Teoría de redes y análisis de redes sociales

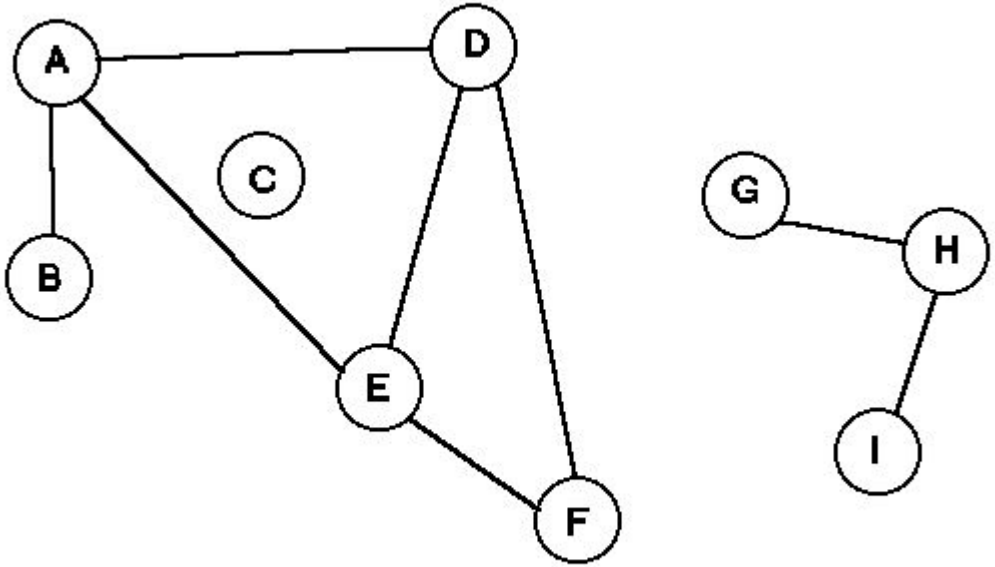
Una introducción rápida y limpia.

Sesión 2

Análisis y modelización de redes

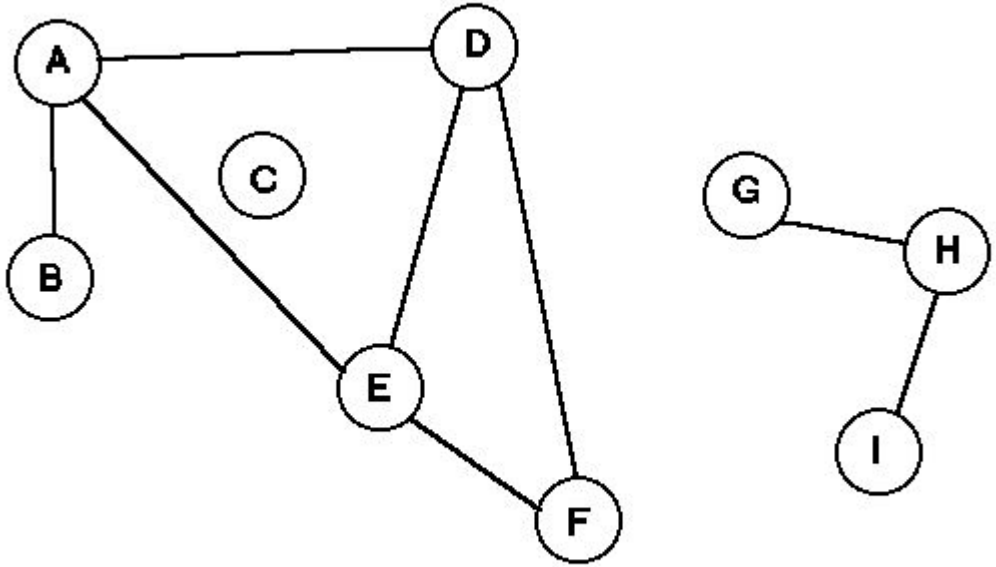
- ¿Qué quiere decir caracterizar una red?
- Lo que puede un modelo
- Centralidad
- Roles
- Comunidades

Representar una red: lista de enlaces



A	B
A	D
A	E
D	E
D	F
G	H
H	I

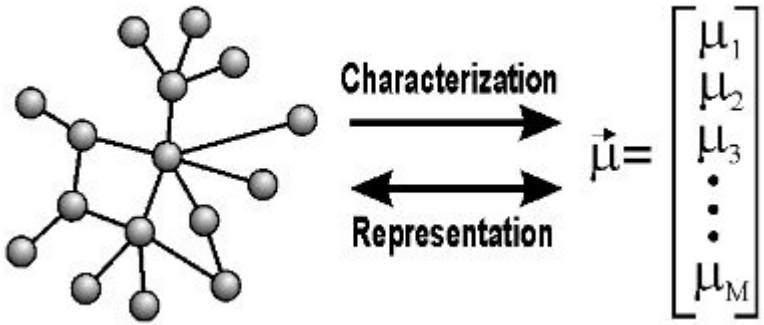
Representar una red: matriz de adjacencia

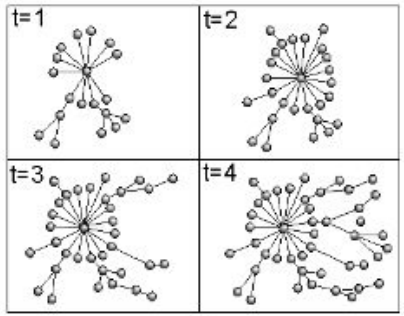


	A	B	C	D	E	F	G	H	I
A	0	1	0	1	1	0	0	0	0
B	1	0	0	0	0	0	0	0	0
C	0	0	0	0	0	0	0	0	0
D	1	0	0	0	1	1	0	0	0
E	1	0	0	1	0	1	0	0	0
F	0	0	0	1	1	0	0	0	0
G	0	0	0	0	0	0	0	1	0
H	0	0	0	0	0	0	1	0	1
I	0	0	0	0	0	0	0	1	0

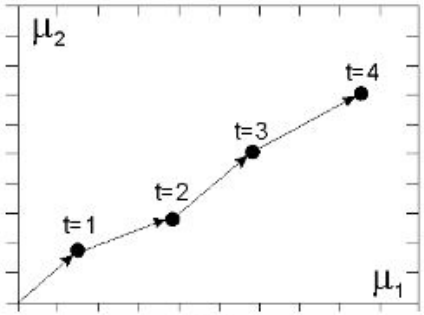
Caracterización y modelado

Caracterizar una red

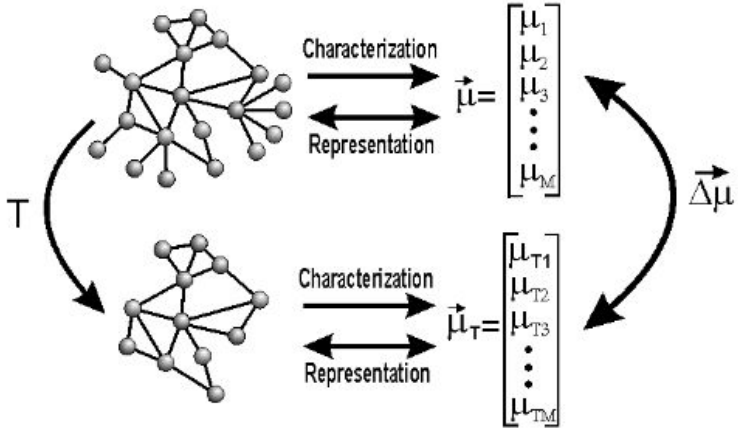




(a)



(b)



Modelar una red

Una red aleatoria es un conjunto estadístico (restringido):

cada miembro = configuración de nodos y enlaces

realizado con la misma probabilidad prescrita(= ponderaciones estadísticas)

