

Redes y Sistemas Complejos
Cuarto Curso del Grado en Ingeniería Informática

Tema 3: Redes Sociales. Centralidad

Oscar Cordon García

Dpto. Ciencias de la Computación e Inteligencia Artificial
ocordon@decsai.ugr.es

DEFINICION Y PARTICULARIDADES DE LAS REDES SOCIALES

Una **red social** es un **tipo concreto de red** que modela las relaciones o interacciones existentes entre un conjunto de **entidades sociales**, tales como personas, grupos u organizaciones

Mathematics	Computer Science
Vertex/vertices	Node
Edge	Link/connection



Según Newman, las redes reales se pueden categorizar en cuatro tipos principales:

- redes sociales,
- redes de información o conocimiento,
- redes tecnológicas y
- redes biológicas

M.E.J Newman. The structure and function of complex networks. SIAM Reviews 2003, 45:167–228

Las redes sociales son las que resultan de interacciones humanas y sociales, como las redes de amistad, las redes informales de comunicación entre empresas y las redes de colaboración (p.e., la red de actores de Hollywood y las redes de coautoría científica), entre otras

HISTORIA (1)

Orígenes del Análisis de Redes Sociales

El **Análisis de Redes Sociales** (*Social Network Analysis*, SNA) se centra en el descubrimiento de patrones de interacción entre **actores sociales** en redes sociales

Es un área de investigación metodológica interdisciplinar con contribuciones de la **Sociología**, la **Psicología Social**, la **Antropología**, la Física, las Matemáticas y la Informática, entre otras

Los orígenes del SNA, como una base para el desarrollo de conceptos sociológicos útiles, puede fecharse a **comienzos de la década de 1930**, cuando Moreno desarrolló el **enfoque sociométrico** como una forma de conceptualizar la estructura de las relaciones sociales establecidas entre pequeños grupos de individuos

Estos lazos interpersonales entre miembros de un grupo fueron representados mediante los denominados **sociogramas**, que pueden definirse como gráficos en los que los individuos se representan como nodos y las relaciones entre ellos como líneas

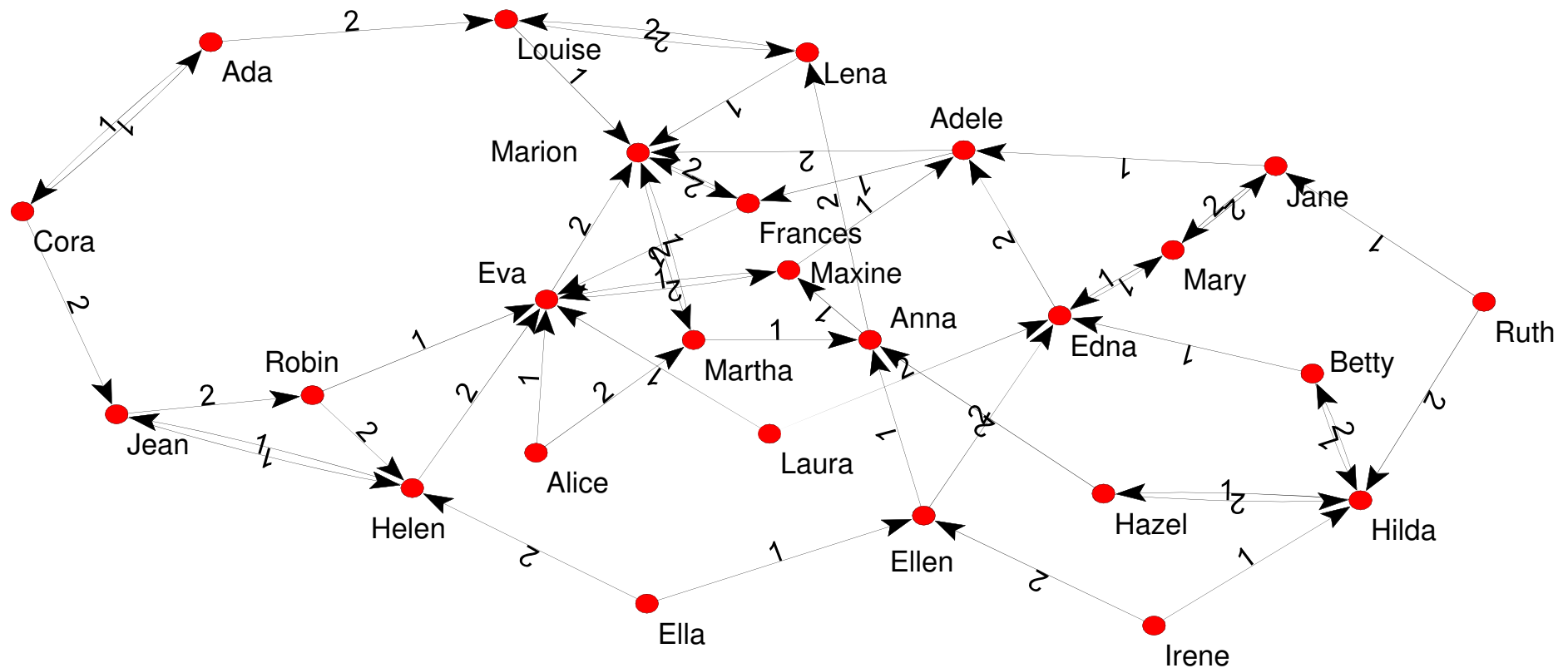
Esos diagramas resultaron ser muy útiles para descubrir las estructuras ocultas de los grupos mediante la identificación de protagonistas, alianzas y subgrupos, entre otras cosas

Moreno JL. Who Shall Survive? New York: Beacon House; 1953

HISTORIA (2):

Ejemplo clásico de red social

1ª y 2ª elección de compañeras de mesa en una residencia femenina
(Moreno, *The sociometry reader*, 1960)



TAREAS DE ANÁLISIS DE REDES SOCIALES

El foco principal del SNA se centra en las **relaciones establecidas entre las entidades** sociales en lugar de en las propias entidades en sí

Su objetivo principal es examinar tanto los contenidos como los patrones de relación en redes sociales para entender las relaciones entre los actores y las implicaciones de esas relaciones

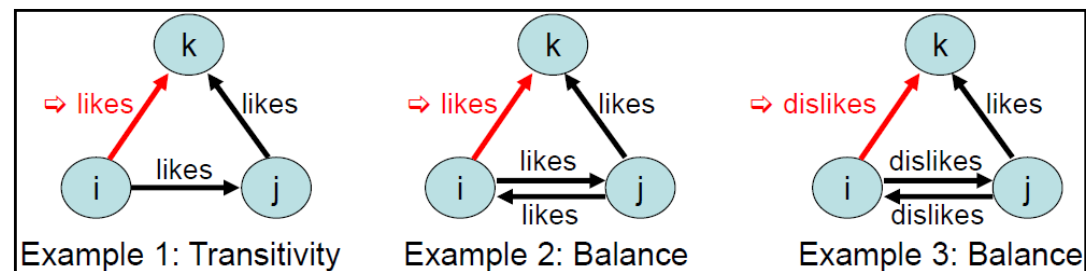
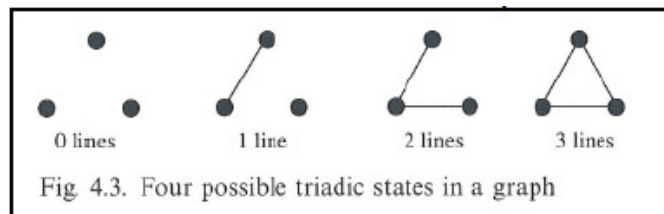
Son tareas habituales del SNA:

- identificar los actores más **influyentes**, **prestigiosos** o **centrales**, mediante medidas estadísticas,
- identificar **hubs** y **autoridades**, usando algoritmos de análisis de enlaces, y
- Descubrir grupos de actores cohesionados, empleando técnicas de detección de **comunidades**

Wasserman S., Faust, K. **Social Network Analysis. Methods and Applications.** Cambridge University Press; 2008

PRINCIPIOS DEL ANÁLISIS DE REDES SOCIALES

- El SNA se centra tanto en **actores individuales** como en grupos de actores (**díadas** (*dyads*), **triadas** (*triads*) y **sistemas más grandes** (subgrupos o la red completa)) y en **sus relaciones**
- Los actores (y sus acciones) se consideran **interdependientes** en lugar de unidades autónomas independientes
- Las ligaduras relacionales entre actores son canales para la transferencia o el **flujo** de recursos (materiales o inmateriales)
- Las redes centradas en individuos ven el entorno estructural de la red como la **provisión de oportunidades para o la restricción de acciones individuales**
- Los modelos de redes conceptualizan la estructura (social, económica, política, etc.) de **patrones duraderos de relaciones** entre actores



TIPOS DE REDES SOCIALES

Las redes sociales pueden caracterizarse por la naturaleza de los actores y las propiedades de las relaciones entre ellos

- El número de **modalidades** (*modes*) de una red se refiere al número de tipos distintos de entidades sociales existentes en ella (Wasserman/Faust 2008, pp. 35)
 - **Redes unimodales/de relación:** consideran un único conjunto de actores
 - **Redes bimodales/de afiliación:** se centran en dos conjuntos de actores, o en un conjunto de actores y otro de eventos
- Asimismo, se pueden distinguir el caso en que se considere un único actor o un conjunto de ellos:
 - **Redes egocéntricas:** se centran en un único actor
 - **Redes completas:** consideran conjuntos de actores

TIPOS DE REDES SOCIALES:

Redes unimodales (1)

→ **Redes unimodales:** consideran un único conjunto de actores

Factores importantes en una red unimodal:

- Actores
- Relaciones
- Atributos de los actores

Los **actores** pueden ser de muchos tipos distintos:

- Personas
- Grupos de personas
- Organizaciones
- Colectivos: Comunidades (formadas por grupos de personas), naciones-estados (entidades más grandes que contienen muchas organizaciones y subgrupos)

TIPOS DE REDES SOCIALES:

Redes unimodales (2)

Las **relaciones** también pueden presentar muchas formas:

- Valoraciones individuales: amistad, vinculación, respeto → “medidas de afecto positivo o negativo de una persona hacia otra”
- Transacciones o transferencias de recursos materiales: préstamos/apropiaciones; compra/venta, contactos hechos por un actor a otro para asegurar recursos valiosos, transferencia de mercancías, conexiones de soporte social
- Transferencias de recursos no materiales: comunicaciones, envío/recepción de información
- Comunicaciones frecuentes entre actores, donde las relaciones representan mensajes transmitidos o información recibida (p.ej. cotilleos)
- Interacciones: interacciones físicas entre actores o su presencia en el mismo sitio a la misma hora (p.ej. sentarse uno enfrente del otro, ir a la misma fiesta, visitar la casa de una persona)
- Movimiento: físico (migraciones de un lugar a otro), social (cambio de ocupación o de status)
- Roles formales (p.ej. dictados por poder y autoridad en un entorno de gestión)
- Familiares: matrimonio, descendencia

TIPOS DE REDES SOCIALES:

Redes unimodales (3)

Las personas pueden presentar distintos **atributos** susceptibles de ser incluidos en una red social:

- Edad
- Género
- Raza
- Estatus socio-económico
- Lugar de residencia
- Nivel académico
- Etc.

- **Redes bimodales: se centran en dos conjuntos de actores, o en un conjunto de actores y otro de eventos**
- Las relaciones miden conexiones entre actores de un conjunto y del otro
- Se describe como redes bimodales duales (*dyadic*) porque los actores del primer conjunto son distintos de los del segundo

TIPOS DE REDES SOCIALES:

Taxonomía resumen

	Completas	Egocéntricas
Relación (unimodal)	Relaciones entre todos los miembros de un grupo determinado (Población)	Relaciones de los miembros de la población con un individuo específico (<i>Ego</i>)
Afiliación (bimodal)	Relaciones entre los miembros de un grupo con los miembros de otro completamente distinto	Relaciones de dos grupos distintos de entidades con un individuo específico

EJEMPLOS DE APLICACIÓN DEL ANÁLISIS DE REDES SOCIALES

- **Maximización de la difusión “boca a boca” de productos** de una compañía dirigiéndose a los clientes de mayor valor en la red (aquellos con una mayor influencia y soporte)
- Análisis de las redes de llamadas telefónicas en compañías de telecomunicaciones para **identificación de perfiles de los usuarios y recomendación de tarifas personalizadas** de acuerdo a dichos perfiles
- Uso para **predicción de la deserción de clientes** (*churn prediction*) identificando cambios en sus patrones de contactos telefónicos
- **Detección de fraudes**, por ejemplo, en comunicaciones organizacionales (conjunto de datos de Enron) para analizar la frecuencia y la dirección de los envíos de e-mails formales/informales que pueden revelar los patrones de comunicación entre empleados y jefes
Estos patrones pueden ayudar a identificar personas implicadas en actividades fraudulentas

MEDIDAS PARA EL ANÁLISIS DE REDES SOCIALES

TIPOS DE MEDIDAS DE ANÁLISIS DE REDES SOCIALES

De la misma forma que en el análisis de redes en general, el análisis de la estructura de las redes sociales pretende entender el comportamiento de los sistemas complejos (en este caso, *sistemas sociales*) que generan dichas redes

Existen dos tipos de medidas:

- **Medidas locales (a nivel de actores):** Todas ellas están basadas en el concepto general de **centralidad** (redes no dirigidas) o **prestigio** (redes dirigidas), una medida general de la posición de un actor en la estructura global de la red social

