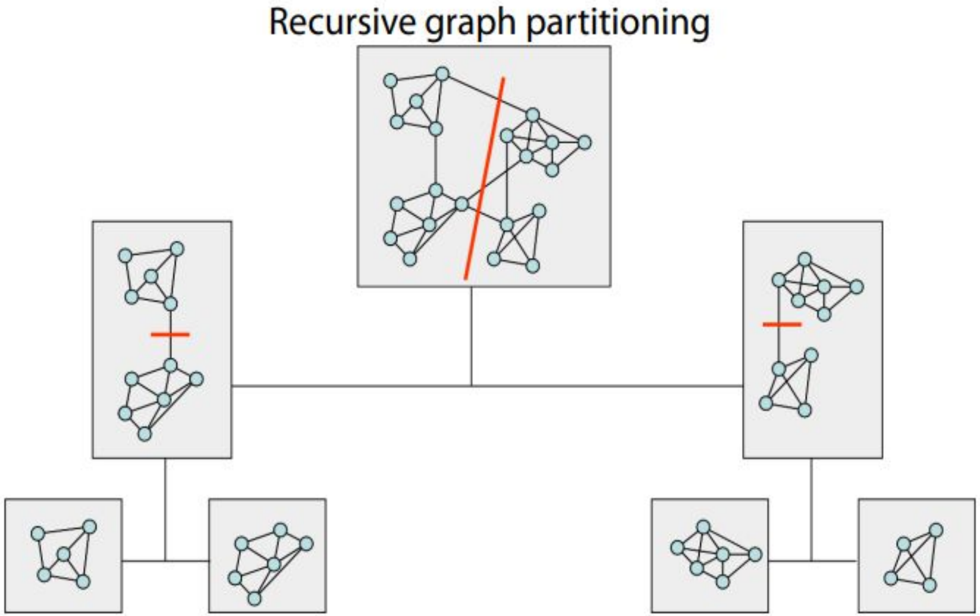


Science 2002

Detección de Comunidades

División jerárquica (Girvan-Newman)

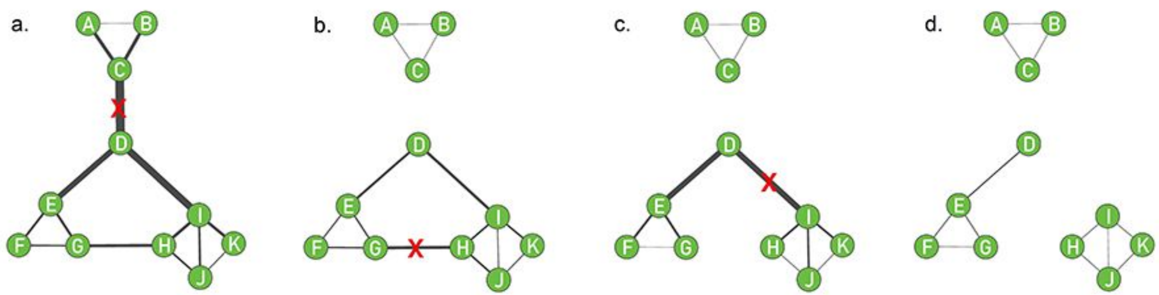
Idea:
“Desagregar” comunidades borrando iterativamente los enlaces que las unen

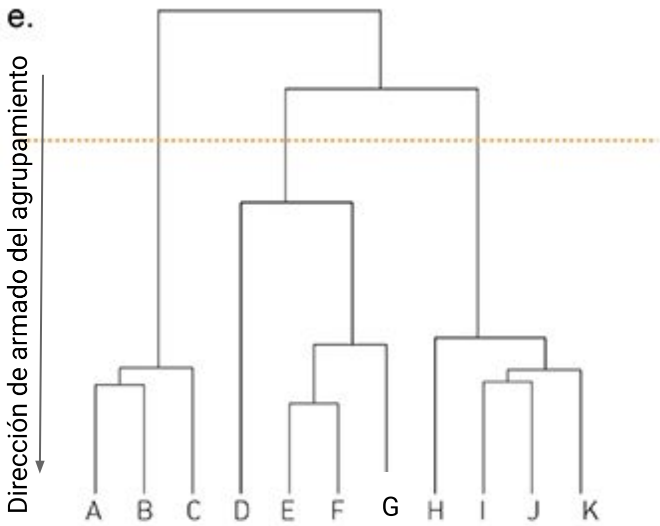


Barabasi cap 9

Detección de Comunidades

División jerárquica (Girvan-Newman)