Modelo de configuración

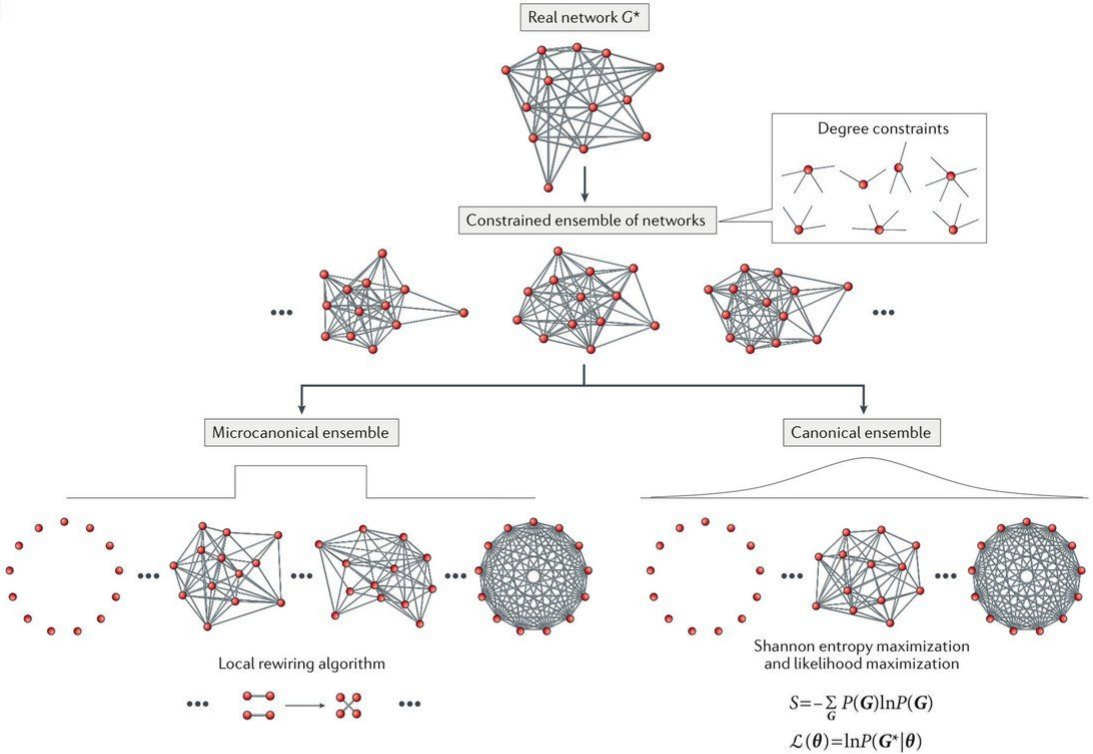


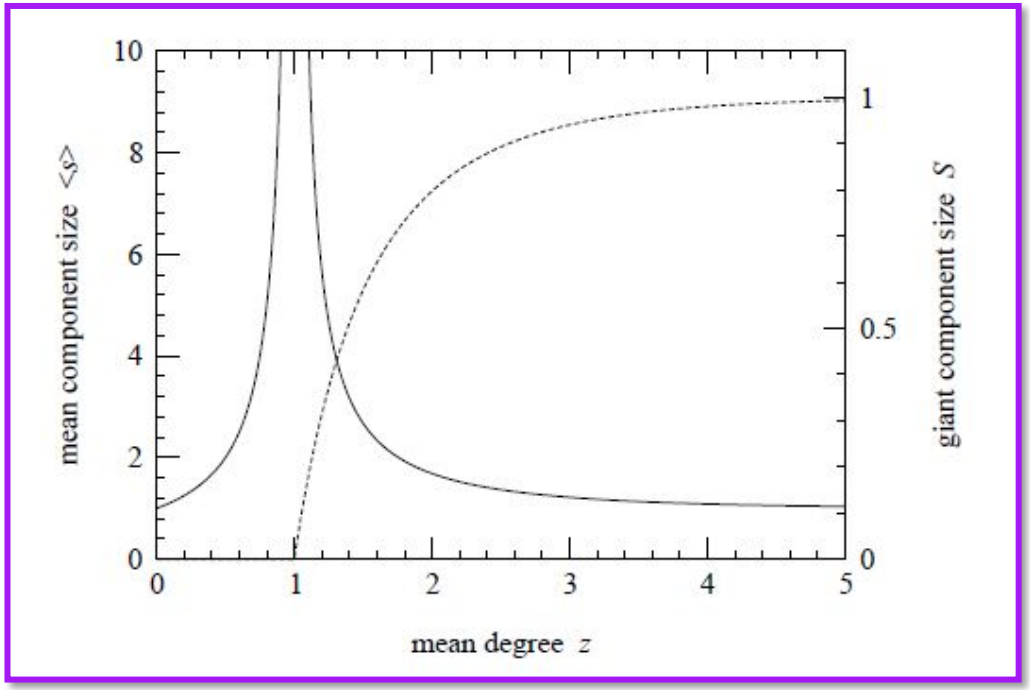
Fig. 1 | **Construction of the microcanonical and canonical ensemble of networks from local constraints.** The real network G^* is the source of constraints on the ensemble. In this case, the constraints are the degrees k^* of G^*. The microcanonical approach relies on the link rewiring method to numerically generate several network configurations, each with exactly the same degree sequence of G^*. The probability $P(G)$ of a network G in the ensemble is non-zero only for the subset of graphs that realize the enforced constraints exactly, as indicated by the schematic probability distribution.

Modelo Erdos-Renyi

Conjunto (canónico) de todos los grafos con n nodos y el mismo grado medio.

Reproduce el efecto de mundo pequeño.

No reproduce nada más, ej.
Transitividad
Distribución de grado

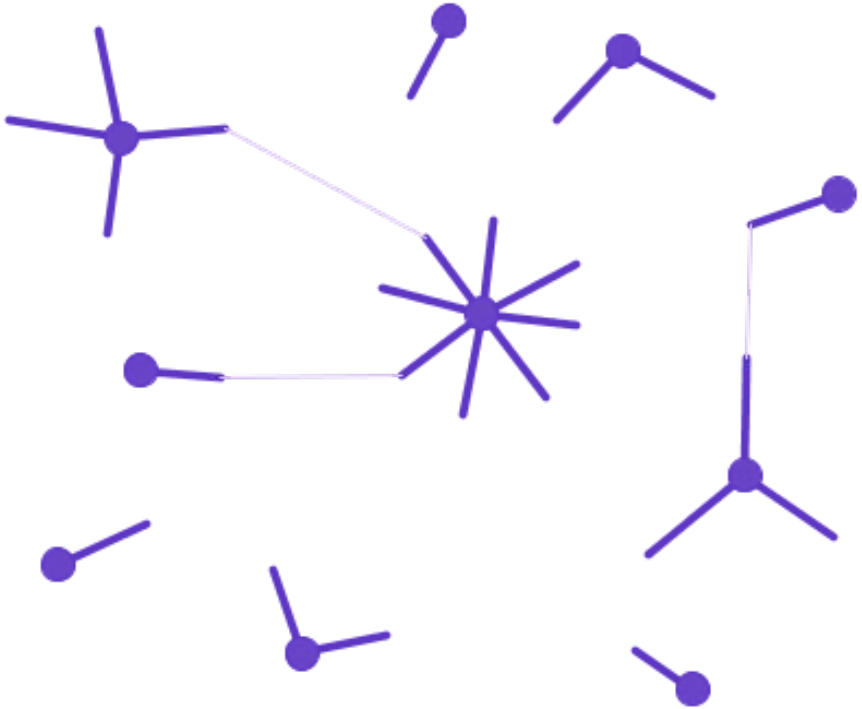


Modelo de configuración

Conjunto (canónico/microcanónico) de todos los grafos con n nodos y sus grados.

Reproduce la distribución de grado.

No reproduce correlaciones:
Modelo Nulo



Grafos aleatorios exponenciales

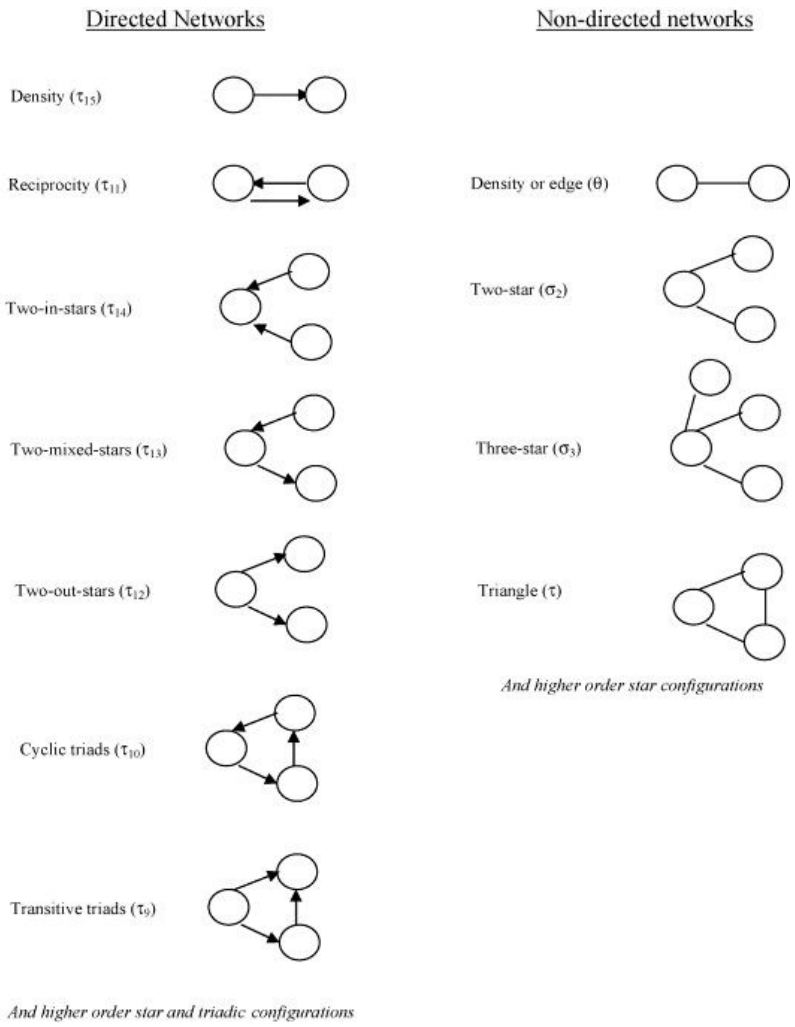
Conjunto (canónico/microcanónico) de todos los grafos con n y:

- número de enlaces
- número de enlaces de un grado dado
- número de triángulos

  ....

Estimación de parámetros.

Correlaciones.