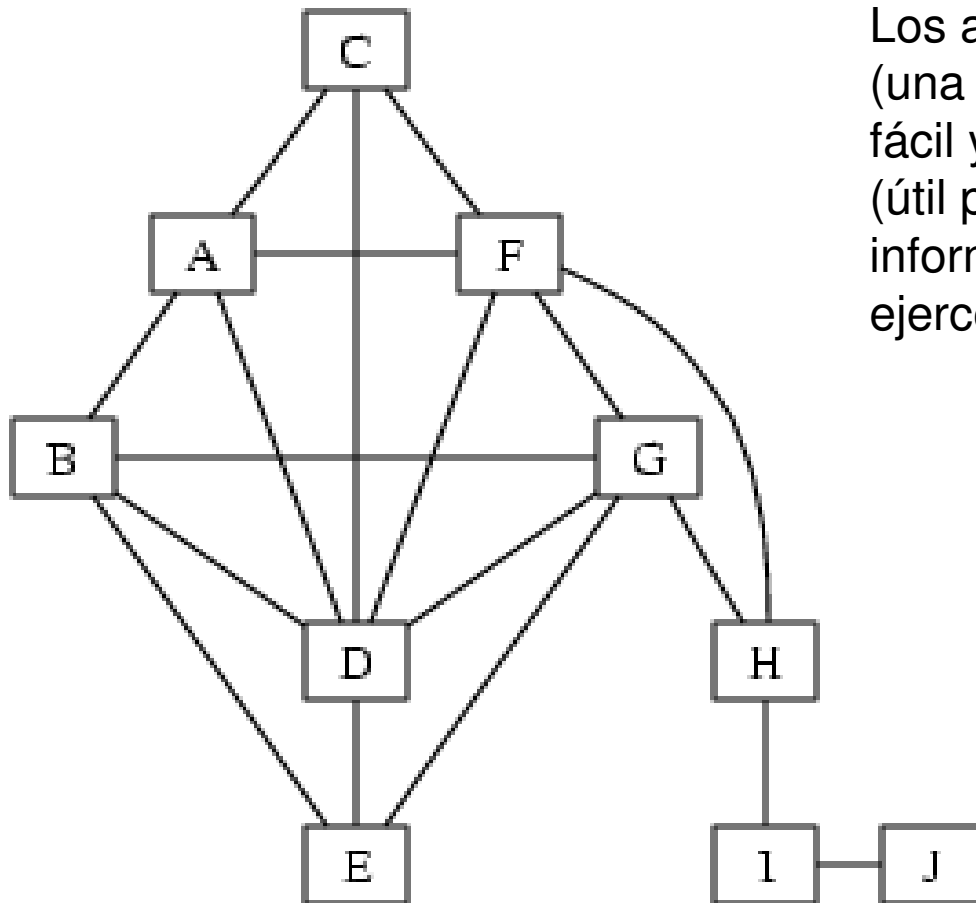
Se usan para identificar los **actores clave** de la red. Muestran como las relaciones se concentran en unos pocos individuos, dando una idea de su *poder social*

- **Medidas globales (a nivel de red):** Proporcionan información más compacta que permite evaluar la estructura global de la red, proporcionando información sobre propiedades importantes de los fenómenos sociales subyacentes

MEDIDAS LOCALES PARA EL ANÁLISIS DE REDES SOCIALES

MEDIDAS LOCALES DE CENTRALIDAD

Concepto de Centralidad

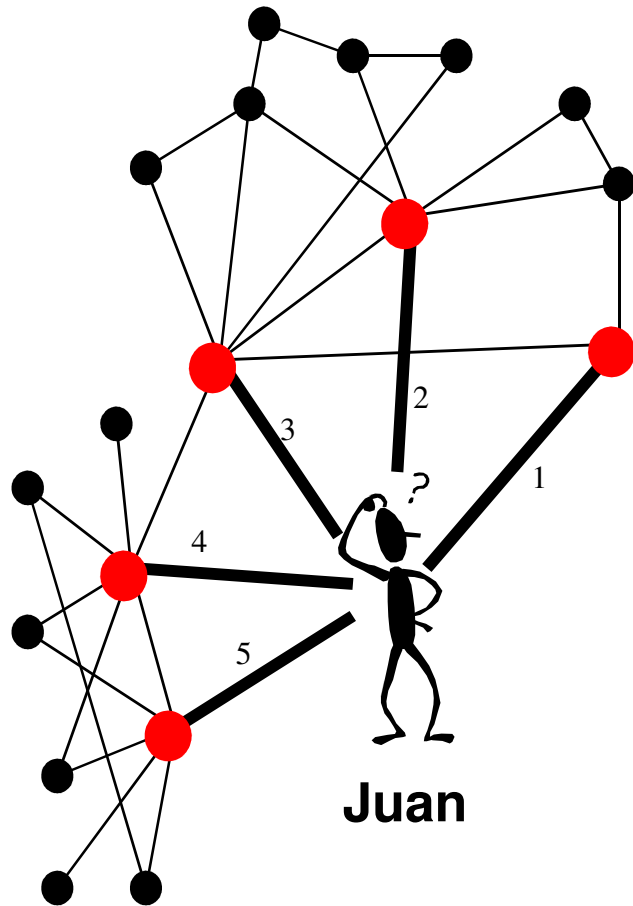


Los actores con una **“posición más central”** (una mayor centralidad) tienen un acceso más fácil y rápido a los demás actores de la red (útil para acceder a recursos como información) y una mayor capacidad para ejercer un control del flujo entre ellos

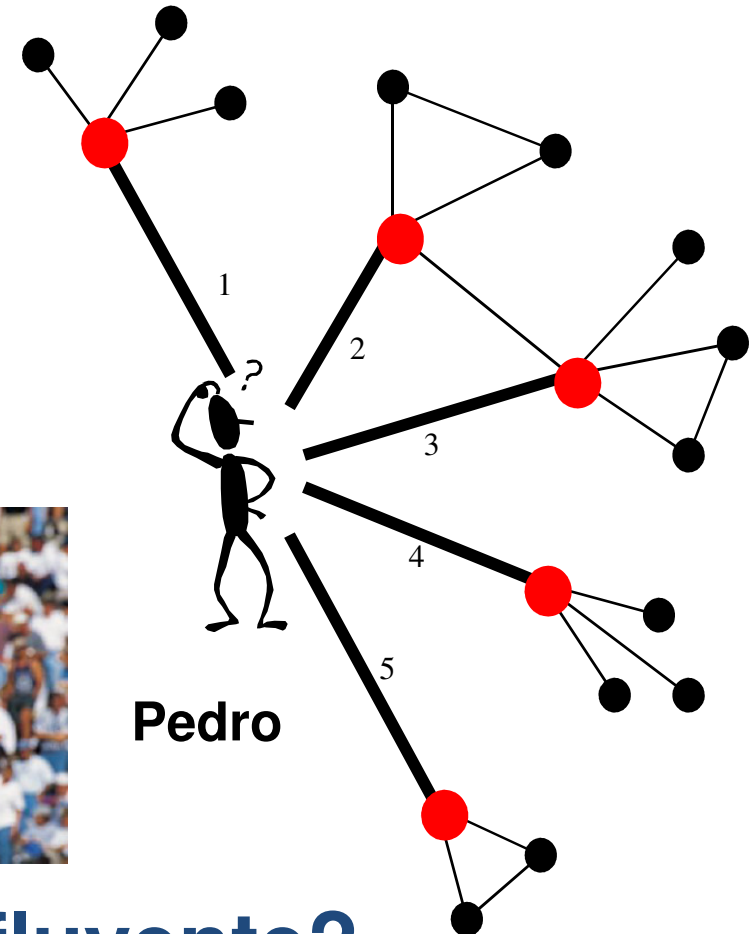
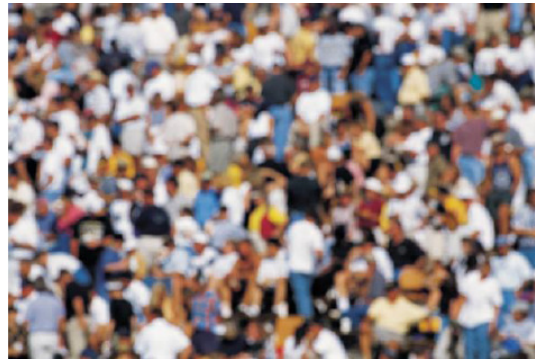
¿Quién es central en esta red?

MEDIDAS LOCALES DE CENTRALIDAD

Concepto de Centralidad



Juan

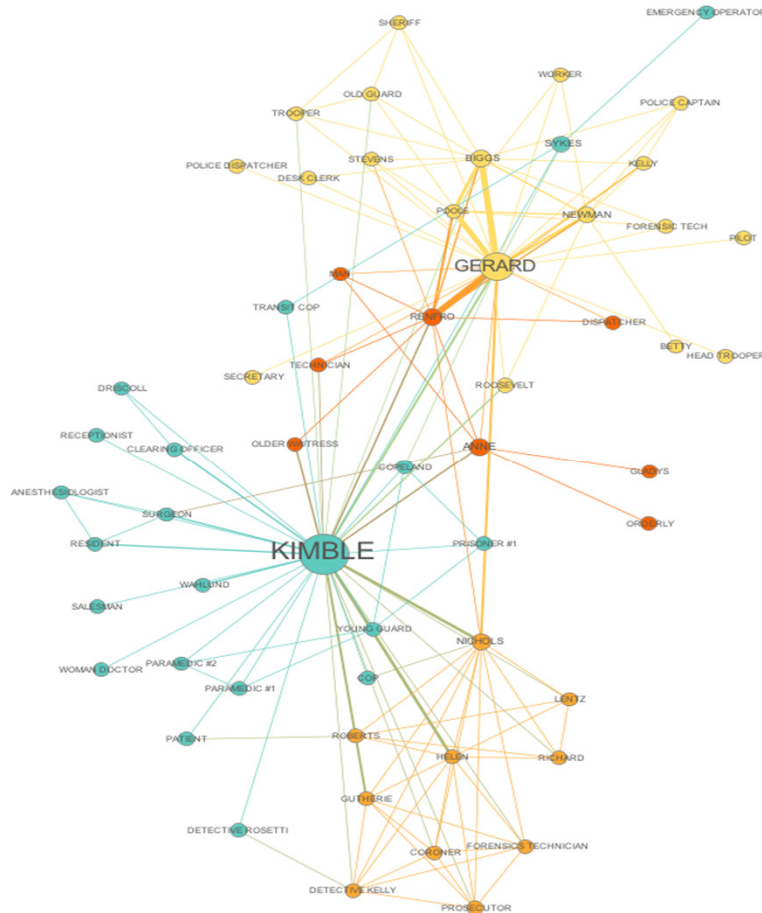


Pedro

¿Quién es más influyente?

MEDIDAS LOCALES DE CENTRALIDAD

The Fugitive (1993)



Ejemplos de Centralidad (1)

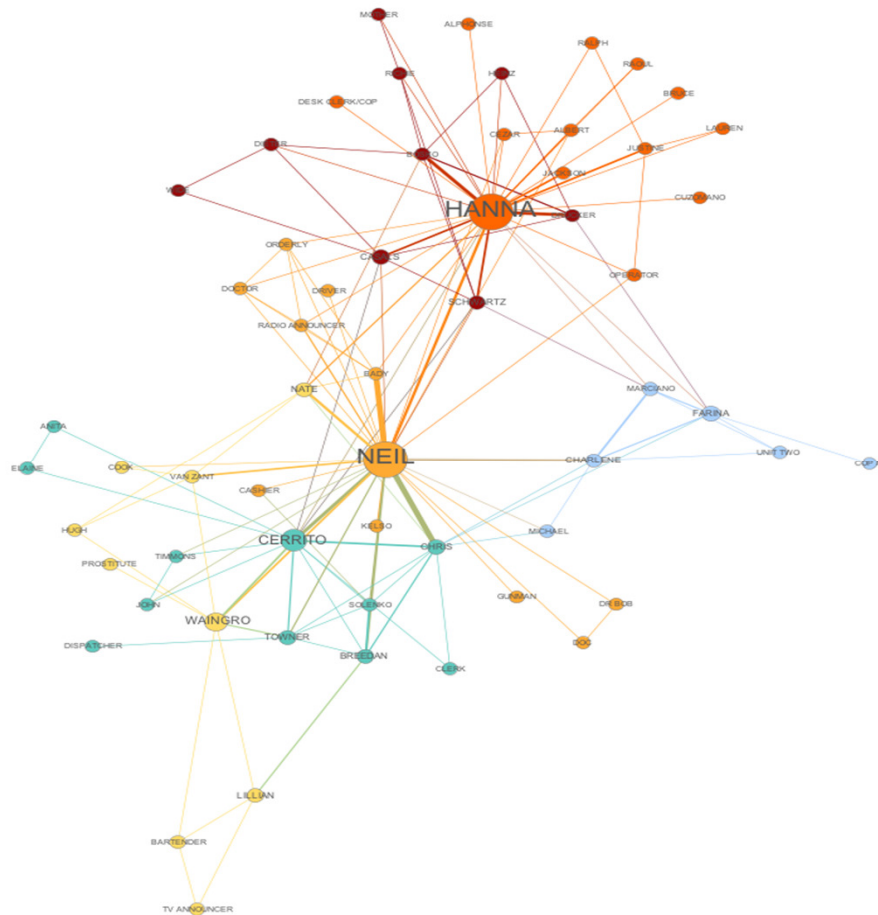
http://es.wikipedia.org/wiki/El_fugitivo_pel%C3%ADcula_de_1993



moviegalexies.com

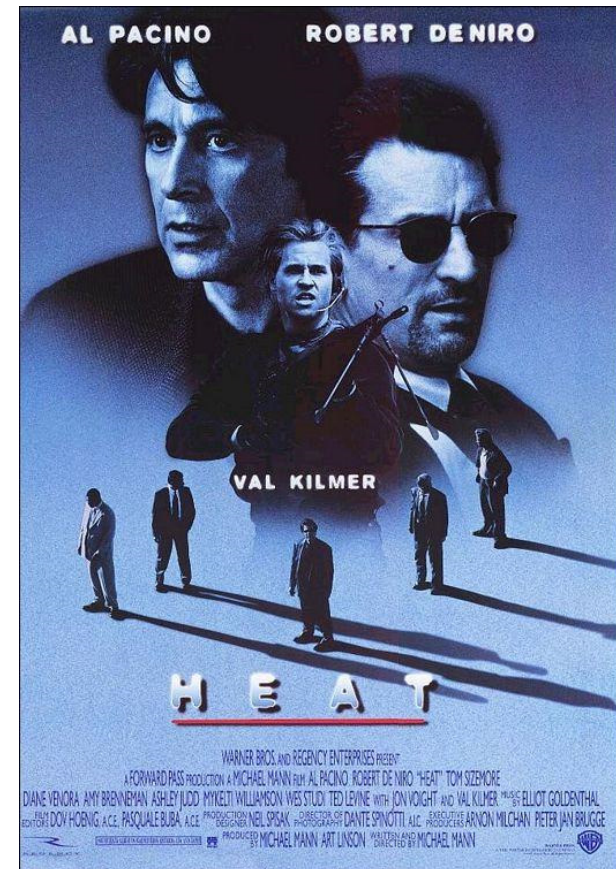
MEDIDAS LOCALES DE CENTRALIDAD

Heat (1995)



Ejemplos de Centralidad (2)