Define medida de centralidad de enlaces:
Link Betweenness Centrality (proporción de caminos cortos entre todos nodos que atraviesen el enlace *(i,j)*)

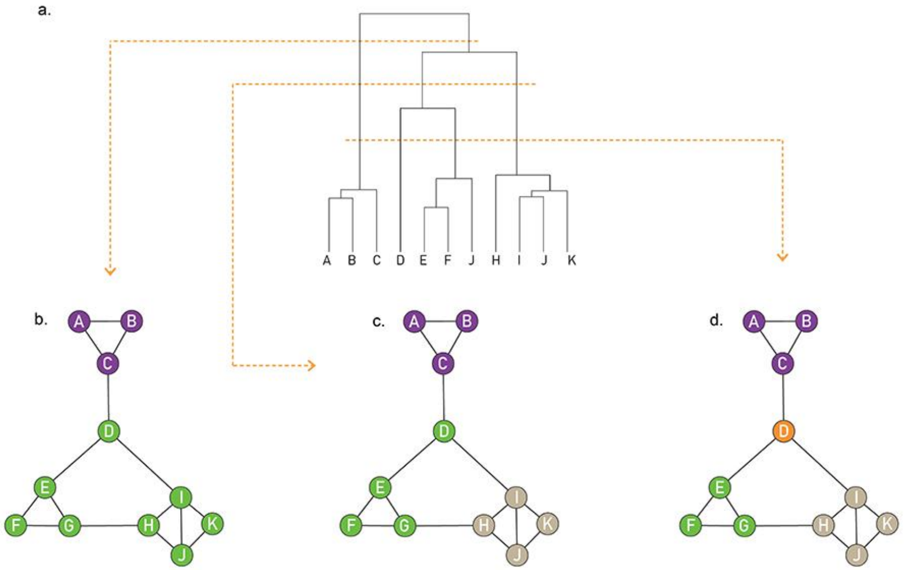
Va iterativamente borrando los enlaces de mayor Link Betweenness Centrality
(Obs: en cada iteración tiene que recalcular la centralidad de los enlaces… lento).

Barabasi cap 9

Detección de Comunidades

División jerárquica o (Girvan-Newman)

Agrupamiento jerárquico (Ravasz)



¿Dónde corto?

Solución:
Maximizar "modularidad"

Barabasi cap 9

Detección de Comunidades

Modularidad

Dada una partición en módulos, medir el exceso de enlaces intra-modulares con respecto a una red al azar (Poissoniana)

$$Q = \frac{1}{2m} \sum_{i,j}^{N} \left(A_{i,j} - \frac{k_i k_j}{2m} \right) \delta_{i,j}$$