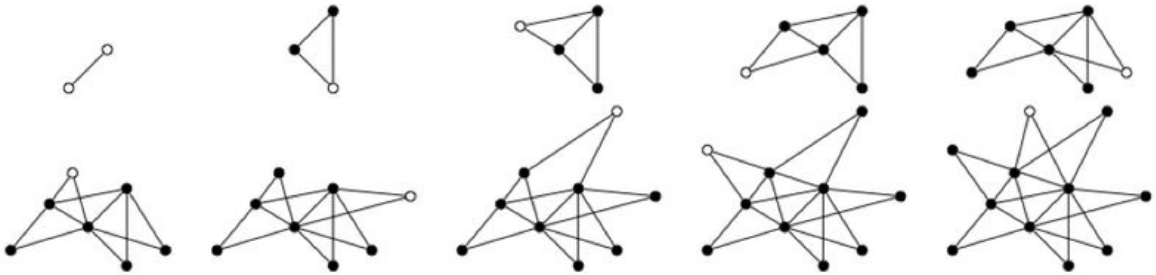


Modelos de crecimientos (o generativos)

Modelos discutidos hasta ahora: toman las propiedades observadas e intentan crear redes que incorporen esas propiedades

No ayuda a comprender cómo las redes llegan a tener esas propiedades

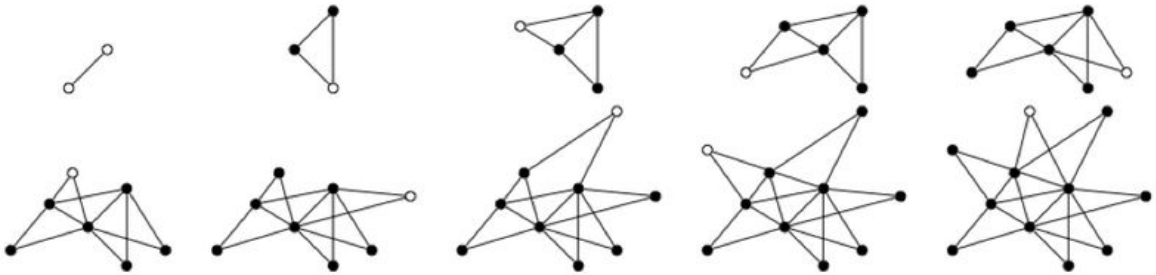
Modelos cuyo objetivo es explicar las propiedades de la red



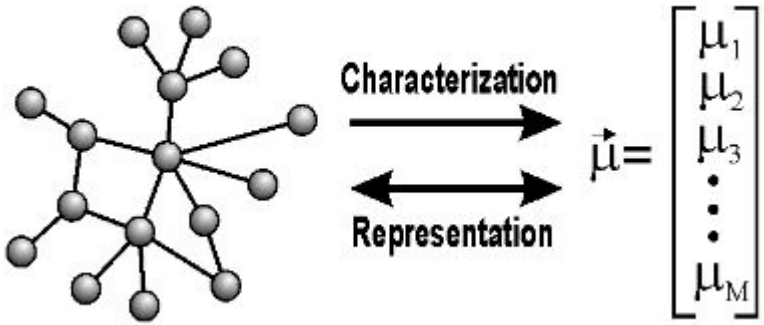
Apego preferencial

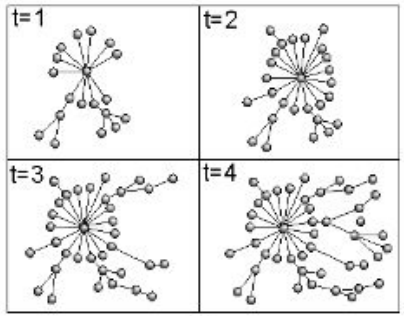
Clase de modelos enfocados a explicar el origen de las distribuciones de grado altamente asimétricas.

Los ricos se hacen más ricos: la cantidad que obtienes aumenta cor la cantidad que ya tienes

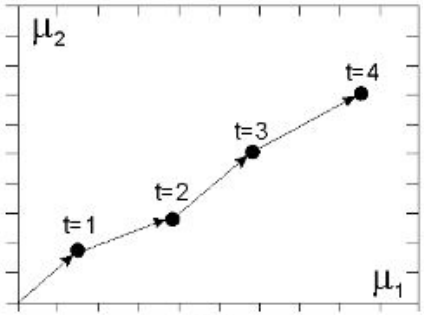


Caracterizar una red

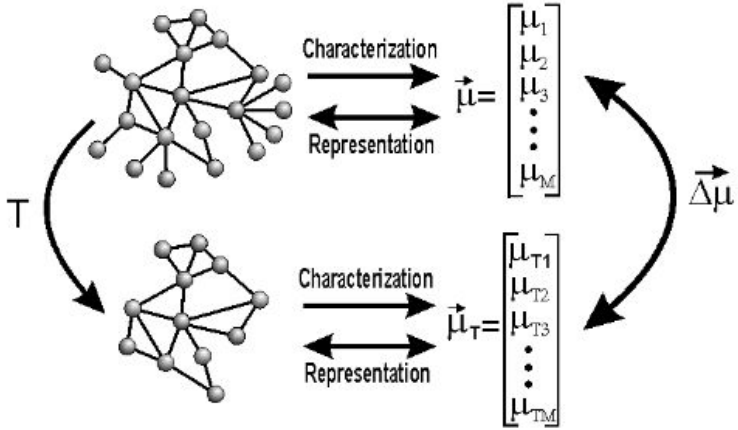




(a)



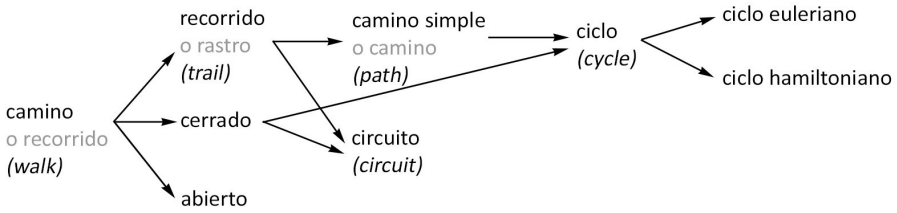
(b)



Medidas de centralidad

Lexico

- Un **recorrido** (en inglés, *trail*, a veces traducido como **rastro**)[1] es un camino sin aristas repetidas.
- Un **camino cerrado** es un camino cuyo vértice inicial y final coinciden.
- Un **camino abierto** es un camino cuyo vértice inicial y final no coinciden.
- Un **camino simple** (en inglés, *path*, a veces traducido como **camino**)[1] es un camino sin vértices repetidos, salvo quizás el primero y el último (por lo tanto, es un tipo especial de recorrido, pues tampoco tiene aristas repetidas).
- Un **circuito** (en inglés, *circuit*) es un recorrido que además es un camino cerrado.
- Un **ciclo** (en inglés, *cycle*) es un camino simple que además es un camino cerrado.

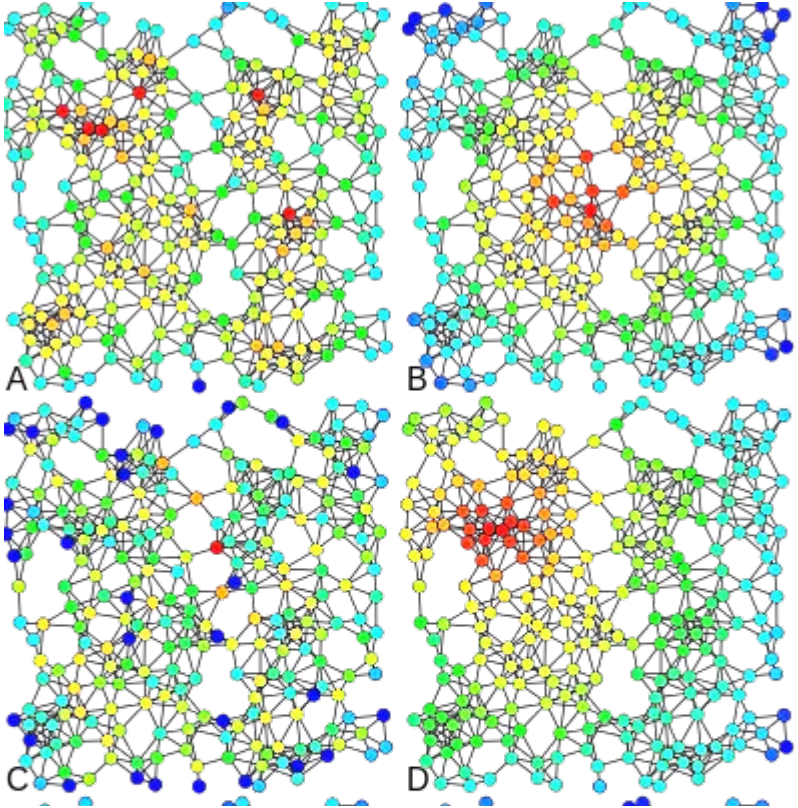


Medidas de centralidad

¿Qué mide una medida de centralidad?

Borgatti & Everett, 2006

Borgatti, 2005



Medidas de centralidad

- Prominencia del nodo
- Importancia estructural
- Poder
- Autonomia
- Control
- Riesgo
- Influencia
- ...

- Centralidad y flujo de red: hacer explícito lo implícito. Los tipos de trayectorias que puede seguir el tráfico (geodésicas, paths, trails o walks) y el método de propagación (transmisión, replicación en serie o transferencia).
- Centralidad y estructura de red: tipo de implicación nodal (radial vs medial) y la propiedad de los caminos evaluada (volumen frente a la longitud)

Medidas de centralidad (flujo): tipo de trayectoria

Las fórmulas para estas diferentes medidas hacen suposiciones **implícitas** sobre la forma en que “las cosas” fluyen en una red.

- Closeness (cercanía) y betweenness (intermediación): sólo tienen en cuenta las geodésicas (caminos más cortos).

- Eigenvector y PageRank: asumen que un nodo puede ser visitado múltiples veces.

Medidas de centralidad (flujo): tipo de propagación

Las fórmulas para estas diferentes medidas hacen suposiciones **implícitas** sobre la forma en que “las cosas” fluyen en una red.

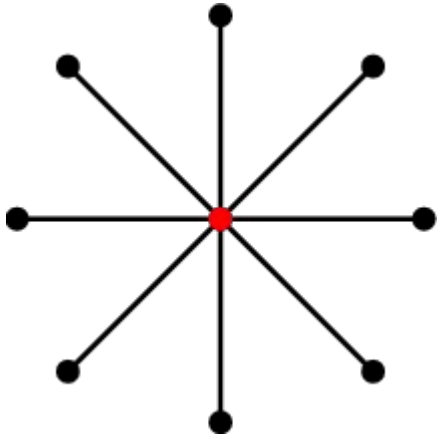
- Betweenness: asume que lo que fluya de nodo a nodo es indivisible (como un paquete) y solo puede seguir un camino.

- Eigenvector: asume que lo que fluye de nodo a nodo se puede multiplicar (como la información o un virus) y puede seguir múltiples caminos.

Medidas de centralidad (estructura)

Perspectiva estructural: cualquier medida de centralidad debería alcanzar sus valores máximos para el centro de una red en forma de estrella.

Lo que es definitorio es cómo se calcula.