<http://es.wikipedia.org/wiki/Heat>



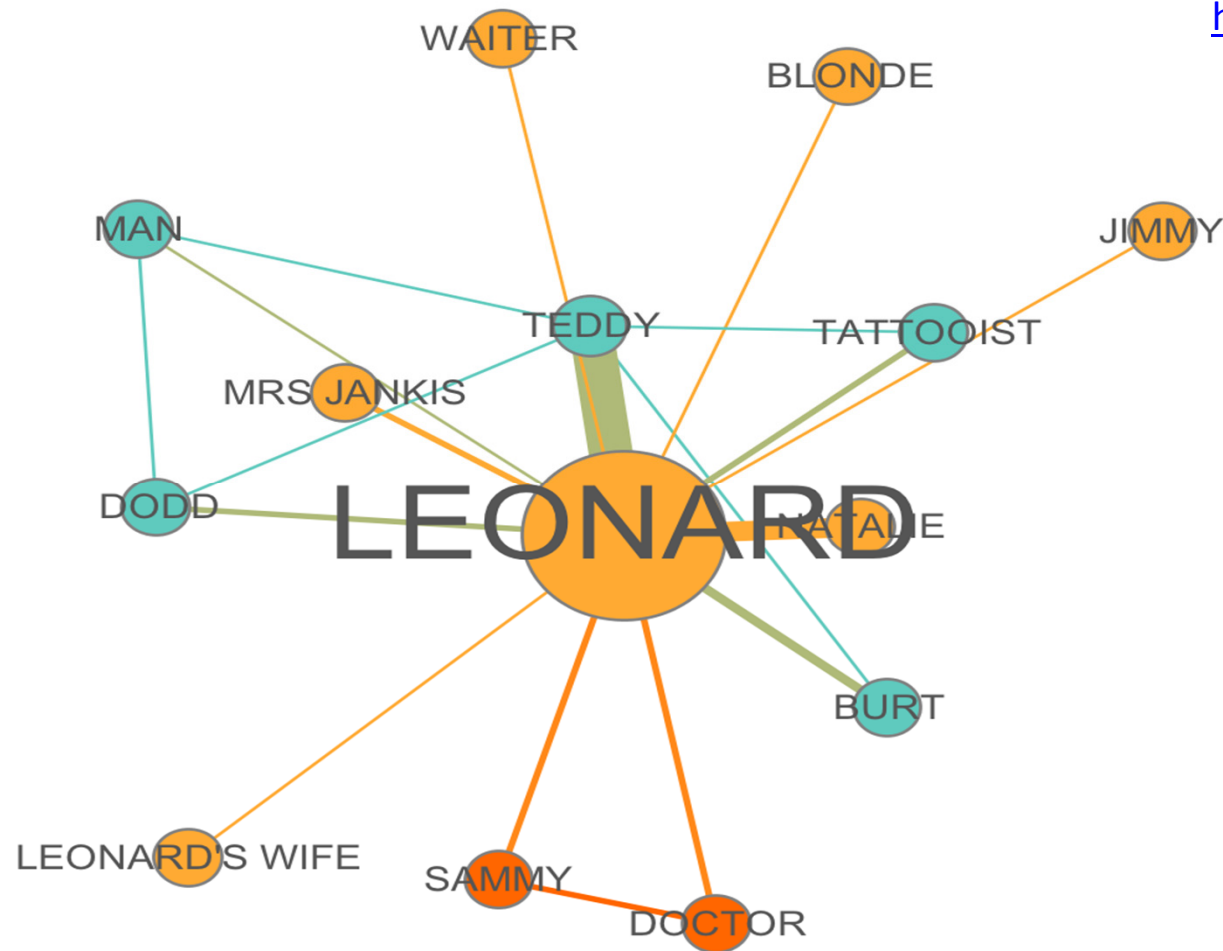
moviegalaxies.com

MEDIDAS LOCALES DE CENTRALIDAD

Ejemplos de Centralidad (3)

Memento (2000)

<http://es.wikipedia.org/wiki/Memento>

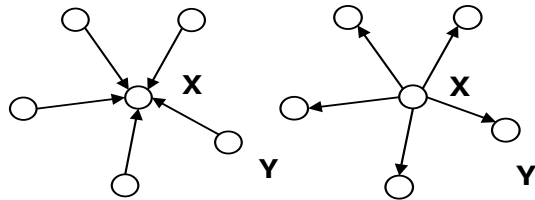


moviegalaxies.com

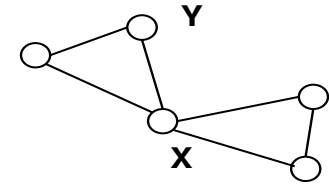
MEDIDAS LOCALES DE CENTRALIDAD (1)

Existen varias medidas distintas de centralidad:

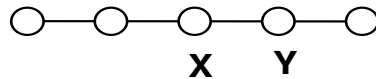
1. Grado:



2. Intermediación (*betweenness*):



3. Cercanía (*closeness*):



4. Excentricidad

5. Centralidad de vector propio

6. Coeficiente de clustering

Las tres primeras fueron propuestas por Freeman para redes no ponderadas (Freeman LC. *Centrality in Social Networks*. *Soc. Netw.* 1 (1979) 215-239). Brin y Page las han extendido recientemente a redes ponderadas (*Soc. Netw.* 32 (2010) 245-251)

La quinta fue propuesta por Bonacich en 1987 (*Am. J. Sociol.* 92 1170-1182)

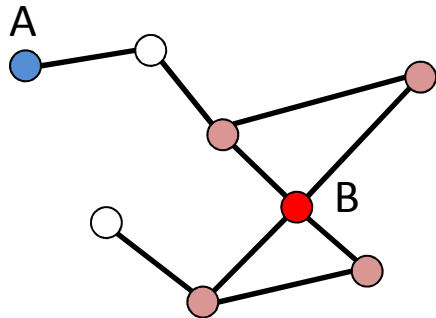
MEDIDAS LOCALES DE CENTRALIDAD (2)

Power Aspect Name	Definition	Influences
Degree	Number of ties for an actor	Having more opportunities and alternatives
Closeness	Length of paths to other actors	Direct bargaining and exchange with other actors
Betweenness	Lying between each other pairs of actors	Brokering contacts among actors to isolate them or prevent connections

MEDIDAS LOCALES DE CENTRALIDAD

Centralidad de grado

No dirigida

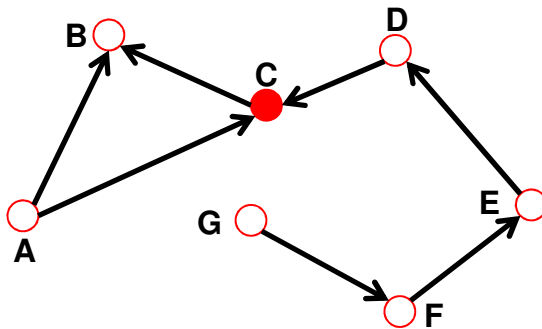


Centralidad de grado de un actor (C_D): número de enlaces que lo conectan con otros

$$C_D(A) = k_A = 1 \quad C_D(B) = k_B = 4$$

$C_D(i)$ se define en $\{0, g-1\}$, siendo g el número de nodos de la componente conexa

Dirigida



En redes dirigidas, se define el **Prestigio de entrada** (*in-degree*), denominado **Soporte**, y el **Prestigio de salida** (*out-degree*), denominado **Influencia**:

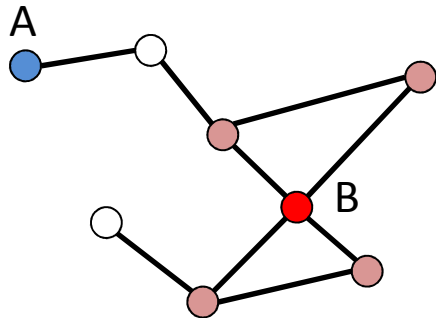
$$P_D^{in}(C) = k_C^{in} = 2 \quad P_D^{out}(C) = k_C^{out} = 1$$

Ambos se definen en $\{0, g-1\}$

MEDIDAS LOCALES DE CENTRALIDAD

Centralidad-Grado en Redes no Dirigidas

No dirigida



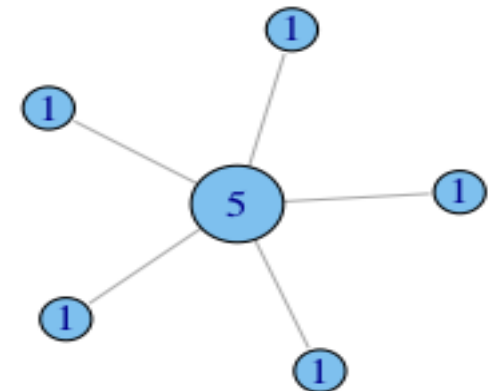
Centralidad de grado de un actor (C_D): número de enlaces que lo conectan con otros

$$C_D(A) = k_A = 1 \quad C_D(B) = k_B = 4$$

$C_D(i)$ se define en $\{0, g-1\}$, siendo g el número de nodos de la componente conexa

Interpretación: *Los actores con más amigos son más centrales*

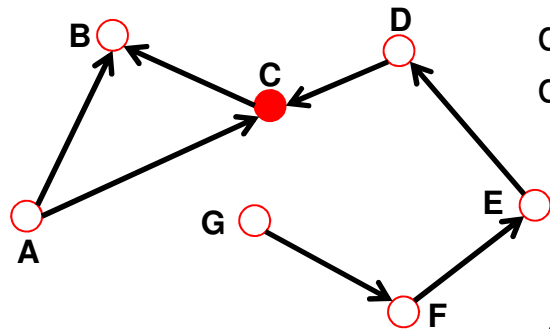
Sólo mide la importancia con respecto a los vecinos más cercanos. Se asume que las conexiones de los amigos no importan, sólo importa lo que ellos pueden hacer directamente (p.ej. “ir a tomarse una cerveza contigo”, “ayudarte a hacer una práctica”, etc.)



MEDIDAS LOCALES DE CENTRALIDAD

Centralidad-Grado en Redes Dirigidas

Dirigida



En redes dirigidas, se define el **Prestigio de entrada** (*in-degree*), denominado **Soporte**, y el **Prestigio de salida** (*out-degree*), denominado **Influencia**:

$$P_D^{in}(C) = k_C^{in} = 2 \quad P_D^{out}(C) = k_C^{out} = 1$$

Ambos se definen en $\{0, g-1\}$

Interpretación Soporte: *Los actores que reciben muchos enlaces son prominentes*

La idea básica es que muchos actores procuran tener enlaces directos a ellos, por lo que se puede considerar como una medida de importancia

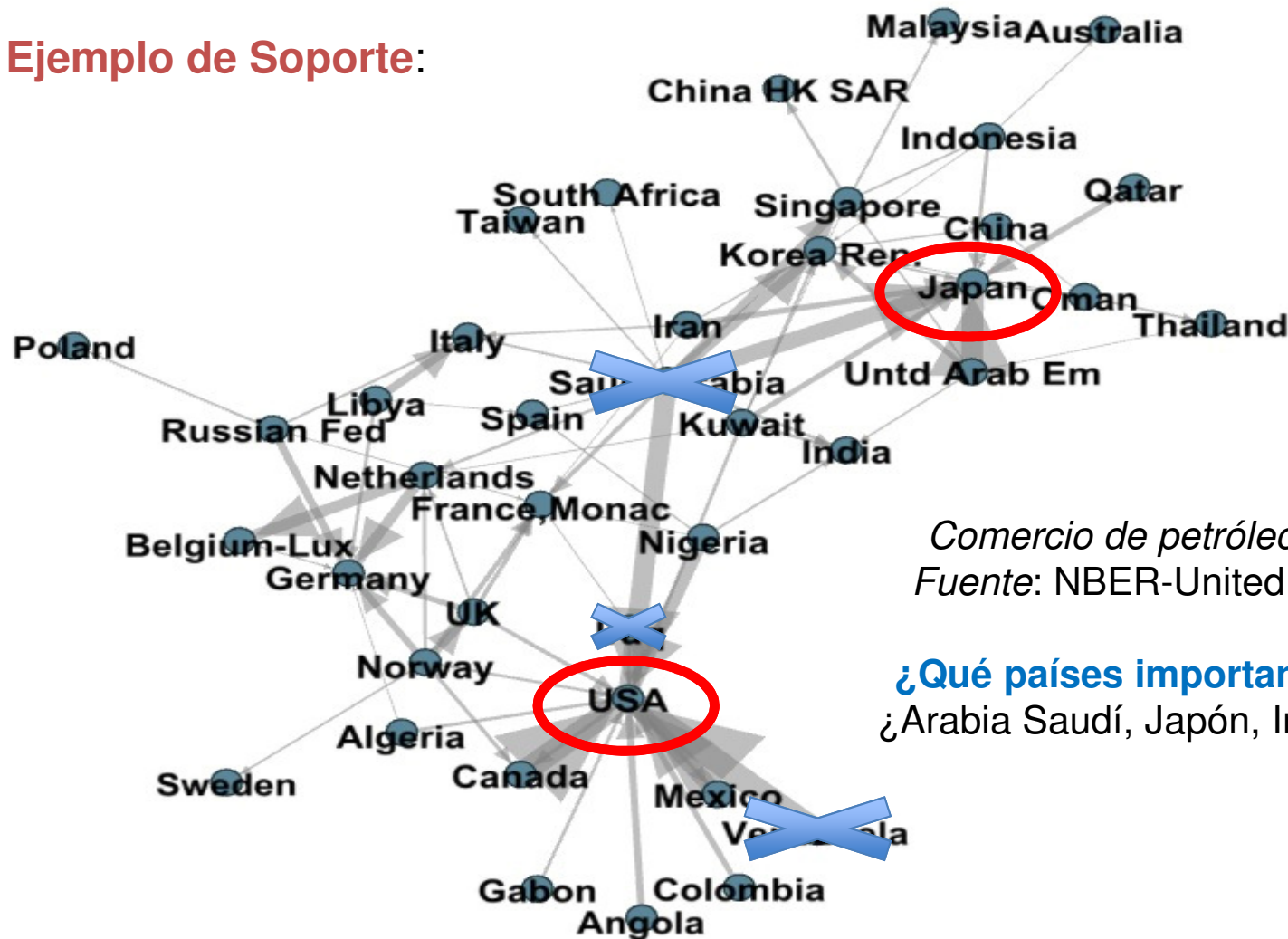
Interpretación Influencia: *Los actores que tienen muchas conexiones directas con otros son influyentes*

Se entiende que pueden intercambiar o transferir información rápidamente a muchos otros (*argumento de la fortaleza de las conexiones débiles*)