Selecciona pares de nodos que pertenezcan a una misma comunidad

Prob. que par de nodos de grado (k_i, k_j) tengan un enlace en una red al azar

m: número de links

A_{ij}: matriz de adyacencia

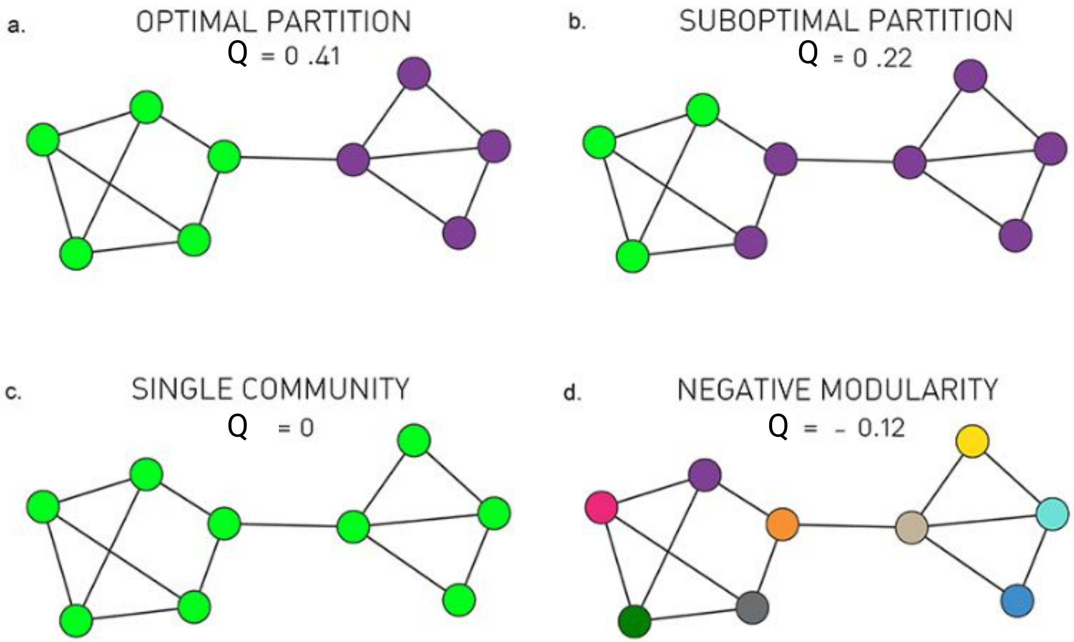
k: grado

$\delta_{i,j}$: 1 si i,j pertenecen al mismo comunidad, 0 si no.

Barabasi cap 9

Detección de Comunidades

Modularidad



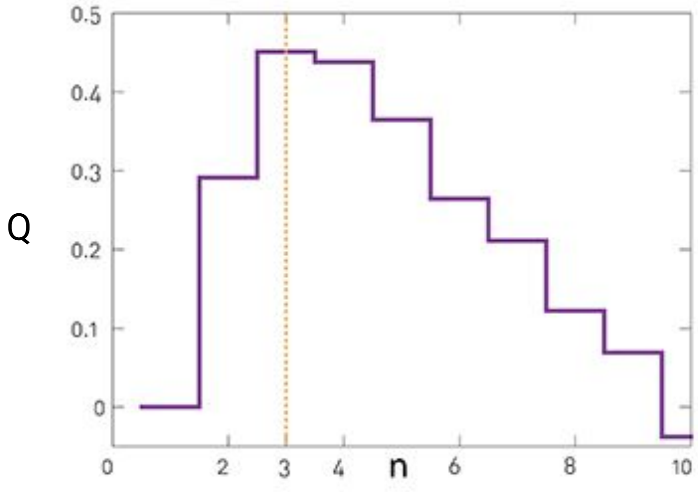
$$Q = \frac{1}{2m} \sum_{i,j}^{N} \left(A_{i,j} - \frac{k_i k_j}{2m} \right) \delta_{i,j}$$