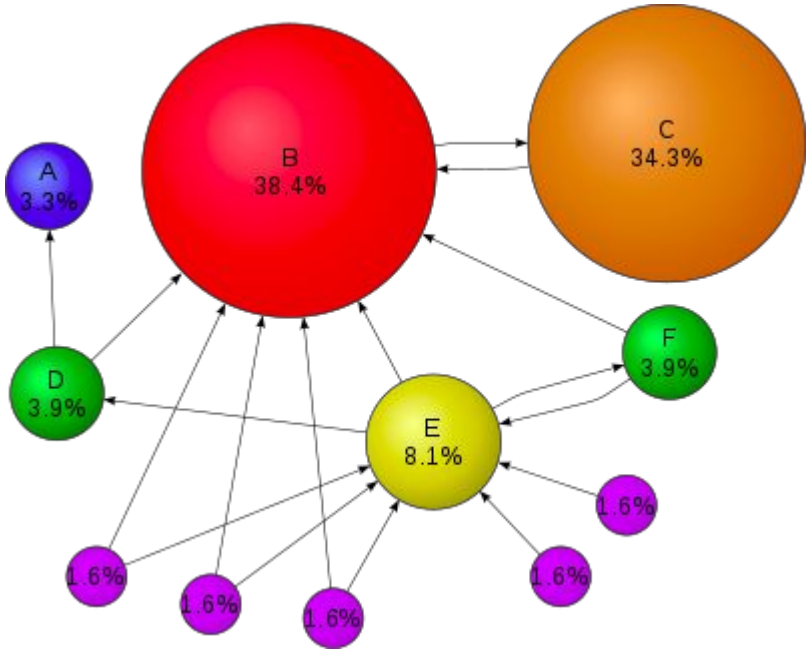
Medidas de centralidad (estructura) de tipo grado

Grado: conta el número de enlaces

K-path (variaciones): conta el número de paths

Centralidad de Katz: conta el número de walks (ponderados)

Centralidad de autovector/PageRank: la centralidad de un nodo debe ser una función de la centralidad de las personas con las que está asociada.



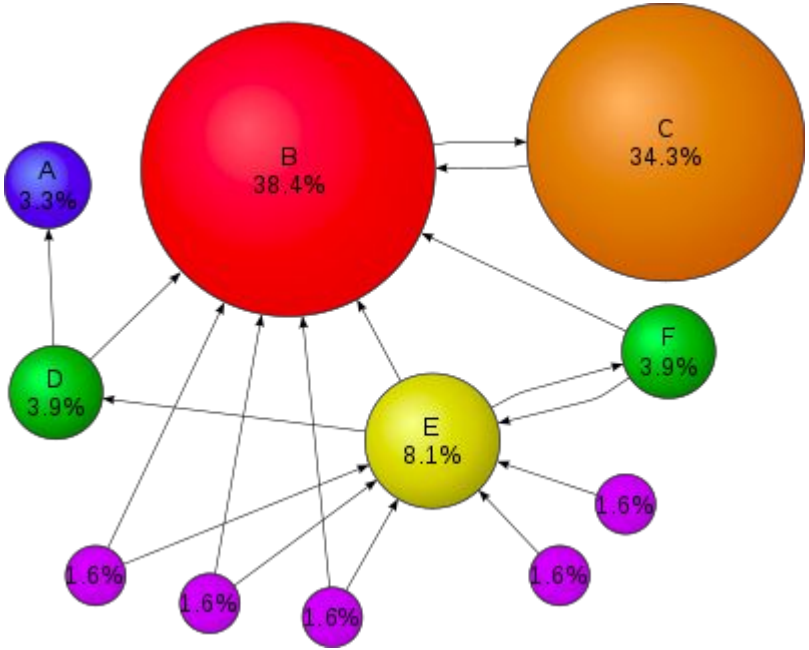
Medidas de centralidad (estructura) de tipo grado

Grado: conta el número de enlaces

K-path (variaciones): conta el número de paths

Centralidad de Katz: conta el número de walks (ponderados)

Centralidad de autovector:



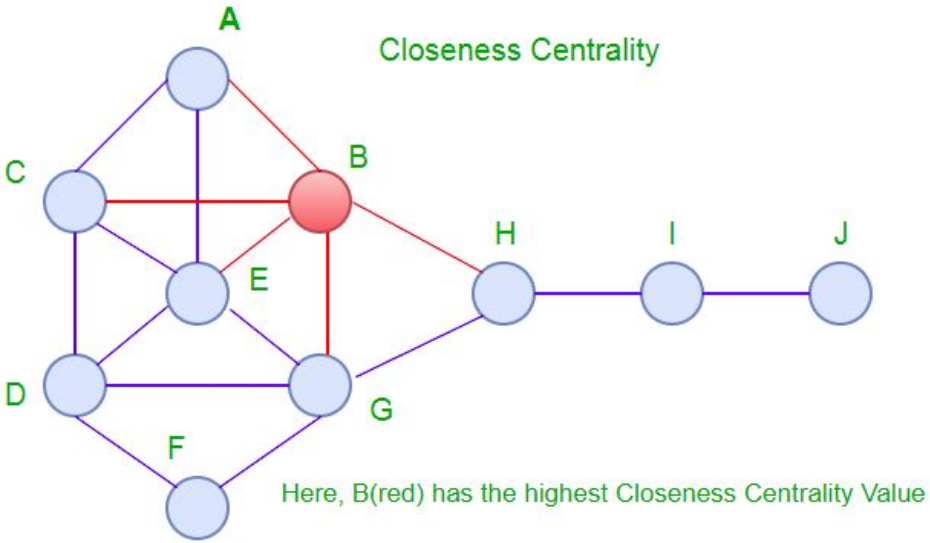
Medidas de volumen

Medidas de centralidad (estructura) de tipo cercanía

Closeness: la distancia geodésica total desde un nodo dado a todos los demás nodos (distancia promedio del nodo de cualquier otro nodo, cuanto más grande menos central).

Friedkin: tiene en cuenta la longitud promedic de todos los caminos

Centralidad de centroide: identifico uno o más nodos centrales y tengo en cuenta las distancia desde este/estos



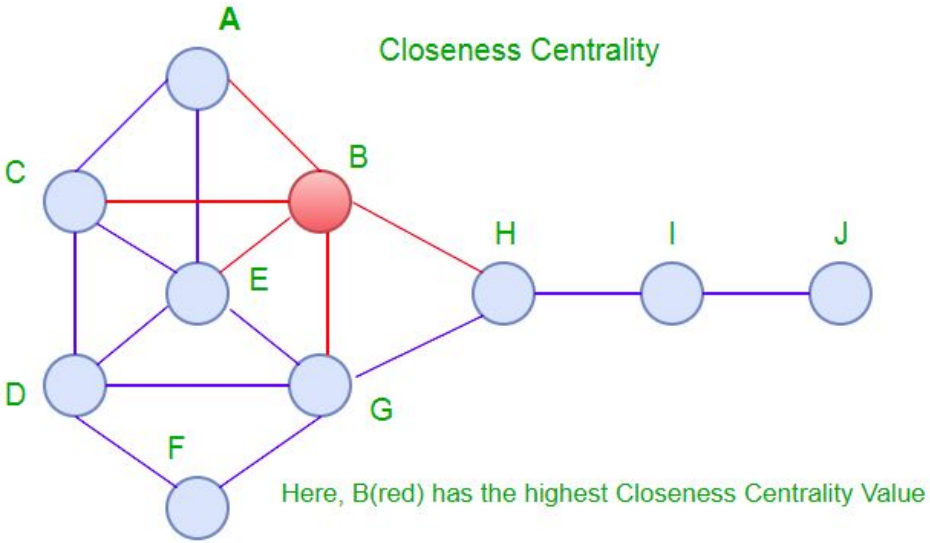
Medidas de centralidad (estructura) de tipo cercanía

Closeness: la distancia geodésica total desde un nodo dado a todos los demás nodos (distancia promedio del nodo de cualquier otro nodo, cuanto más grande menos central).

Friedkin: tiene en cuenta la longitud promedic de todos los caminos

Centralidad de centroide: identifico úno o más nodos centrales y tengo en cuenta las distancia desde este/estos

Medidas de longitud



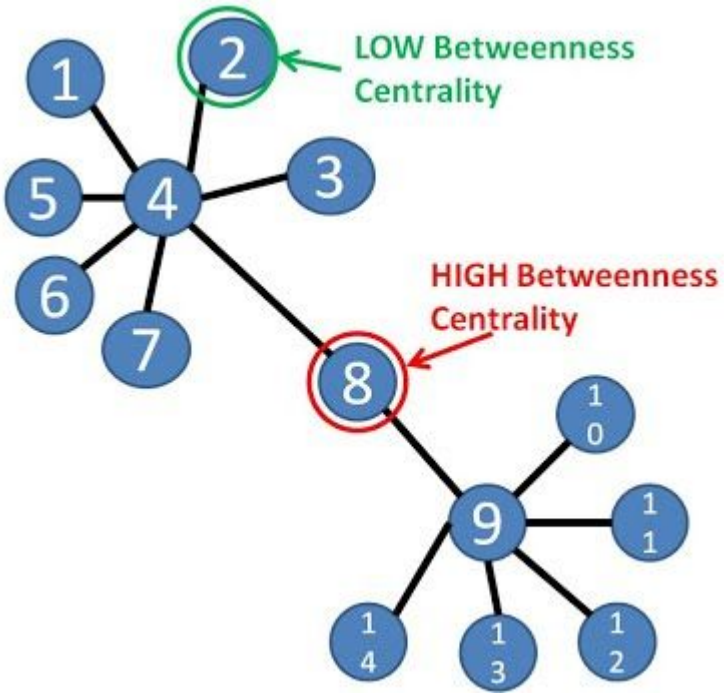
Medidas de centralidad (estructura) de tipo intermediación

Betweenness: el número de veces que cualquier actor necesita a un actor dado para llegar a cualquier otro actor (geodésicas).