MEDIDAS LOCALES DE CENTRALIDAD

Soporte

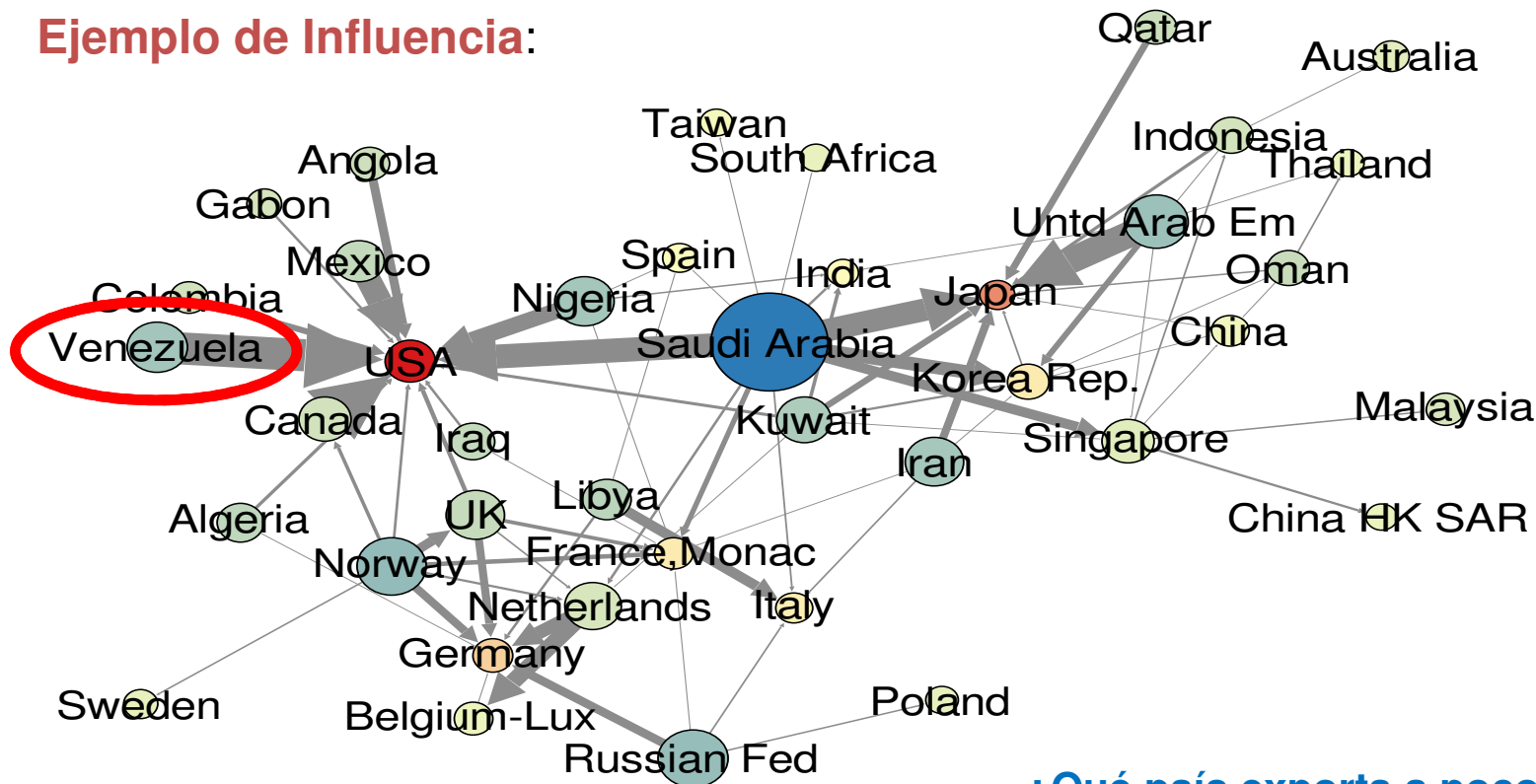
Ejemplo de Soporte:



MEDIDAS LOCALES DE CENTRALIDAD

Influencia

Ejemplo de Influencia:

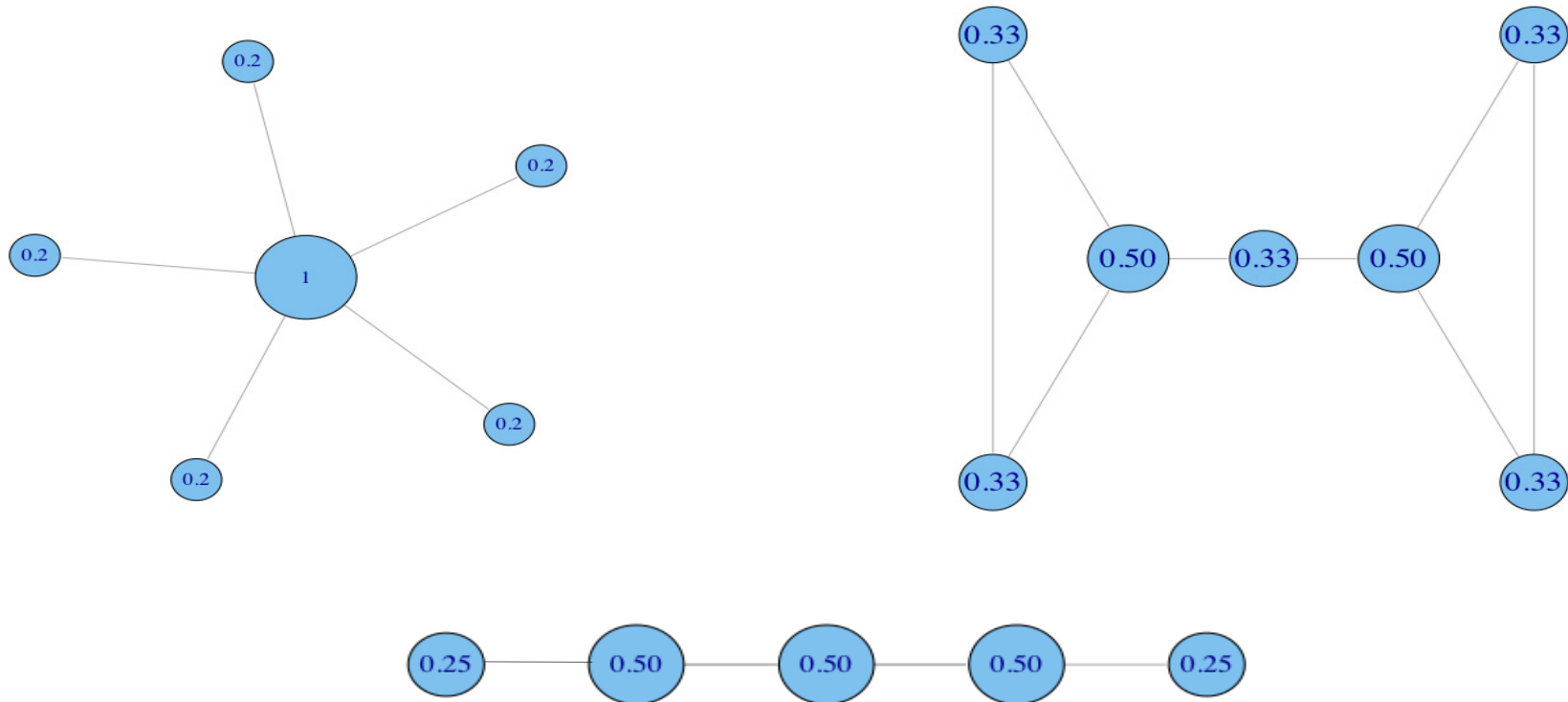


¿Qué país exporta a pocos países (*out-degree* bajo) pero lo hace en gran cantidad (grosor de los arcos = volumen exportado)?
¿Arabia Saudí, Japón, Irak, USA, Venezuela?

MEDIDAS LOCALES DE CENTRALIDAD

Normalización del Grado

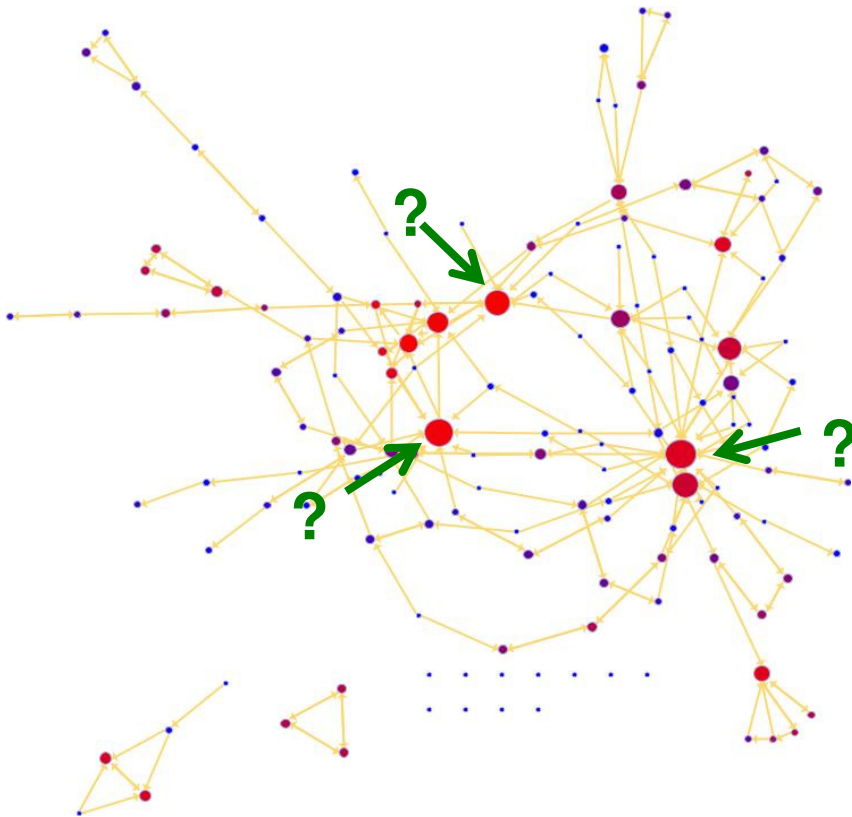
Es habitual normalizar los valores de las Centralidades de grado (C'_D , P'^{in}_D , P'^{out}_D) dividiéndolos por un valor máximo $g-1$ (siendo g el número de nodos de la componente conexa en la que está situado el actor)



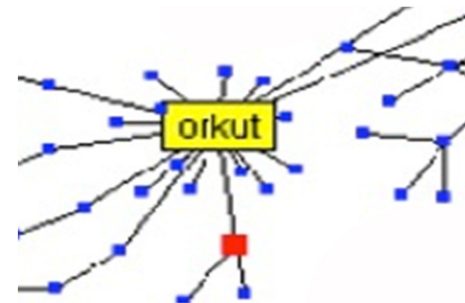
MEDIDAS LOCALES DE CENTRALIDAD

Problemas del Grado (1)

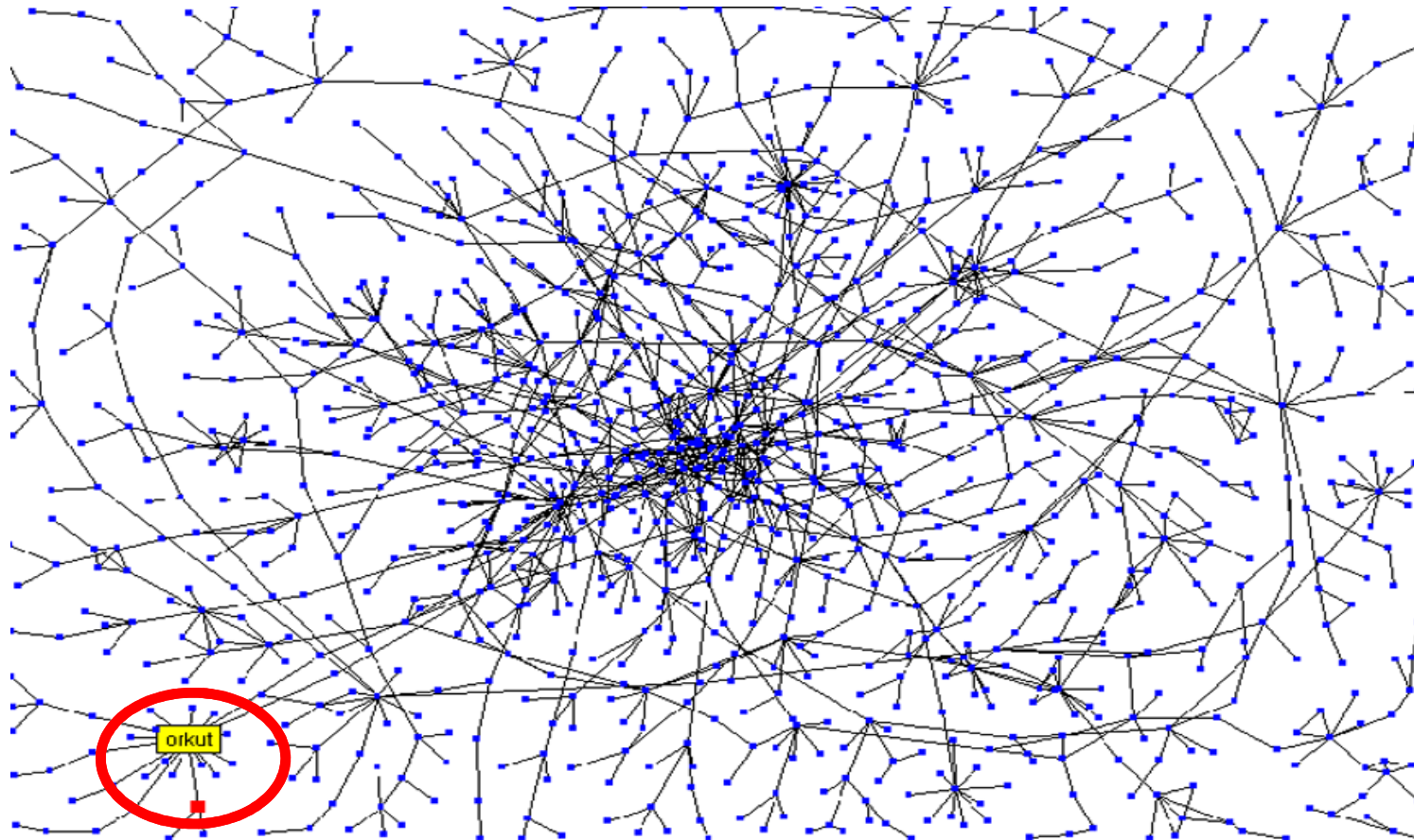
El grado es una medida eficaz para evaluar la centralidad de un actor en una red social **pero desde una perspectiva muy local**: *mide la importancia y la influencia de dicho actor con respecto a sus vecinos más cercanos*