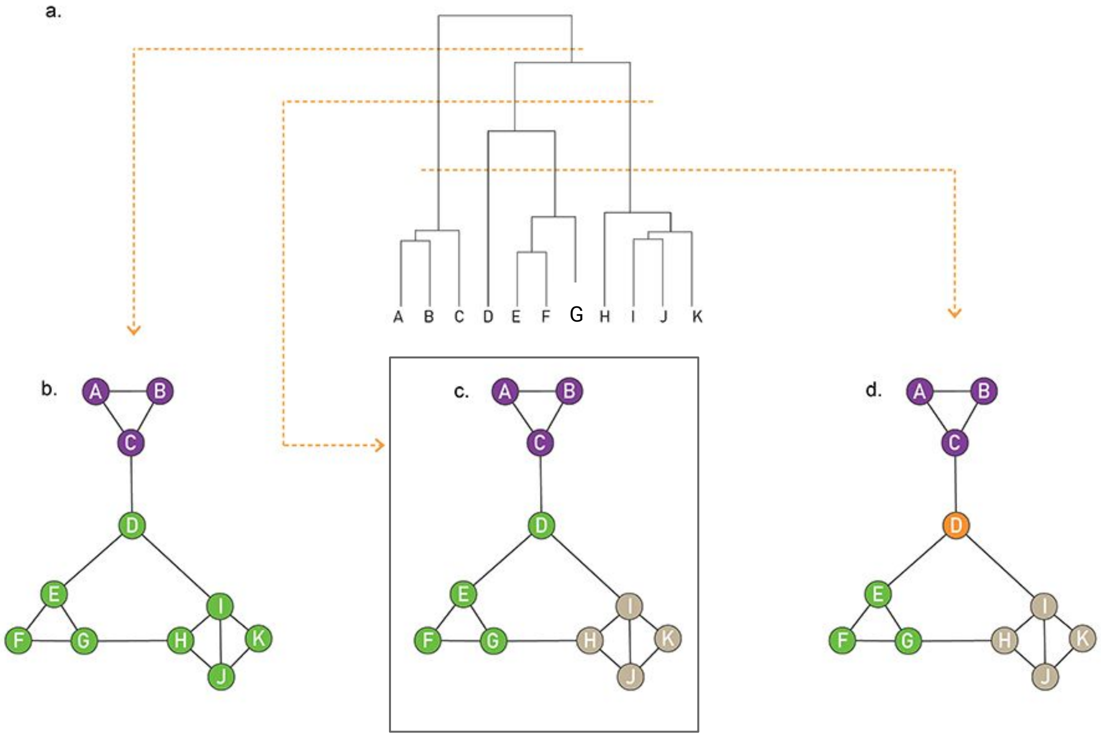
m: número de links
A_{ij}: matriz de adyacencia
k: grado
$\delta_{i,j}$: 1 si i,j pertenecen al mismo comunidad, 0 si no.

Q varía en el rango [-0.5, 1]

Detección de Comunidades

División jerárquica (Girvan-Newman)





Barabasi cap 9

Detección de Comunidades

Método de Louvain

Journal of Statistical Mechanics: Theory and Experiment
An IOP and SISSA journal

Fast unfolding of communities in large networks

Vincent D Blondel[1], Jean-Loup Guillaume[1,2], Renaud Lambiotte[1,3] and Etienne Lefebvre[1]

[1] Department of Mathematical Engineering, Université Catholique de Louvain, 4 avenue Georges Lemaitre, B-1348 Louvain-la-Neuve, Belgium
[2] LIP6, Université Pierre et Marie Curie, 4 place Jussieu, F-75005 Paris, France
[3] Institute for Mathematical Sciences, Imperial College London, 53 Prince's Gate, South Kensington Campus, London SW7 2PG, UK
E-mail: vincent.blondel@uclouvain.be, jean-loup.guillaume@lip6.fr, r.lambiotte@imperial.ac.uk and pixetus@hotmail.com

Received 18 April 2008
Accepted 3 September 2008
Published 9 October 2008

Método de optimización de la modularidad

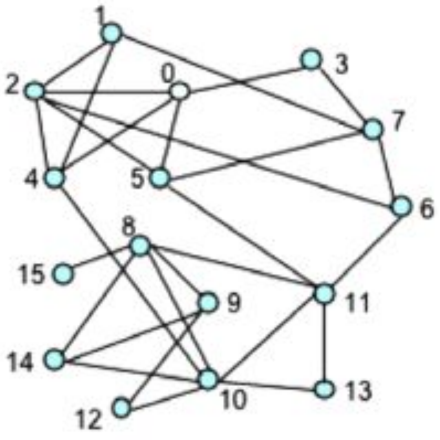
Encuentra particiones de alta modularidad

Estructura jerárquica

Es muy rápido

Detección de Comunidades

Método Louvain



Paso inicial

Optimización de la modularidad

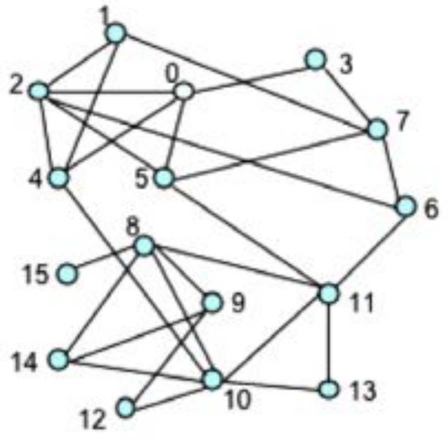
1.a Elegir un nodo al azar y calcular cambio en modularidad (ΔQ) que resultaría de juntarlo en una comunidad con cada uno de sus vecinos