K-betweenness (variaciones): limita la longitud de los caminos

Centralidad de Katz: conta el número de walks (ponderados)

Centralidad de autovector:



Una tipología de medidas

	Radial	Medial
Volume	Freeman degree, Sade k-path, Bonacich eigenvector, Katz status, Hubbell status, Hoede status, Doreian iterated, Hubbell, Markovsky et al. GPI, Friedkin TEC, Coleman power, Bonacich power, Burt prestige	Anthonisse rush, Freeman betweenness, Freeman et al. flow betweenness, Friedkin MEC, Newman RWB
Length	Freeman closeness, Stephenson-Zelen information, Friedkin IEC	Borgatti DF

Las medidas **radiales** evalúan el grado de **pertenencia** a un grupo, las medidas **mediales** evalúan la el grado de **intermediación.**

Una tipología de medidas

Todas las medidas evalúan la participación de un nodo en la estructura de caminos de una red.

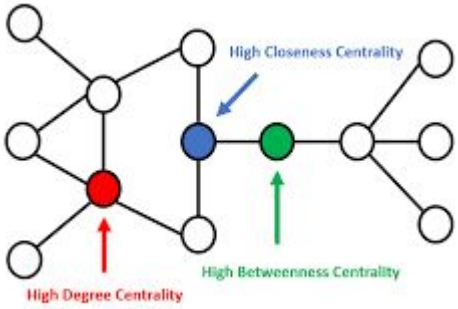
Cuatros dimensiones básicas:

El tipo de caminos

Las propiedades medidas de los caminos (volumen o longitud)

El tipo de participación del nodo (radial o medial)

El tipo de “resumen” (sumas o promedios)



Una tipología de medidas

Todas las medidas evalúan la participación de un nodo en la estructura de caminos de una red.

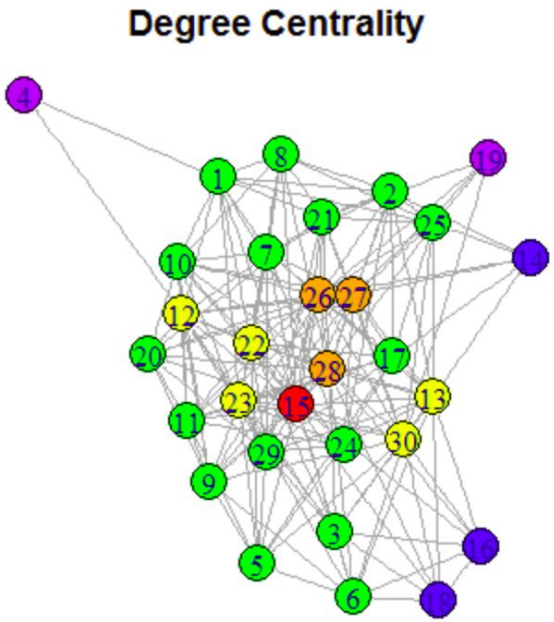
Cuatros dimensiones básicas:

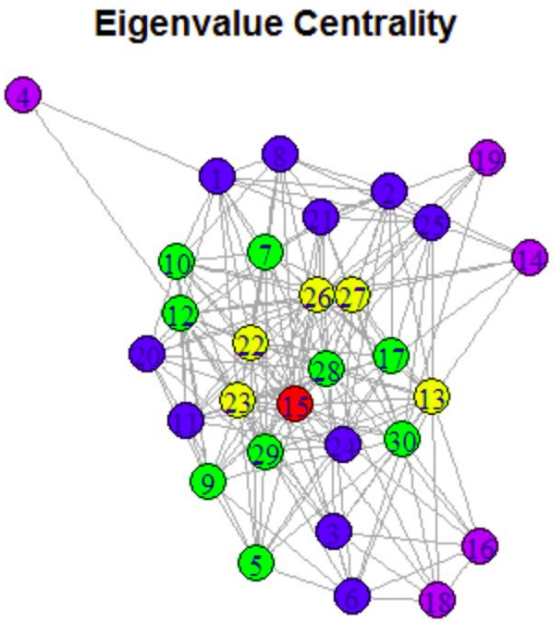
El tipo de caminos

Las propiedades medidas de los caminos (volumen o longitud)

El tipo de participación del nodo (radial o medial)

El tipo de “resumen” (sumas o promedios)

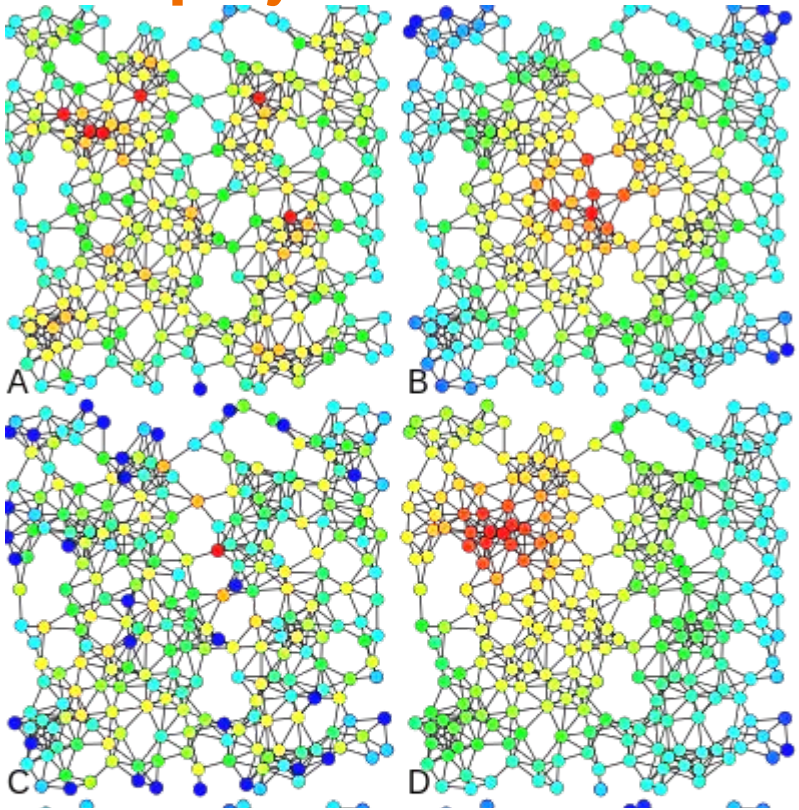




Medidas de centralidad (flujo): tiempo y frecuencia

Medidas de tipo cercanía: el tiempo que tarda en llegar lo que está fluyendo en la red.

Medidas de tipo intermediación: la frecuencia con la que llega

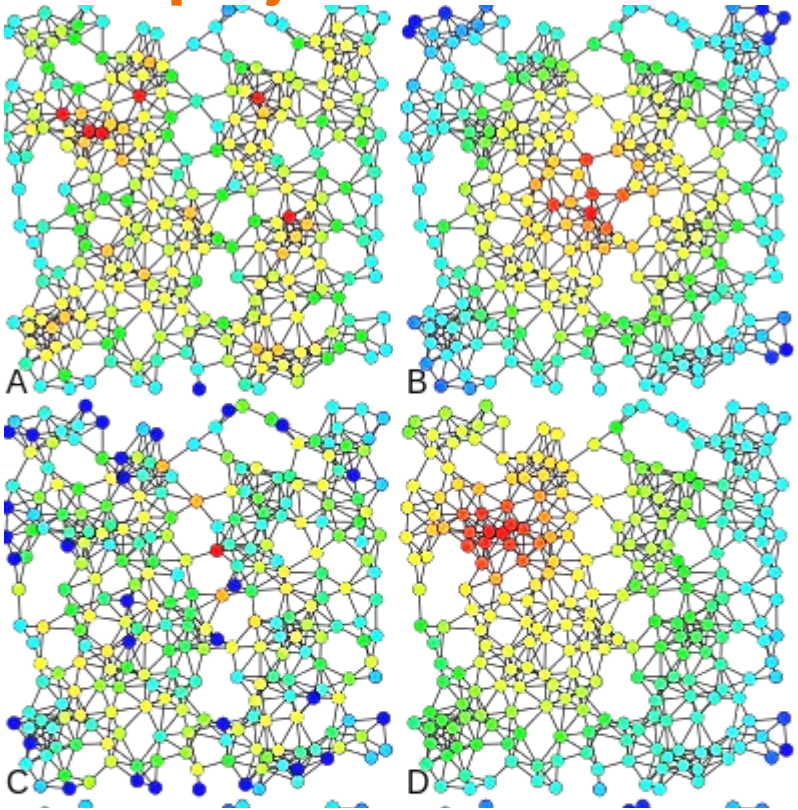


Medidas de centralidad (flujo): tiempo y frecuencia

Medidas de tipo cercanía: el tiempo que tarda en llegar lo que está fluyendo en la red.

Medidas de tipo intermediación: la frecuencia con la que llega

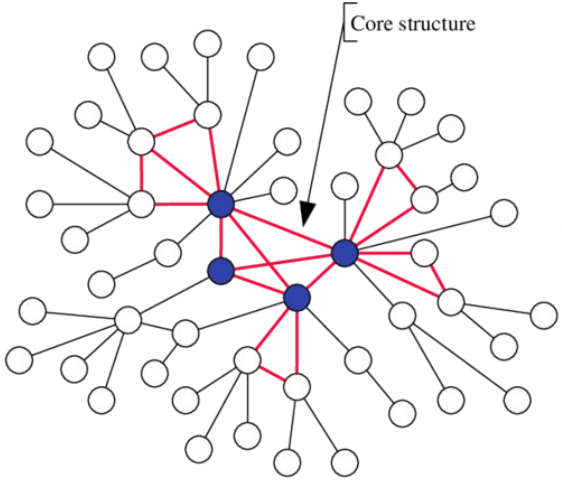
Las medidas basadas en frecuencia son medidas de volumen y mediales, mientras que las medidas basadas en el tiempo corresponden a medidas de longitud y radiales.