Limitación importante: no tiene en cuenta la estructura global de la red



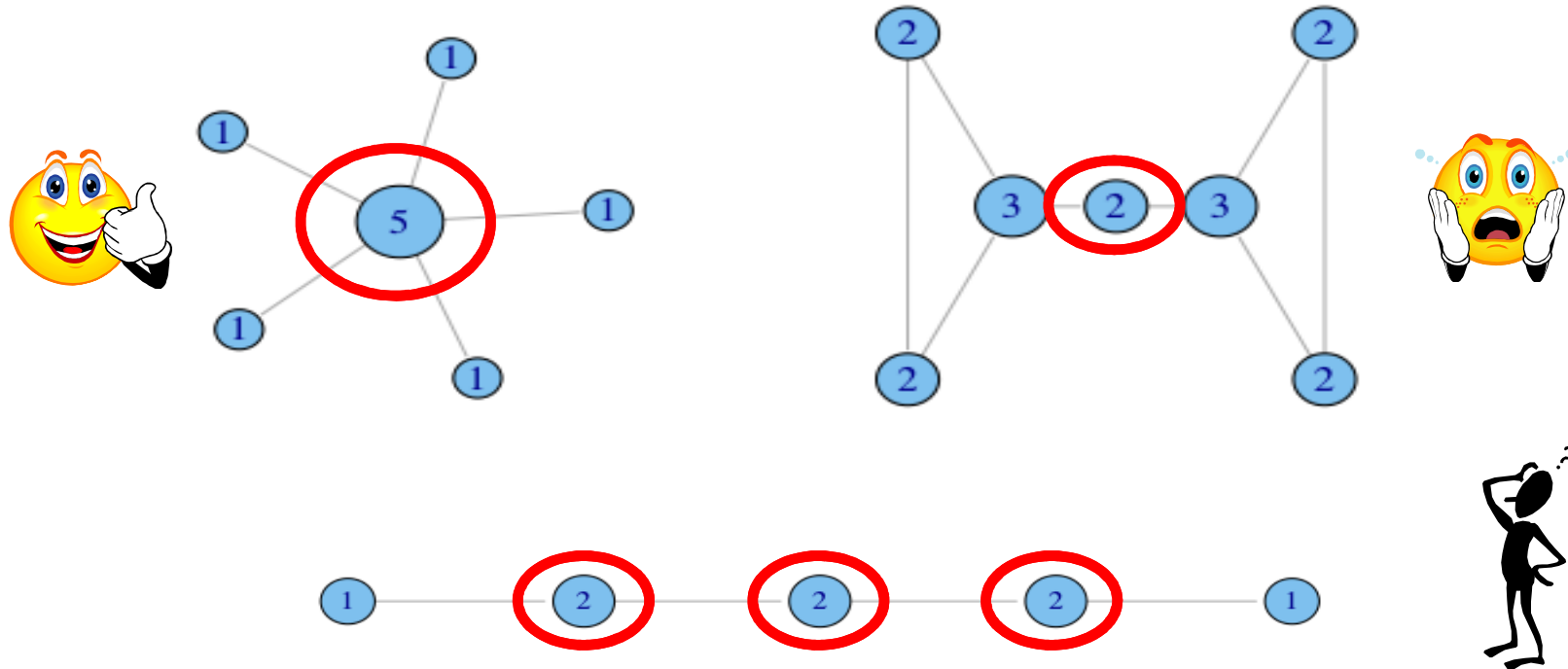
Stanford Social Web (ca. 1999)



Red de páginas web personales en Stanford

MEDIDAS LOCALES DE CENTRALIDAD

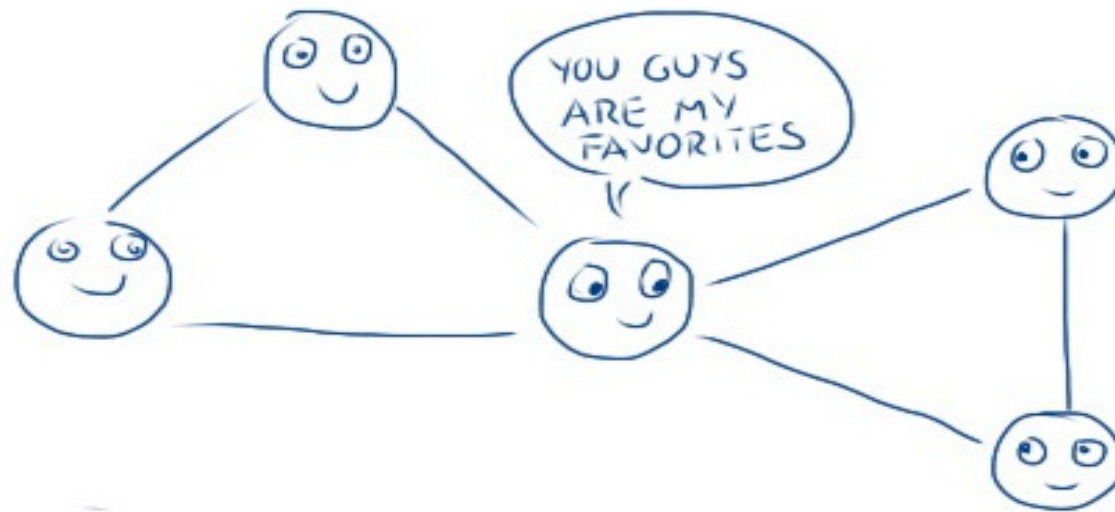
Problemas del Grado (3)



¿Qué problemas presenta el grado para capturar la centralidad intuitiva en estos grafos?



El grado no captura las “corredurías”
(*brokerage*)



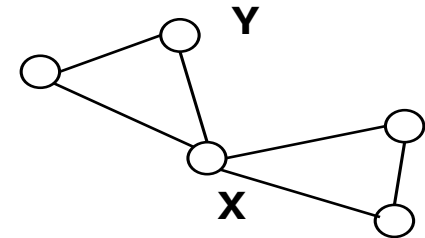
El grado no captura las “corredurías”
(*brokerage*)

MEDIDAS LOCALES DE CENTRALIDAD

Intermediación (1)

La **intermediación** es una medida pensada para capturar la correduría:

$$C_B(i) = \sum_{j,k \in V(G) / v} g_{jk}(i) / g_{jk}$$



donde g_{jk} es el número de caminos mínimos que conectan cualquier par de nodos j y k (normalmente 1) y $g_{jk}(i)$ es el número de esos caminos que incluyen al actor i . $C_B(i)$ se define en $\{0, (g-1) \cdot (g-2)\}$ en redes dirigidas y en $\{0, (g-1) \cdot (g-2)/2\}$ en no dirigidas

Intuición: Ver al actor con una posición más favorable en la medida en que dicho actor esté situado entre los caminos geodésicos de todos los demás. En otras palabras, cuantos mas nodos que necesiten pasar por mi para hacer sus conexiones indirectas por los caminos más cortos, más central seré yo

Es habitual considerar la medida normalizada (redes no dirigidas):

$$C'_B(i) = \frac{C_B(i)}{(g-1)(g-2)/2}$$

Número de pares de actores excluyendo el propio nodo i

MEDIDAS LOCALES DE CENTRALIDAD

Intermediación (2)

Los actores con una intermediación alta ocupan roles críticos en la estructura de la red puesto que suelen ocupar una posición que les permite trabajar con **interfaces entre subgrupos de actores fuertemente unidos**

Son elementos vitales en la conexión entre distintas regiones de la red

En la perspectiva de las redes sociales, **las interacciones entre dos actores no adyacentes pueden depender de otros actores del conjunto, especialmente de aquellos situados en los caminos entre ambos**

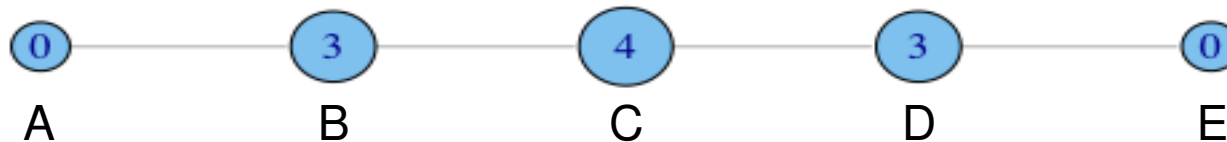
Estos actores se denominan también **porteros** (*gatekeepers*) porque tienden a controlar el flujo de información entre comunidades

La intermediación también puede calcularse para las relaciones, midiendo el grado en que hacen posible otras conexiones. Determina **puentes locales**, enlaces con intermediaciones altas:
$$C'_B(e) = \sum_{j,k \in V(G)} g_{jk}(e) / g_{jk}$$

MEDIDAS LOCALES DE CENTRALIDAD

Ejemplos de Intermediación (1)

Valores de $C_B(i)$ sin normalizar:

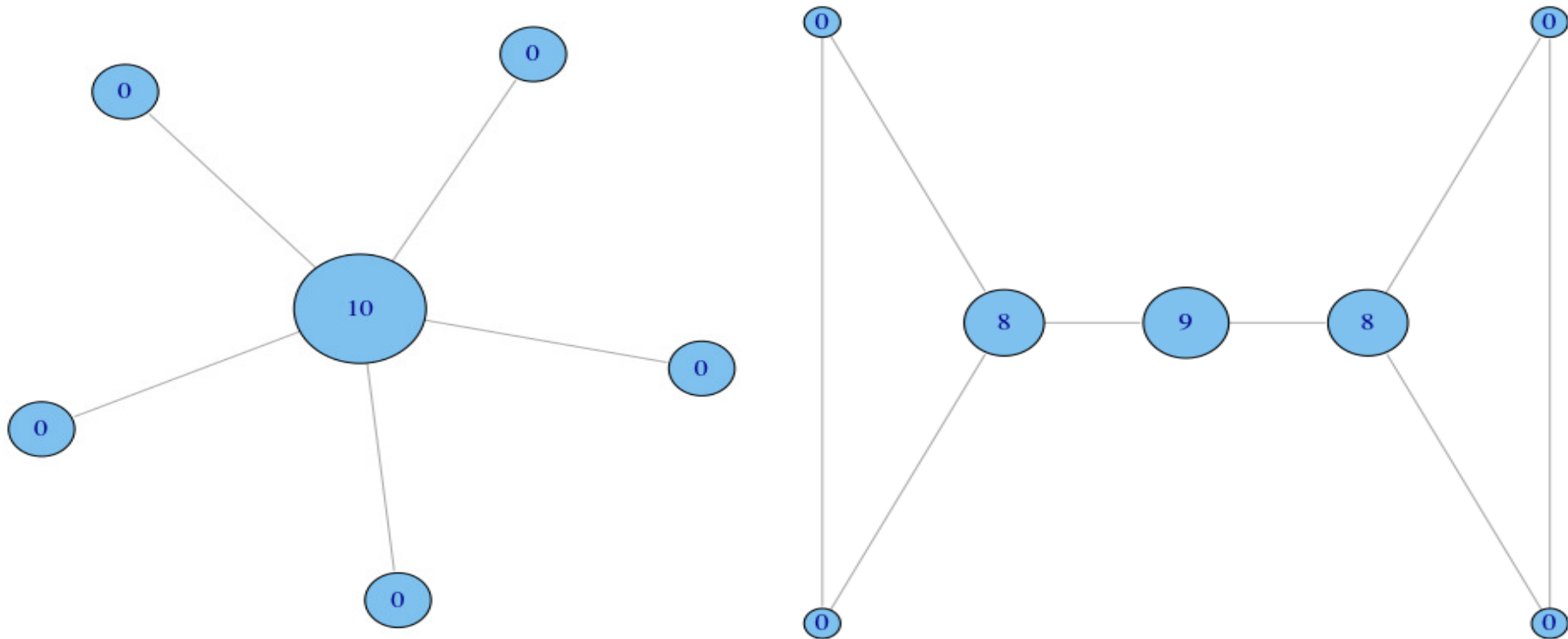


- A y E no están situados entre ningún par de actores
- B está entre A y otros 3 vértices (los caminos mínimos entre A y C, D y E tienen que pasar por él)
- C está entre 4 pares de vértices: (A,D), (A,E), (B,D), (B,E)
Como no hay caminos alternativos para estos pares de nodos, C recibe todo el crédito

MEDIDAS LOCALES DE CENTRALIDAD

Ejemplos de Intermediación (2)

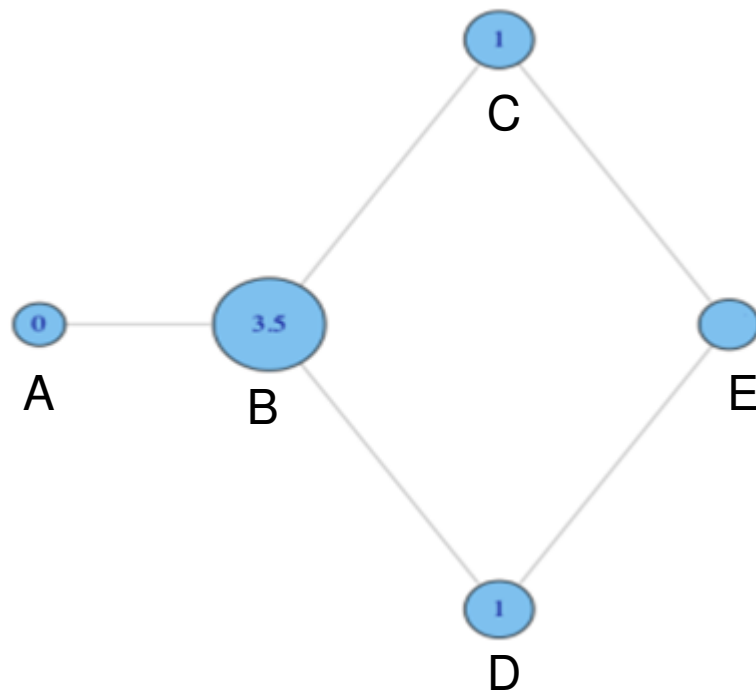
Valores de $C_B(i)$ sin normalizar:



MEDIDAS LOCALES DE CENTRALIDAD

Ejemplos de Intermediación (3)