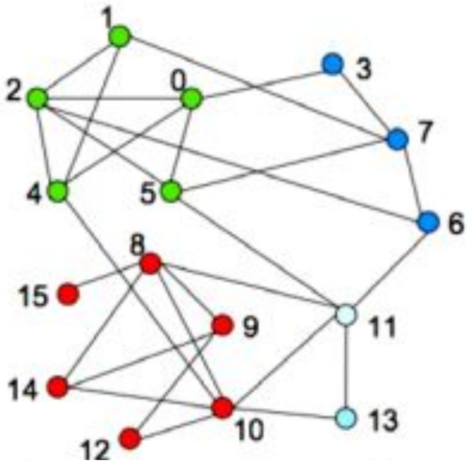
1.b El nodo se junta al nodo vecino con mayor ΔQ

1.c Se repite el proceso para todos los nodos, hasta que no haya mejoría

Detección de Comunidades

Método Louvain





Paso inicial

Optimización de la modularidad

1.a Elegir un nodo al azar y calcular cambio en modularidad (ΔQ) que resultaría de juntarlo en una comunidad con cada uno de sus vecinos

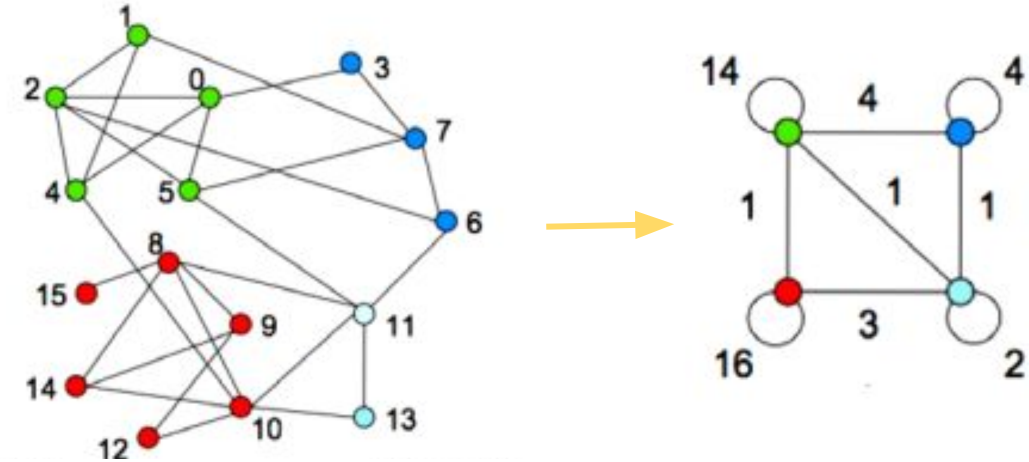
1.b El nodo se junta al nodo vecino con mayor ΔQ

1.c Se repite el proceso para todos los nodos, hasta que no haya mejoría

Solo nos estamos fijando pares de nodos. ¡Es rápido!