Medidas de centralidad

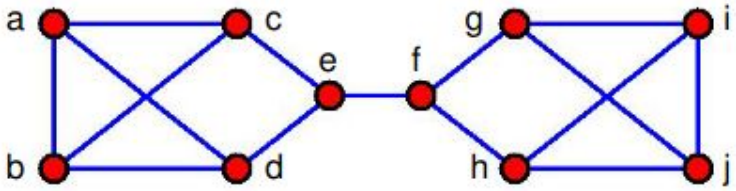
¿En qué condiciones tiene sentido resumir, con un solo valor la cohesión de un nodo con todos los demás?

Centro(s)-Periferia(s)

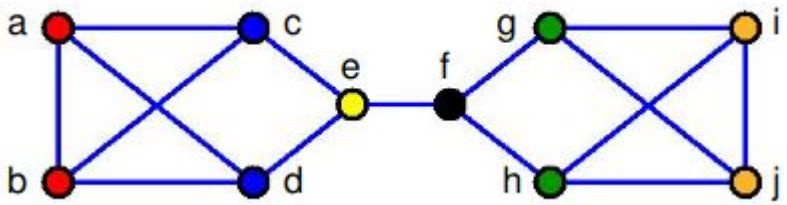


Roles y equivalencias estructurales

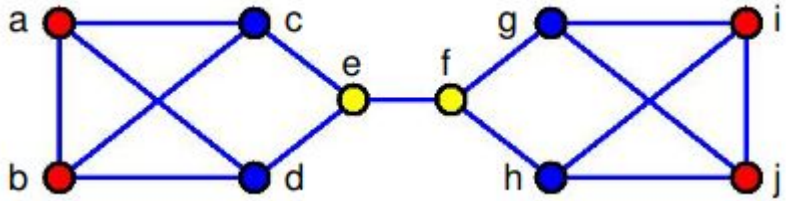
Análisis de roles



Análisis de roles



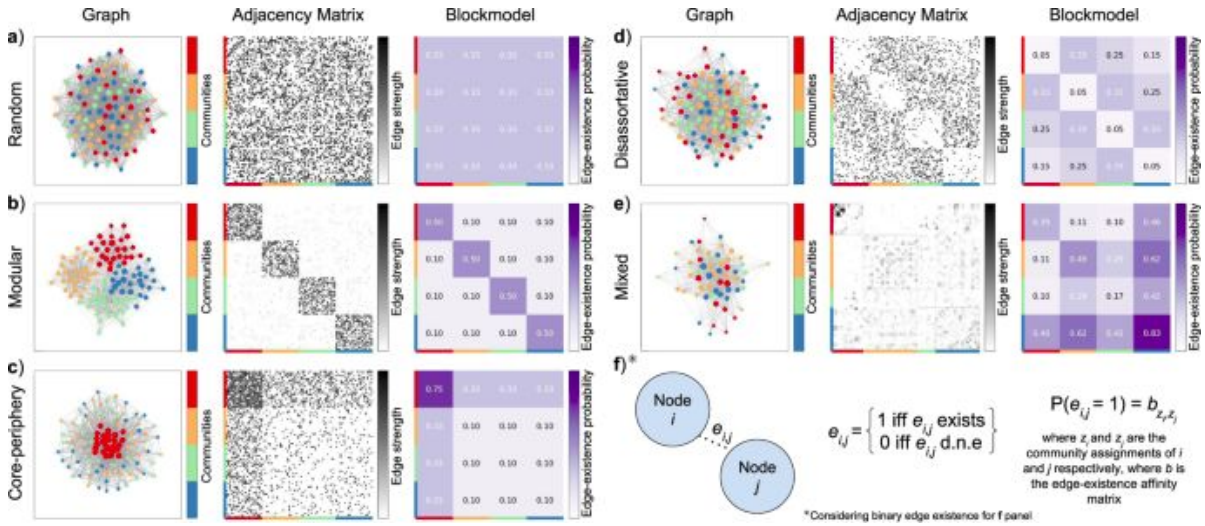
Análisis de roles



Roles y modelos de bloques

La idea del modelado de bloques es sacar algunas características principales de la red dividiendo (particionando) los nodos en categorías de nodos "equivalentes".

La gran pregunta es cuáles son los tipos significativos de equivalencia



Estructura de comunidades

Estructura de comunidades

Se refiere a la organización de los nodos/actores en clústeres o módulos, *con muchos enlaces que unen nodos del mismo clúster y comparativamente pocos enlaces que unen nodos de diferentes clúster.*

Detección de comunidades: un problema central en muchas aplicaciones.

Muy difícil y aún no resuelto satisfactoriamente/unívocamente, a pesar del enorme esfuerzo de una gran comunidad interdisciplinaria durante los últimos años.

En las redes reales, la distribución de los enlaces no solo es global, sino también localmente **no homogénea** -> formación de clúster.

Las comunidades son grupos de nodos que probablemente comparten propiedades comunes.

Órigenes

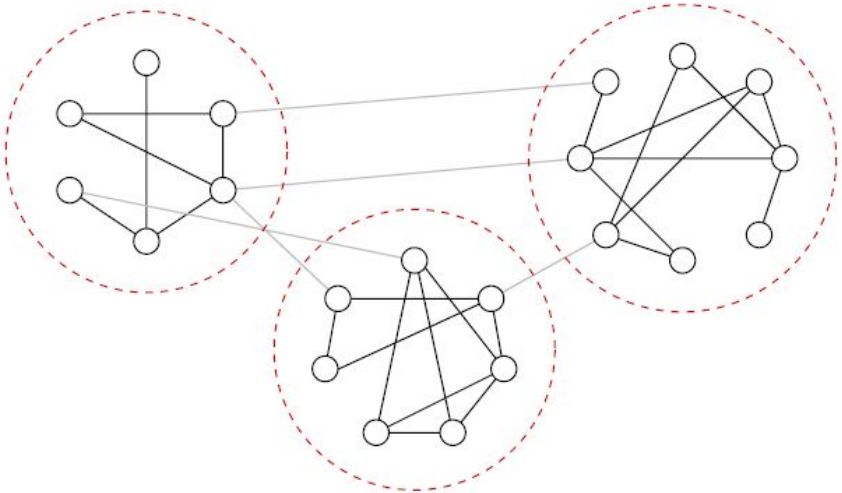
Stuart Rice (1927): grupos de personas en pequeños cuerpos políticos → similitud de patrones de votación

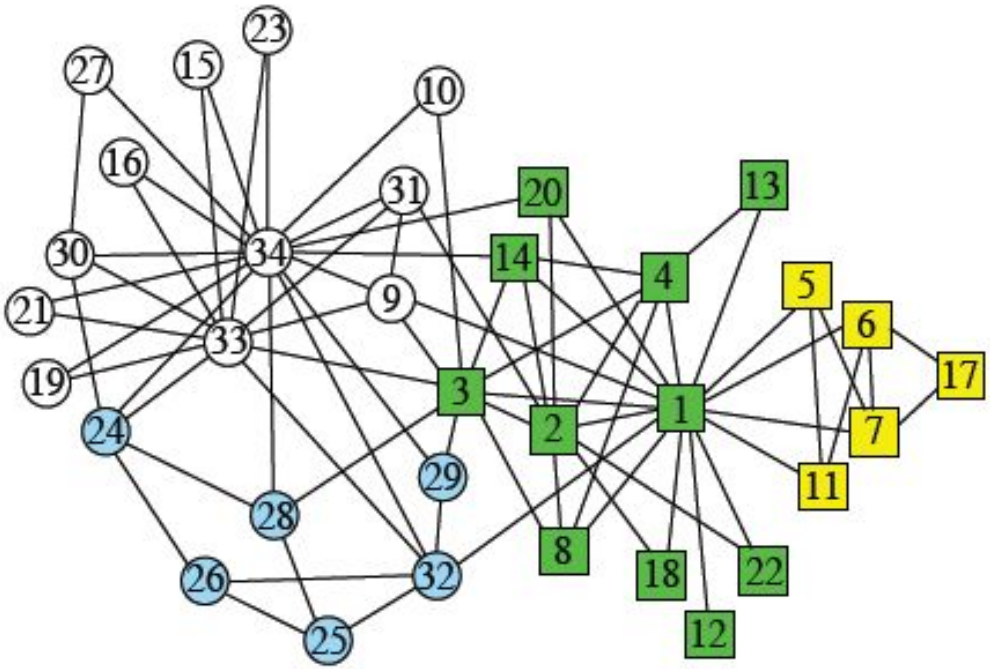
Homans (1950): grupos sociales revelados reordenando filas y columnas de matrices que describen los lazos sociales → forma aproximada de bloques diagonales. Este procedimiento ahora es estándar

Weiss y Jacobson (1955): grupos de trabajo dentro de una agencia gubernamental Matriz de relaciones de trabajo entre miembros identificadas mediante entrevistas privadas. Los grupos de trabajo se separaron eliminando a los miembros que trabajaban con personas de diferentes grupos (conector). Esta idea de cortar los puentes entre grupos está en la base de varios algoritmos modernos de detección de comunidades.