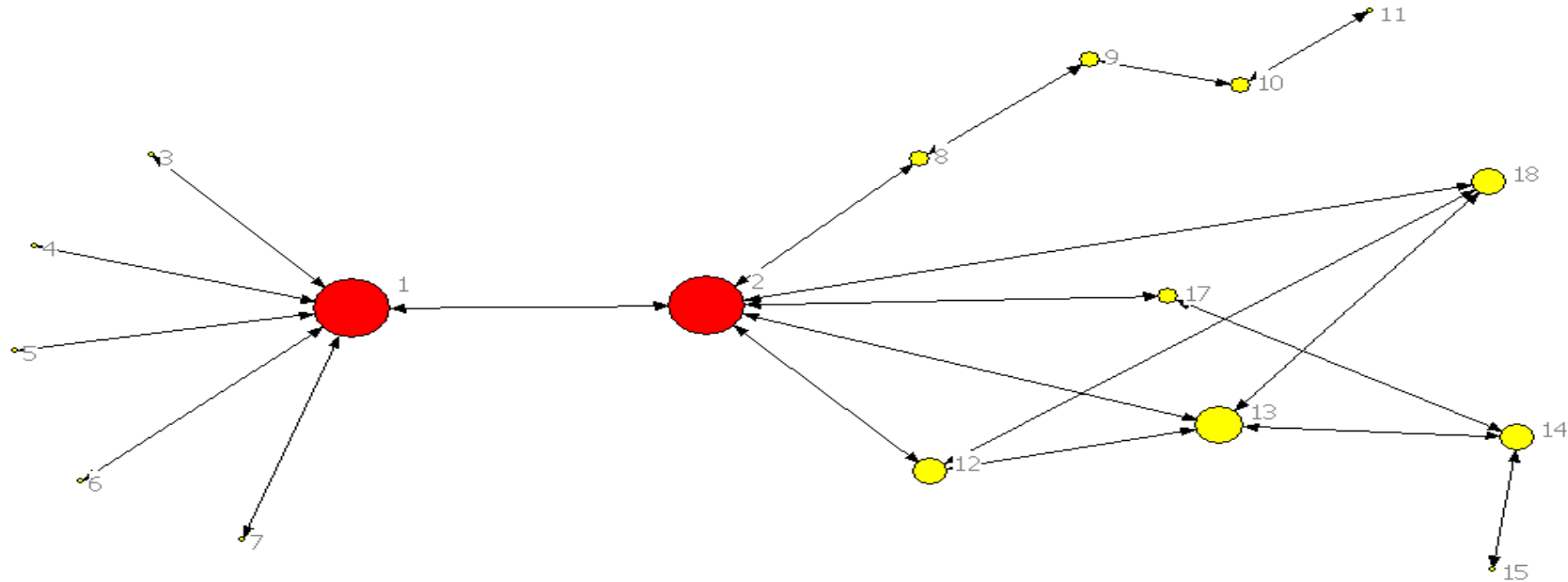
Valores de $C_B(i)$ sin normalizar:



- ¿Por qué C y D tienen intermediación 1?
- Los dos están en los caminos mínimos para los pares (A,E) y (B,E), así que comparten el crédito: $\frac{1}{2} + \frac{1}{2} = 1$
- **¿Cuál es el valor de intermediación de E?**

MEDIDAS LOCALES DE CENTRALIDAD

Grado vs. Intermediación (1)

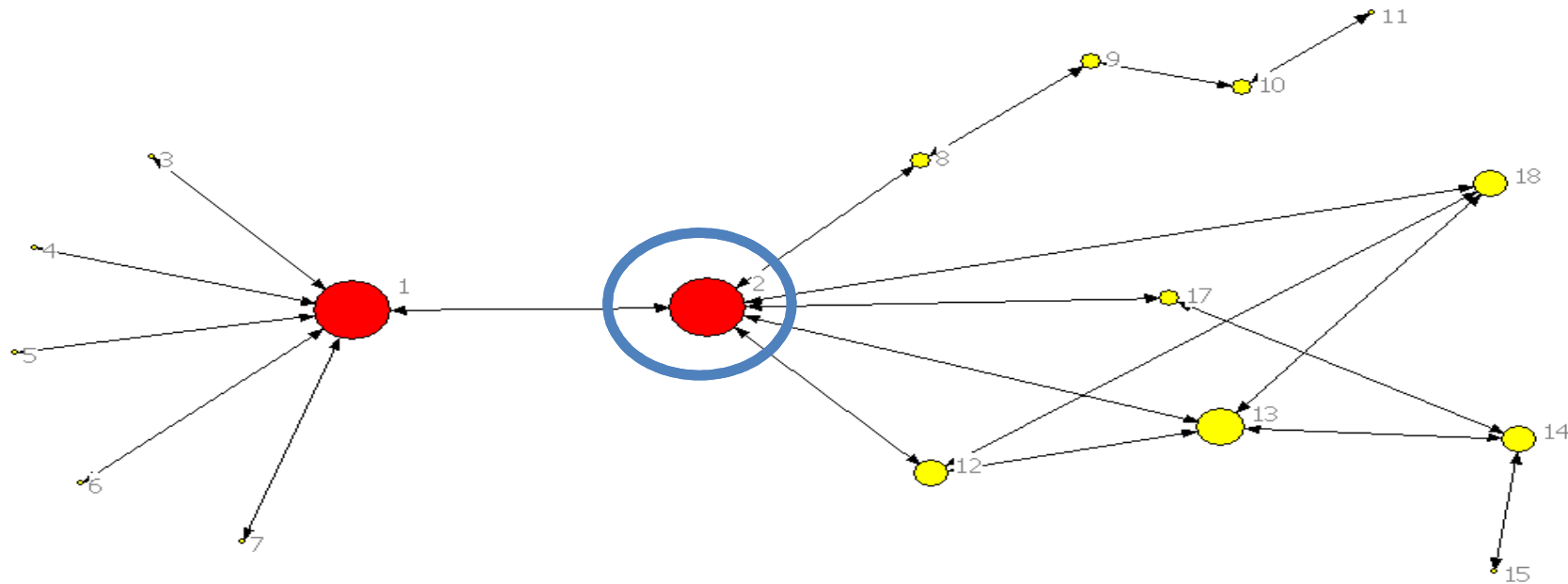


El grado de los nodos en rojo (1 y 2) es el mismo (6) pero, evidentemente, no son igual de importantes

¿Por qué? Por la debilidad de esta medida, que solo toma en cuenta los vínculos inmediatos (a un nivel local) dejando de lado los vínculos indirectos (a nivel global)

MEDIDAS LOCALES DE CENTRALIDAD

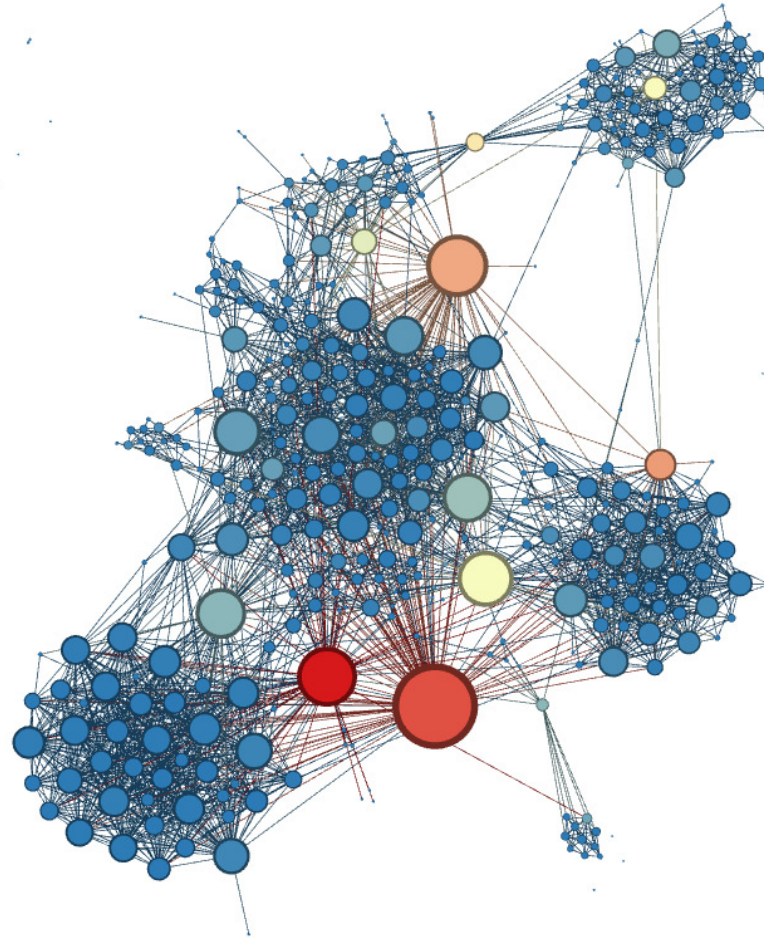
Grado vs. Intermediación (2)



Nombre	Grado (C_D) Max = 17	Intermediación (C_B) Max = $17 \cdot 16 / 2 = 136$
1	6	70.00
2	6	96.50

MEDIDAS LOCALES DE CENTRALIDAD

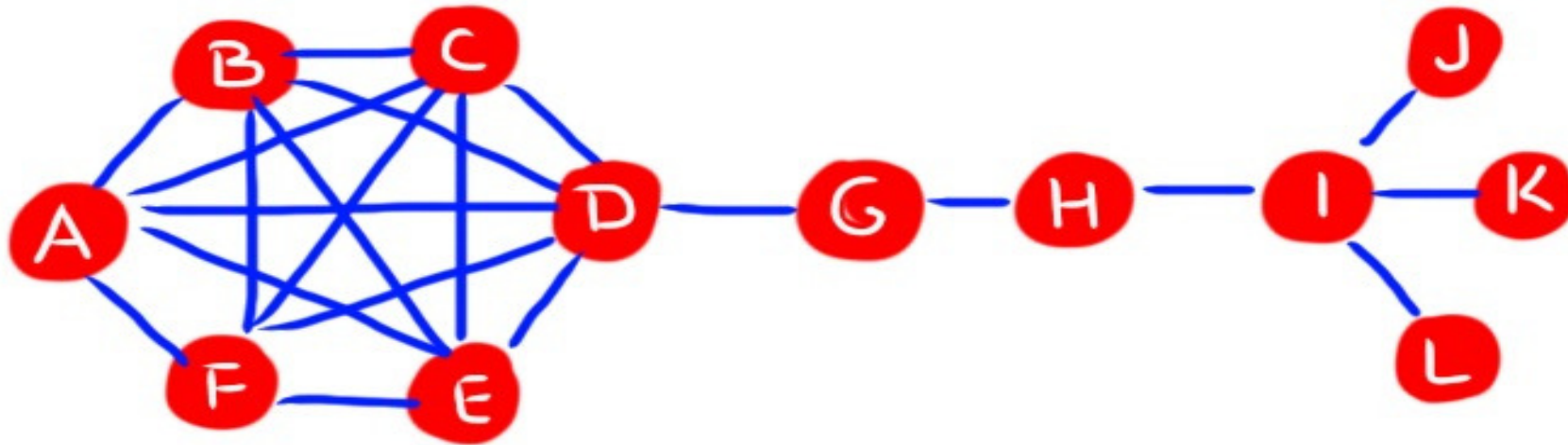
Grado vs. Intermediación (3)



Red Personal de Contactos de Facebook de Oscar Córdón: el tamaño de los nodos indica el **grado** y el color la **intermediación** (más **azul**, menor valor; más **rojo**, mayor valor)

MEDIDAS LOCALES DE CENTRALIDAD

Ejercicio



- Encontrar un actor con un valor alto de intermediación pero con un grado bajo
- Encontrar un actor con un valor bajo de intermediación pero con un grado alto

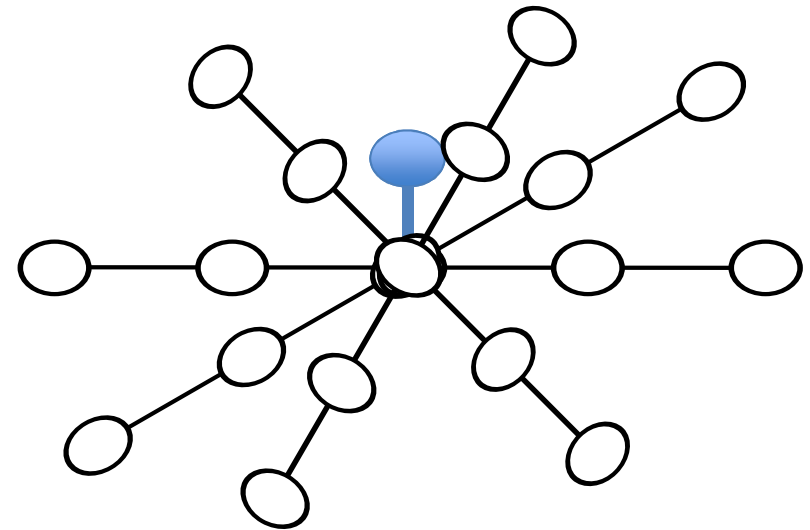
MEDIDAS LOCALES DE CENTRALIDAD

Cercanía (1)

La **cercanía** es una tercera forma alternativa de medir la centralidad que se plantea el hecho de que puede no ser tan importante tener muchos amigos directos ni estar situado “entre” otros actores

En este caso, **se le da importancia a “estar en medio de las cosas”, no demasiado lejos del centro**, para lo cual no es necesario estar en una posición de correduría

Se enfatiza la distancia de un actor a otros en la red al concentrarse en la distancia geodésica de cada actor con todos los demás



La suma de estas distancias geodésicas (distancias de los caminos mínimos) para cada actor es la lejanía de dicho actor al resto. **La inversa de dicha suma es la medida de cercanía**

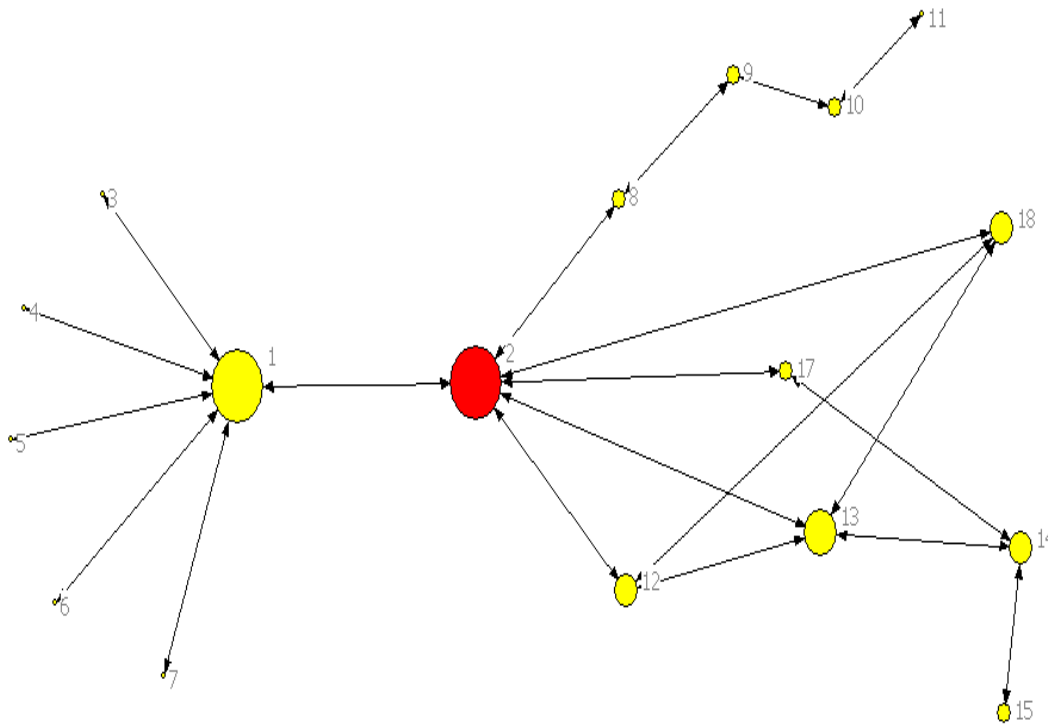
MEDIDAS LOCALES DE CENTRALIDAD

Cercanía (2)