Detección de Comunidades

Método Louvain



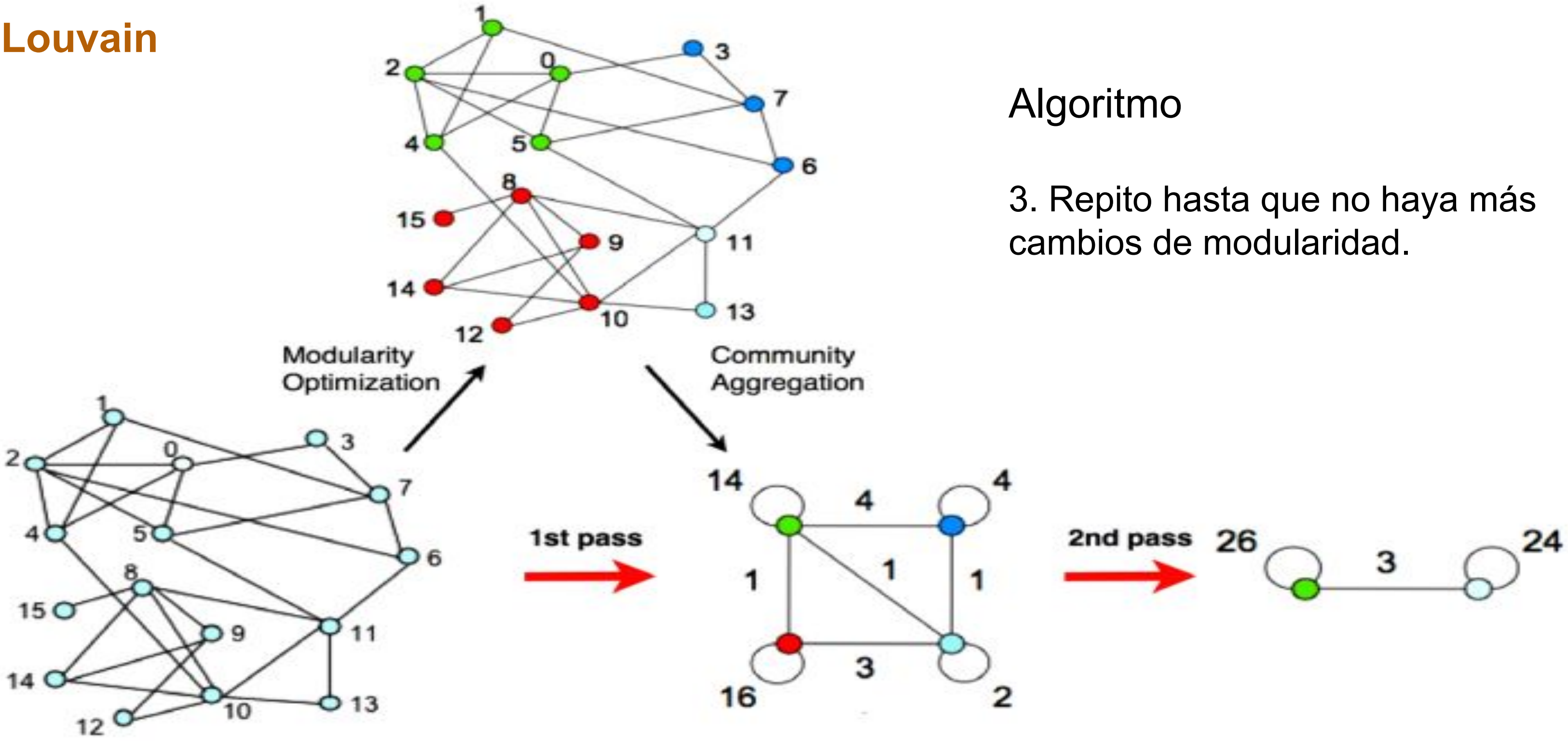
Segundo paso
Agregado de comunidades

2.a. Los nodos de una misma comunidad se unen en un “super” nodo que representa a la comunidad

2.b. Peso los enlaces sumando cantidad de enlaces intra- e inter-módulos

Detección de Comunidades

Método Louvain



Algoritmo

3. Repito hasta que no haya más cambios de modularidad.

Detección de Comunidades

Name	Nature	Comp.
Ravasz	Hierarchical Agglomerative	$O(N^2)$
Girvan-Newman	Hierarchical Divisive	$O(N^2)$
Greedy Modularity	Modularity Optimization	$O(N^2)$
Greedy Modularity (Optimized)	Modularity Optimization	$O(N\log^2 N)$
Louvain	Modularity Optimization	$O(L)$
Infomap	Flow Optimization	$O(N\log N)$
Clique Percolation (CFinder)	Overlapping Communities	$Exp(N)$
Link Clustering	Hierarchical Agglomerative; Overlapping Communities	$O(N^2)$

Barabasi cap 9