Detección de comunidades

Identificar los módulos y su organización jerárquica (utilizando solo la información codificada en la topología de la red)



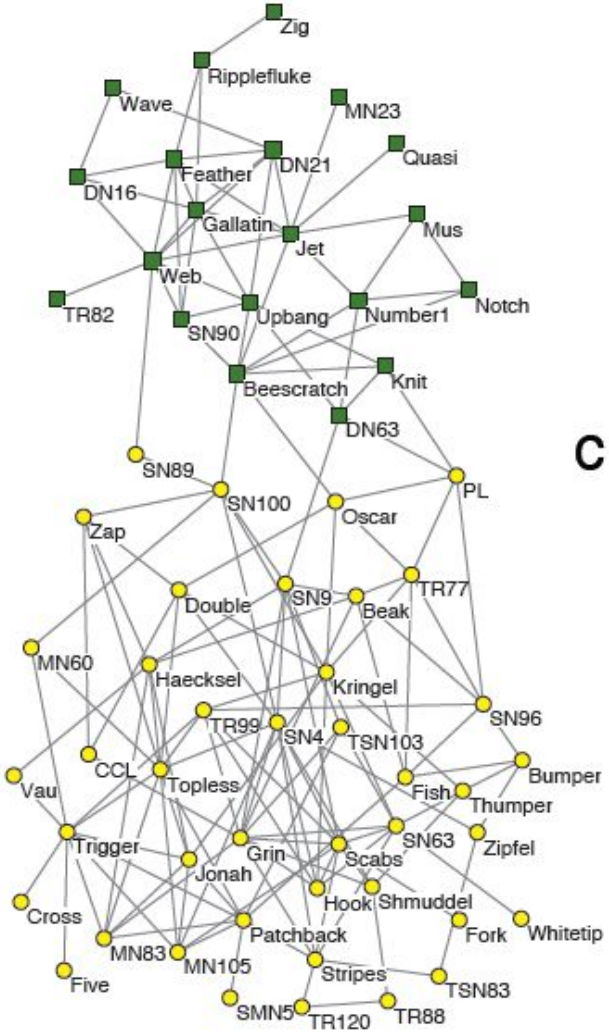

23
27
15
10
16
31
20
13
30
34
14
4
5
21
9
6
33
17
19
3
1
2
7
24
29
11
28
8
18
22
26
32
12
25

Red de delfines que viven en Doubtful Sound (Nueva Zelanda) (Lusseau, 2003).

62 delfines

Enlaces: entre animales que se vieron juntos más a menudo de lo esperado por casualidad.

Los delfines se separaron en dos grupos luego de que un delfín abandonara el lugar por algún tiempo (cuadrados y círculos)



Detección de comunidades

El problema de la partición en comunidades de una red no está bien definido.

Comunidad y partición: cierto grado de arbitrariedad y / o sentido común.

A menudo hay muchas formas igualmente legítimas de resolverlos

En la mayoría de los casos, las comunidades se definen algorítmicamente:

- no hay una definición precisa a priori
- producto final del algoritmo

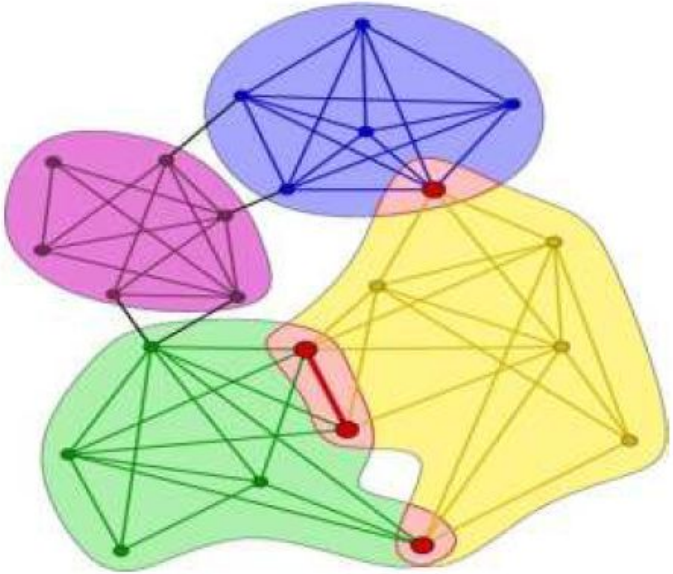
La elección del algoritmo define la respuesta. La pregunta debe orientar la elección del algoritmo.

Métodos de percolación

Desmantelar la red según alguna regla que permita aislar clústeres.

Perspectiva local.

Cliques percolation



Métodos de percolación

Desmantelar la red según alguna regla que permita aislar clústeres.

Perspectiva local

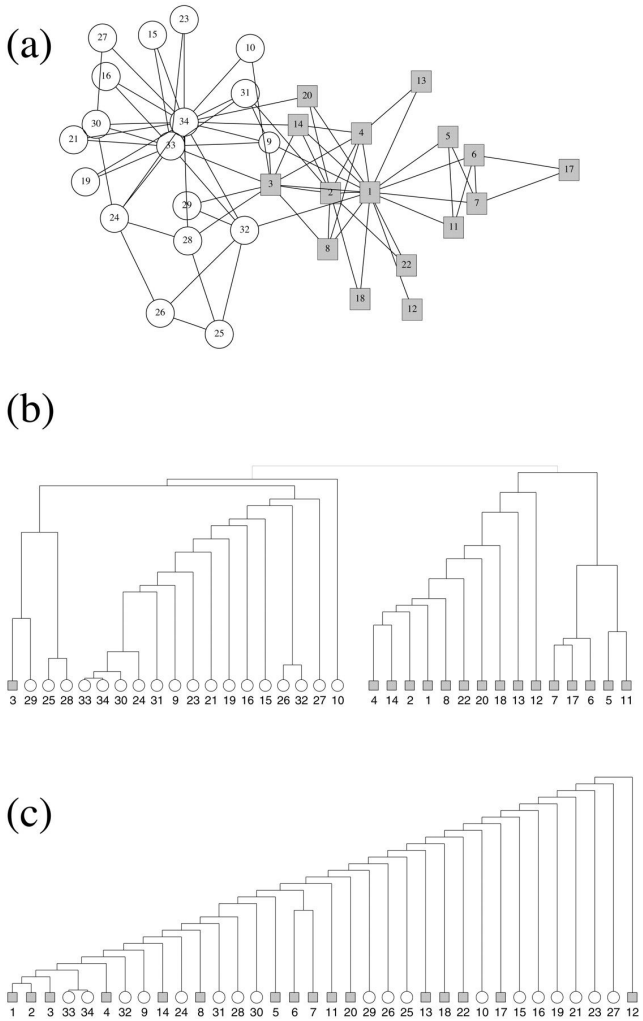
Algoritmo de Girvan-Newman

1. Calcula la betweenness para todos los bordes de la red.

2. Quita el enlace con la betweenness mayor.

3. Vuelve a calcular la betweenness para todos los enlaces afectados por la eliminación.

4. Repite desde el paso 2 hasta que no quedan enlaces.



Métodos de función de calidad

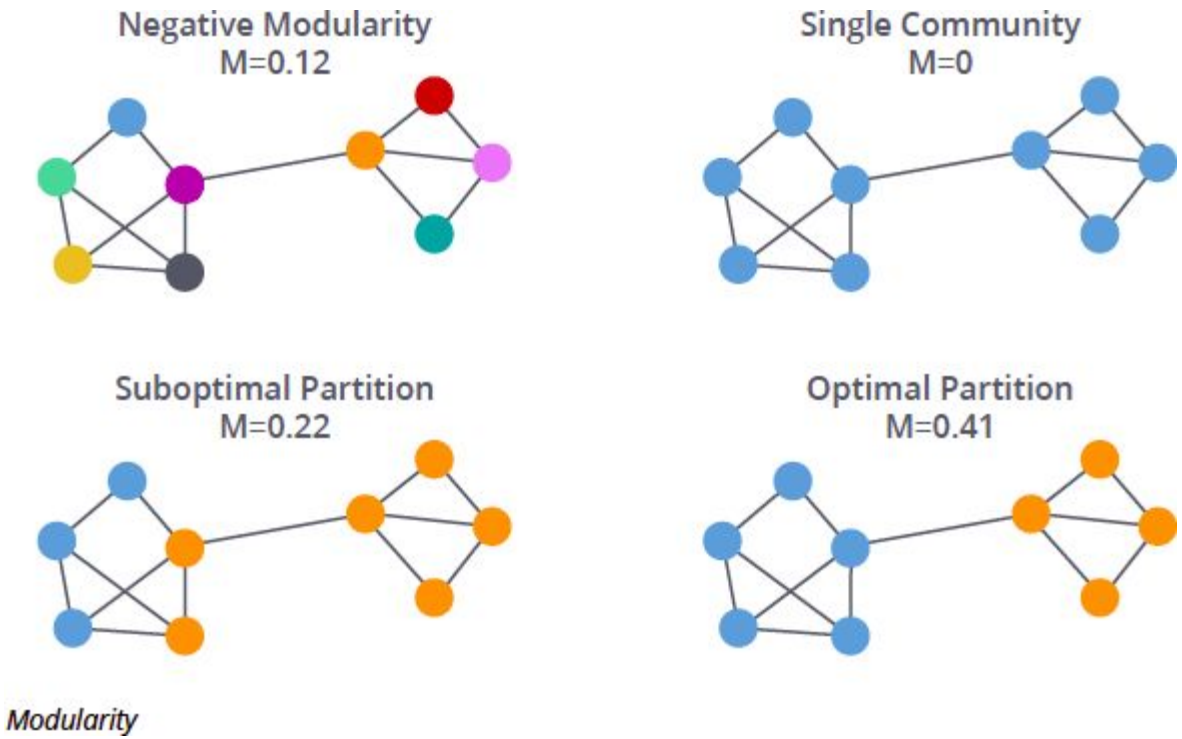
Función de calidad: una función que asigna un número a cada partición de una red. La mejor partición será la que maximize esta función.

Perspectiva Global

Modularidad

Las comunidades son lo que distingue la red en análisis de una red aleatoria.

La posible existencia de comunidades se revela mediante la comparación entre la densidad real de enlaces en un clúster y la densidad que uno esperaría tener en el clúster si los nodos se adjuntaran independientemente de la estructura de comunidades. (Comparación con modelo núlo).



Métodos de función de calidad

Función de calidad: una función que asigna un número a cada partición de una red. La mejor partición será la que maximize esta función.