Fórmulas de la **Centralidad de cercanía** (sin normalizar, C_c , y normalizada, C'_c):

$$C_c(i) = \frac{1}{\sum^g d(i, j)}$$

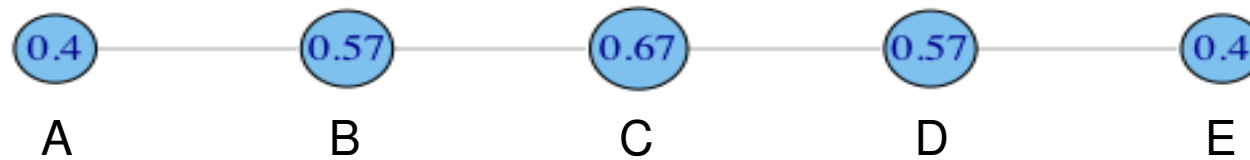
$$C'_c(i) = C_c(i) / g - 1$$



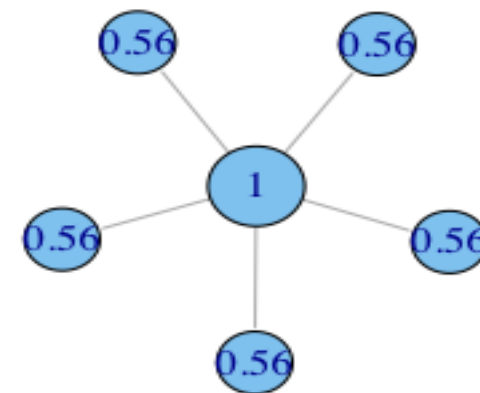
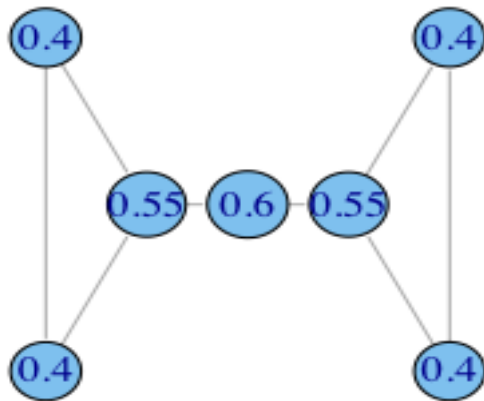
Actor	Lejanía	CercaníaN
2	34.000	50.000
1	40.000	42.500
13	42.000	40.476
17	44.000	38.636
8	44.000	38.636
12	45.000	37.778
18	45.000	37.778
14	52.000	32.692
6	56.000	30.357
5	56.000	30.357
7	56.000	30.357
3	56.000	30.357
4	56.000	30.357
9	56.000	30.357
15	66.000	25.758
10	70.000	24.286
16	82.000	20.732
11	86.000	19.767

MEDIDAS LOCALES DE CENTRALIDAD

Ejemplos de Cercanía



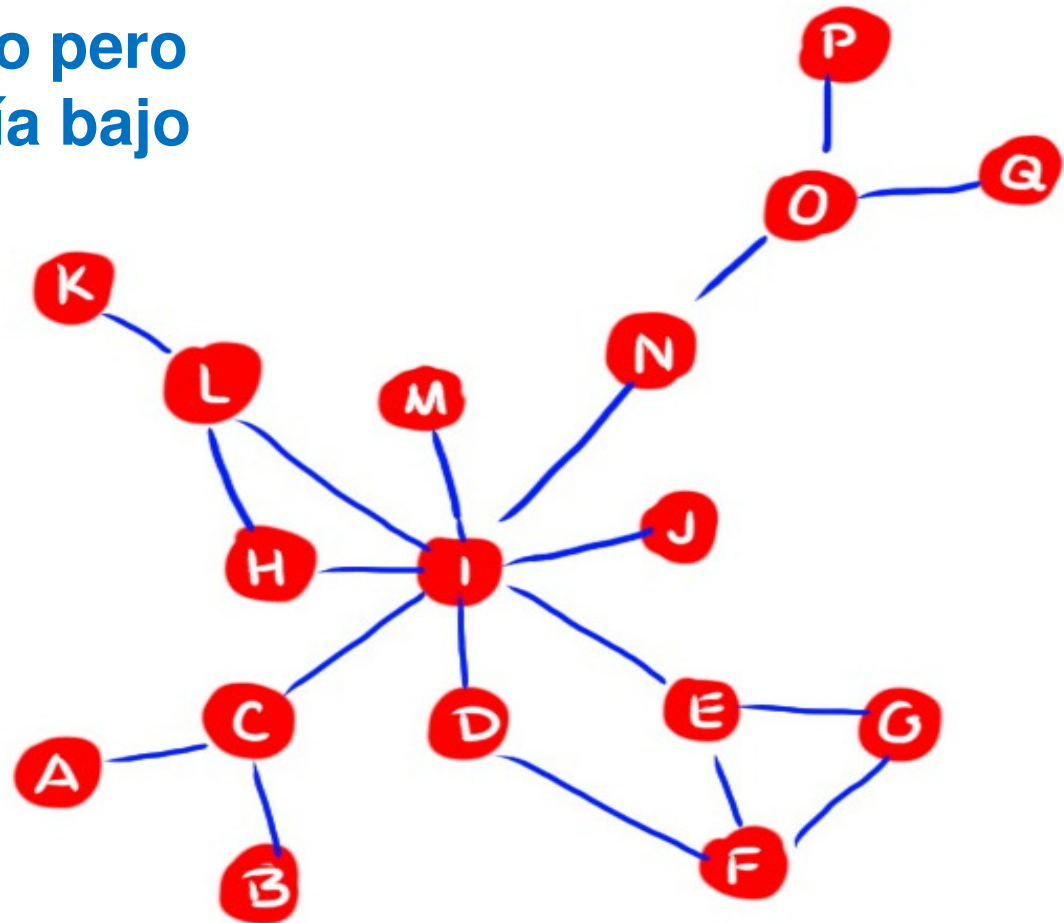
$$C'_c(A) = \left[\frac{\sum_{j=1}^g d(A, j)}{g-1} \right]^{-1} = \left[\frac{1+2+3+4}{4} \right]^{-1} = \left[\frac{10}{4} \right]^{-1} = 0.4$$



MEDIDAS LOCALES DE CENTRALIDAD

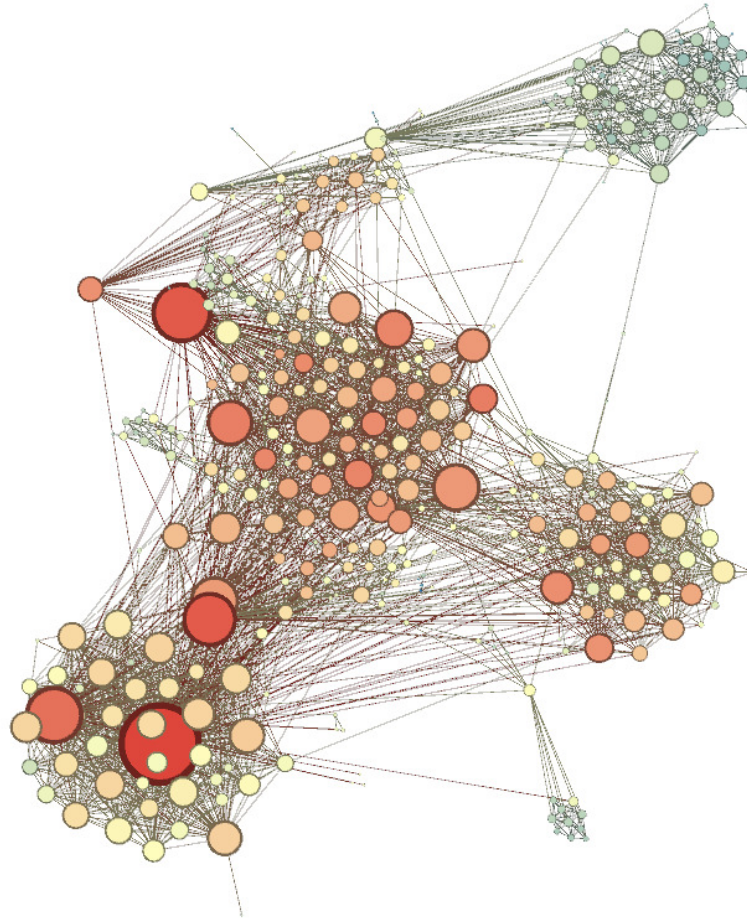
Ejercicio

Encontrar un actor con un grado relativamente alto pero con un valor de cercanía bajo



MEDIDAS LOCALES DE CENTRALIDAD

Grado vs. Cercanía



Red Personal de Contactos de Facebook de Oscar Córdón: el tamaño de los nodos indica el **grado** y el color la **cercanía** (más **azul**, menor valor; más **rojo**, mayor valor)

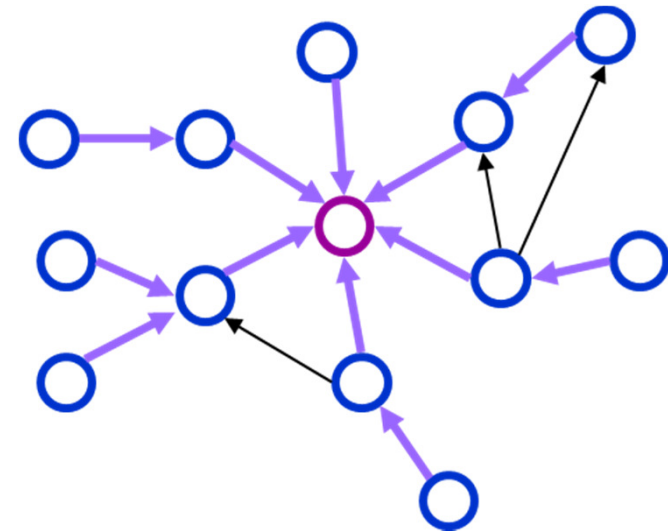
MEDIDAS LOCALES DE CENTRALIDAD

Cercanía en redes dirigidas

En redes dirigidas, se puede aprovechar la información que proporciona la dirección de los enlaces (relaciones) para definir **dos medidas de cercanía** distintas **considerando sólo los enlaces de una única dirección** en cada una de ellas:

- **Cercanía de entrada** (p.ej. prestigio en redes de citación)
- **Cercanía de salida** (alcance de la influencia de un actor)

Normalmente se consideran sólo aquellos actores desde los que el actor del nodo i puede ser alcanzado



Otra medida local de centralidad basada en distancias es la **Centralidad de excentricidad (C_E)**. Se define como la inversa de la **excentricidad** (la máxima distancia geodésica) entre un actor y cualquier otro actor de la red:

$$C_E(i) = \frac{1}{\max_{j \in V(G)/i} d(i, j)}$$

$$C'_E(i) = C_E(i) / g - 1$$

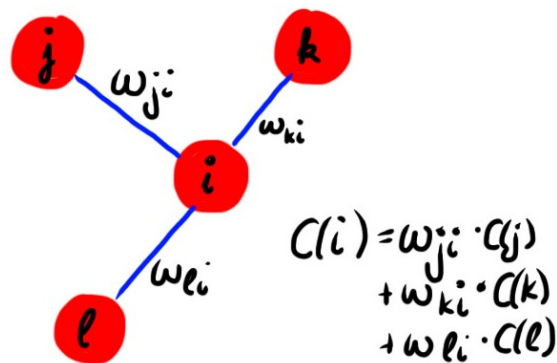
Los actores con un mayor valor de excentricidad se denominan **actores periféricos**, los de menor valor forman el **centro de la red**

MEDIDAS LOCALES DE CENTRALIDAD

Centralidad de vector propio (1)

La **Centralidad de vector propio** se basa en que la centralidad de un nodo concreto depende de cómo de centrales sean sus vecinos (**prominencia**)

La idea básica es que el poder y el status de un actor (**ego**) se define recursivamente a partir del poder y el status de sus vecinos (**alters**)



w_{ij} (a_{ij}) corresponde a la entrada de la matriz de adyacencia. Puede ser binaria $\{0,1\}$ o un peso numérico

La medida es válida para redes dirigidas (**Prestigio de rango**) y no dirigidas

Es una versión más elaborada de la Centralidad de grado al asumir que no todas las conexiones tienen la misma importancia. No se tiene en cuenta la cantidad sino la calidad de las mismas

MEDIDAS LOCALES DE CENTRALIDAD

Centralidad de vector propio (2)

La medida de Centralidad de vector propio, C_{VP} , se define como una combinación lineal (o una **suma**, si los enlaces ponderados no están ponderados) de los valores de todos los actores que apunten a i :

$$C_{VP}(i) = a_{1i} \cdot C_{VP}(1) + a_{2i} \cdot C_{VP}(2) + \dots + a_{ni} \cdot C_{VP}(n)$$

Para calcular los valores de C_{VP} para los n actores se construye un sistema de n ecuaciones con n incógnitas que se representa de forma matricial

Si $\mathbf{C} = (C_{VP}(1), \dots, C_{VP}(n))^T$ es el vector transpuesto que almacena los n valores de C_{VP} ($\mathbf{C}$ es un vector columna) y $\mathbf{A}$ es la matriz de adyacencia, entonces:

$$\mathbf{C} = \mathbf{A}^T \cdot \mathbf{C}$$

Esta ecuación coincide con la **ecuación característica para encontrar los vectores y valores propios de la matriz $\mathbf{A}^T$** . $\mathbf{C}$ es un vector propio de $\mathbf{A}^T$

MEDIDAS LOCALES DE CENTRALIDAD

Centralidad de vector propio (3)