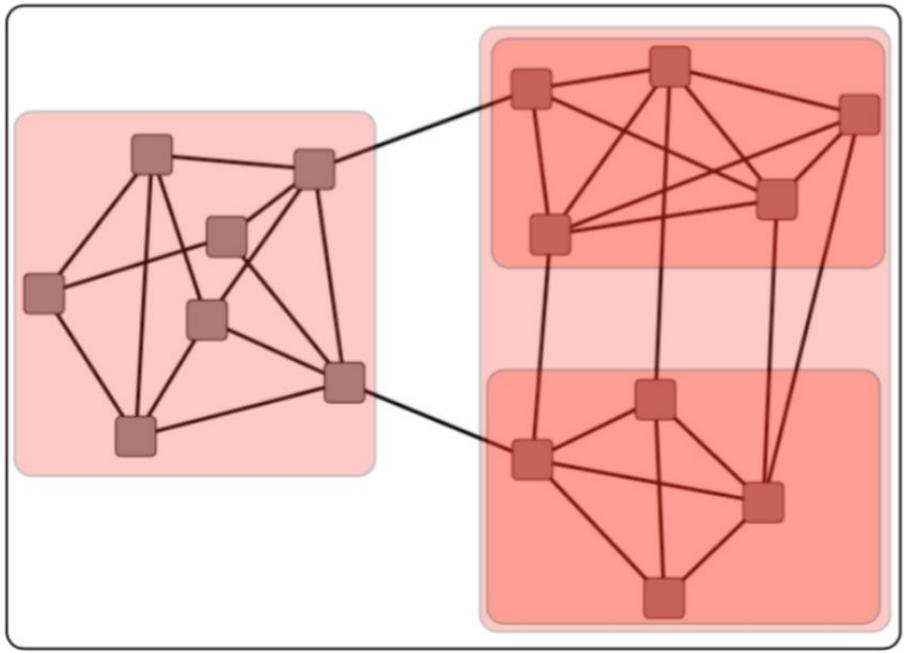
Perspectiva Global

Modularidad

Límite de resolución

Multi-resolución (jerarquia)

$$Q = \frac{1}{2m} \sum_{i,j} \left[A_{ij} - P_{ij} \right] \delta(c_i, c_j)$$



Métodos de función de calidad

Función de calidad: una función que asigna un número a cada partición de una red. La mejor partición será la que maximize esta función.

Perspectiva Global

Modularidad

Límite de resolución

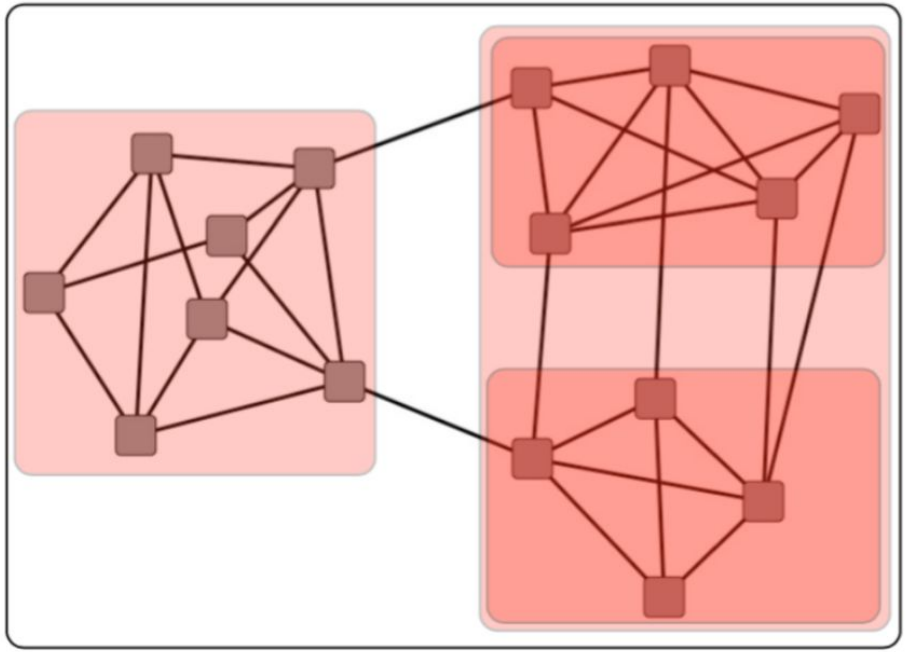
Multi-resolución (jerarquia)

Reichardt & Bornholdt

$$Q_\gamma = \frac{1}{2m} \sum_{i,j} \left[A_{ij} - \gamma P_{ij} \right] \delta(c_i, c_j)$$

Arenas et al.

$$Q(A_{ij} + rI_{ij})$$



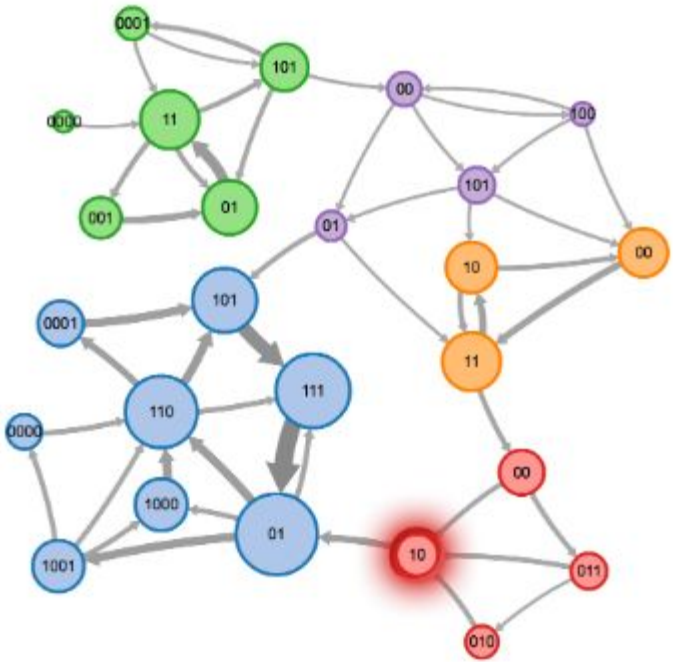
Métodos dínamicos

Los métodos basados en flujos operan sobre la dinámica de la red más que sobre su estructura topológica per se. El fundamento es que la función principal de las redes es capturar el flujo entre los componentes de los sistemas reales que representan. En consecuencia, las comunidades consisten en nodos entre los cuales el flujo persiste durante mucho tiempo una vez ingresado.

Perspectiva Global

InfoMap

Minimizar la información necesaria para describir el recorrido de un caminante aleatorio en la red

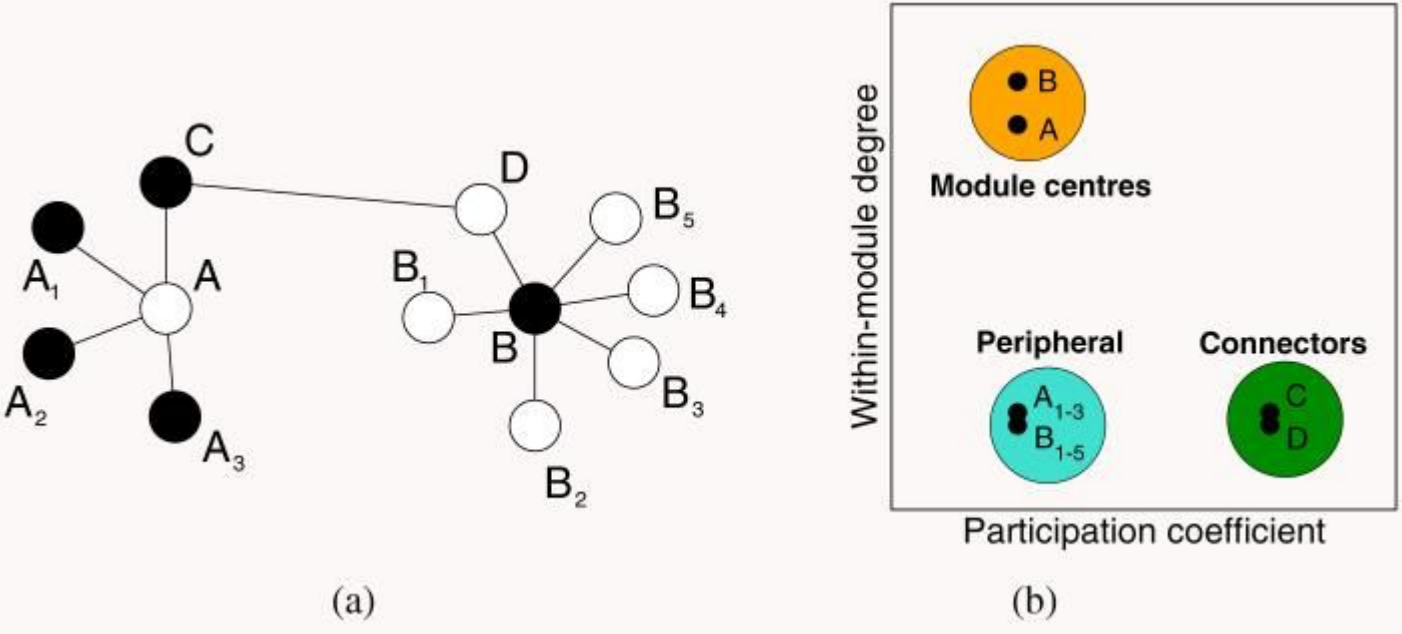


Estructura de comunidades y roles

Roles

roles		INTERACCIONES CON OTRAS COMUNIDADES		
		CASI NULAS	BASTANTE	MUCHAS
ROL EN SU COMUNIDAD	LÍDER	LÍDER DE PROVINCIA	LÍDER CONECTOR	LÍDER DESARRAIGADO
	NO LÍDER	PERIFÉRICO	CONECTOR	DESARRAIGADO
		ULTRA-PERIFÉRICO		

Roles


C
D
A1
A
B1
B5
B4
B
A2
A3
B3
B2
(a)
Within-module degree
B
A
Module centres
Peripheral
Connectors
A1-3
B1-5
C
D
Participation coefficient
(b)

Gephi hands on