Es habitual considerar la medida normalizada:

$$c_i = \frac{1}{\lambda} \sum_{j=1}^n a_{ij} \cdot c_j \quad \mathbf{C} = \frac{1}{\lambda} \mathbf{A}^T \cdot \mathbf{C} \quad \lambda \cdot \mathbf{C} = \mathbf{A}^T \cdot \mathbf{C}$$

donde λ es una constante que equivale al mayor valor absoluto del vector propio dominante de $\mathbf{A}$

Existen distintos algoritmos para calcular ese vector propio $\mathbf{C}$, como el **Método de las Potencias**, usado por Google en el **Pagerank** (redes dirigidas)

http://es.wikipedia.org/wiki/Método_de_las_potencias

Como la fórmula es circular, el Método de las Potencias es iterativo. Sin embargo, se deben cumplir unas condiciones específicas para poder aplicarlo

MEDIDAS LOCALES DE CENTRALIDAD

Centralidad de vector propio (4)

P. Bonacich. Power and centrality: a family of measures. American Journal of Sociology 1987, 92(5): 1170-1182

Una de las medidas de Centralidad de vector propio más extendida es la **Centralidad de Bonacich**:

$$c_i(\beta) = \sum_j (\alpha + \beta c_j) a_{ji} \quad \mathbf{C}(\beta) = \alpha (\mathbf{I} - \beta \mathbf{A})^{-1} \mathbf{A} \mathbf{1}$$

- α es una constante de normalización
- β determina la importancia de la centralidad de los vecinos de i para calcular la centralidad de dicho nodo
- $\mathbf{A}$ es la matriz de adyacencia (binaria o ponderada)
- $\mathbf{I}$ es la matriz identidad
- $\mathbf{1}$ es una matriz con todas las componentes a 1

El parámetro β es un **factor de atenuación** que determina qué actores de la red influyen en el cálculo de la centralidad del nodo i :

- Si β es pequeño $\rightarrow$ atenuación alta: sólo los amigos cercanos influyen y su importancia es muy puntual
- Si β es grande $\rightarrow$ atenuación baja: la estructura global de la red tiene importancia (tus amigos, los amigos de tus amigos, etc.)
- Si $\beta=0 \rightarrow$ la fórmula coincide directamente con la Centralidad de grado:

$$c_i(\beta) = \sum_j (\alpha \quad) a_{ji}$$

MEDIDAS LOCALES DE CENTRALIDAD

Centralidad de vector propio (6)

Además, el **signo** de β determina el comportamiento de la medida:

- Si $\beta > 0 \rightarrow$ los actores tienen una mayor centralidad cuando están conectados a otros actores centrales
- Si $\beta < 0 \rightarrow$ los actores tienen una mayor centralidad cuando están conectados a actores poco centrales

$$\beta = .25$$



$$\beta = -.25$$



MEDIDAS LOCALES DE CENTRALIDAD

Coeficiente Local de Clustering

Transitividad: *Las redes sociales son transitivas por naturaleza, es decir, los amigos de un actor dado también suelen ser amigos entre sí*

La propiedad de transitividad de una red se cuantifica por el **Coeficiente de Clustering** que puede ser global, a nivel de toda la red, o **local**, a nivel de cada nodo individual:

$$C_{CLC}(i) = \frac{2L_i}{k_i(k_i - 1)}$$

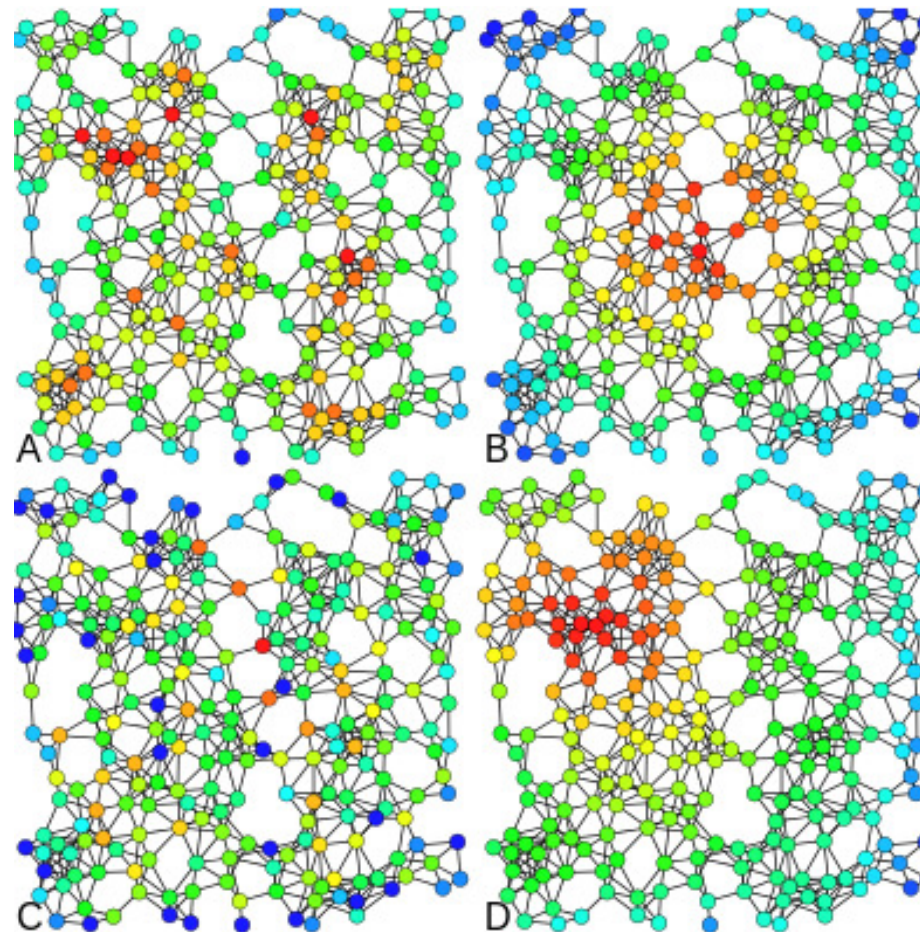
- Nodo i con grado k_i
- L_i = número de enlaces entre los vecinos del nodo i
- $C_i \in [0,1]$. $C_i = 0$ indica que ninguno de los vecinos de i están conectados entre sí, $C_i = 1$ indica que todos están conectados
- A mayor valor, más central se considera el nodo

MEDIDAS LOCALES DE CENTRALIDAD

Comparativa

A) centralidad de grado; B) cercanía; C) intermediación; D) centralidad de vector propio

azul = menor valor



rojo = mayor valor

MEDIDAS GLOBALES PARA EL ANÁLISIS DE REDES SOCIALES

MEDIDAS GLOBALES

Existen varias medidas globales en SNA. La mayoría son las mismas empleadas para analizar cualquier otro tipo de red, que ya hemos estudiado:

1. **Diámetro y Radio**
2. **Distancia media**
3. **Grado Medio**
4. **Densidad**
5. **Coeficiente de Clustering Global**
6. **Reciprocidad**

Otras, denominadas **medidas de Centralización**, son **agregaciones a nivel de grupo de actores** de las medidas locales de Centralidad que permiten comparar redes

MEDIDAS GLOBALES

Diámetro y Radio

Diámetro ($d_{\max}$): longitud del camino mínimo más largo de la red

En redes grandes, se puede determinar con el algoritmo de búsqueda primero en anchura

Equivale al valor máximo de excentricidad para todos los nodos de la red:

$$E(i) = \max_{j \in V(G)/i} d(i, j) \quad d_{\max} = \max \{E(i) : i \in V(G)\}$$

En el contexto del SNA, esta métrica da una **idea de la proximidad entre pares de actores en la red**, indicando cómo de lejos están en el peor de los casos

Las redes más dispersas suelen tener un mayor diámetro que las más densas al existir menos caminos entre cada par de nodos

Radio (r): Valor mínimo de excentricidad para toda la red:

$$r = \min \{E(i) : i \in V(G)\}$$

MEDIDAS GLOBALES

Distancia media

Distancia media ($\langle d \rangle$) para un grafo dirigido:

$$\langle d \rangle \equiv \frac{1}{2L_{\max}} \sum_{i, j \neq i} d_{ij} \quad d_{ij} \text{ es la distancia geodésica entre los nodos } i \text{ y } j$$

En un grafo no dirigido $d_{ij}=d_{ji}$. De este modo, sólo es necesario contar la longitud de los caminos una vez:

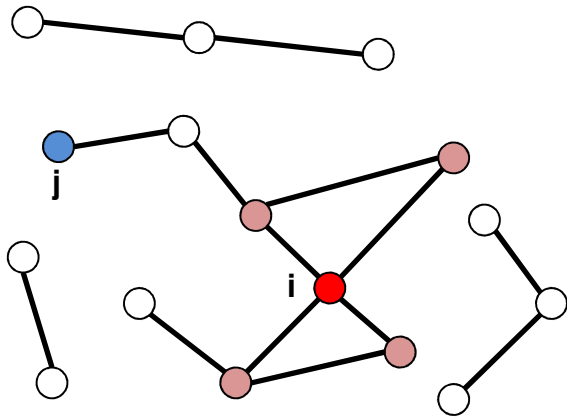
$$\langle d \rangle \equiv \frac{1}{L_{\max}} \sum_{i, j > i} d_{ij}$$

La medida da una idea de cómo de lejos están los distintos actores en promedio. En SNA representa la **eficiencia del flujo de información en la red**

MEDIDAS GLOBALES

Grado medio

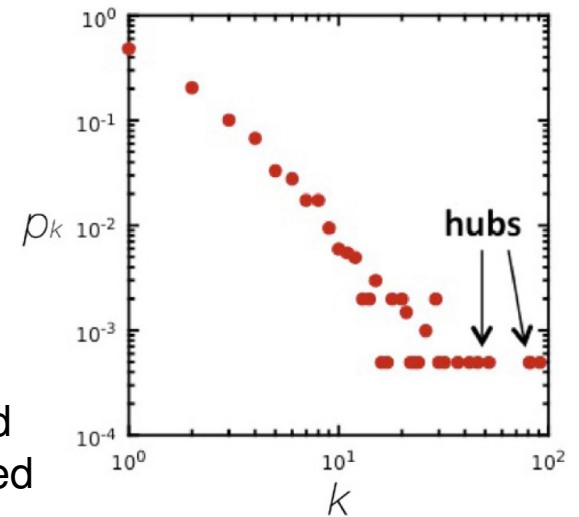
No dirigida



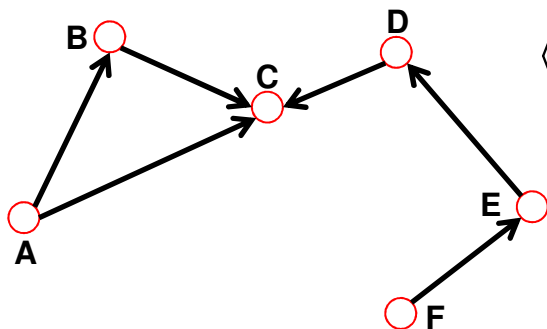
$$\langle k \rangle \equiv \frac{1}{N} \sum_{i=1}^N k_i$$

$$\langle k \rangle \equiv \frac{2L}{N}$$

N – número de nodos de la red
L – número de enlaces de la red



Dirigida



$$\langle k^{in} \rangle \equiv \frac{1}{N} \sum_{i=1}^N k_i^{in}, \quad \langle k^{out} \rangle \equiv \frac{1}{N} \sum_{i=1}^N k_i^{out}, \quad \langle k^{in} \rangle = \langle k^{out} \rangle$$

$$\langle k \rangle \equiv \frac{L}{N}$$

MEDIDAS GLOBALES

Densidad y Coeficiente de Clustering

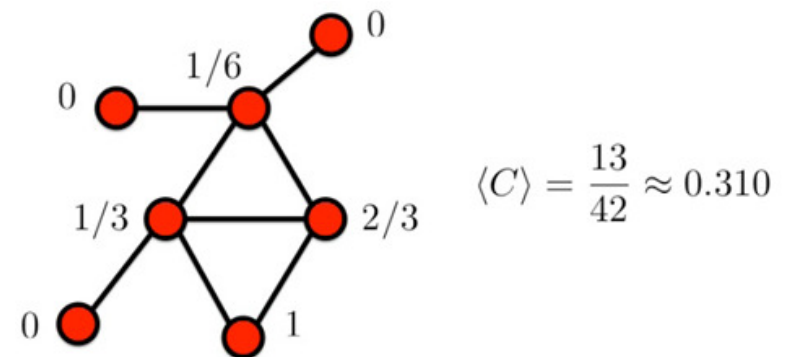
La **densidad** $D = L/L_{max}$ mide el grado de conectividad de la red social a nivel global

El coeficiente de clustering C_i mide la **densidad local**, el ratio de vecinos de un nodo que están conectados entre sí

El **coeficiente de clustering medio** ($\langle C \rangle$) es una **medida global** que indica la probabilidad de que dos vecinos de un nodo de la red escogido aleatoriamente estén conectados entre sí:

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

$$\langle C \rangle = \frac{1}{N} \sum_{i=1}^N C_i$$



Las redes sociales reales son redes de mundos pequeños (*small worlds*) y suelen tener valores de $\langle C \rangle$ altos. Eso implica que la **transitividad** entre nodos aparece más y con más fuerza, incrementando la probabilidad de que se formen **cliques**

MEDIDAS GLOBALES

Reciprocidad

La **reciprocidad (R)** en un **grafo dirigido** mide la tendencia de pares de actores a tener conexiones mutuas entre ellos

La forma más habitual de calcularla es como un ratio entre el número de conexiones (*díadas*) mutuas (*#mut*) y el número de conexiones totales en la red, las mutuas y las asimétricas (*#asim*):

$$R = \frac{\#mut}{\#mut + \#asim}, \quad R \in [0,1]$$

Indica la probabilidad de que dos actores de la red apunten el uno al otro

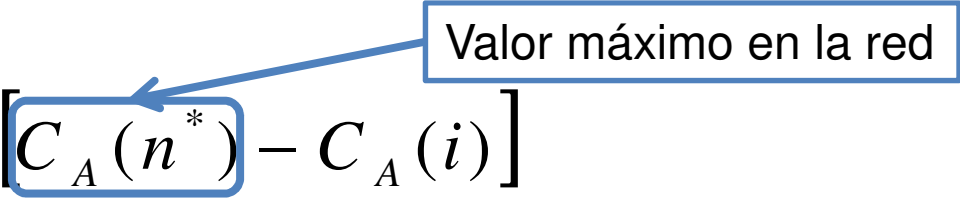
Una *díada asimétrica* es un par de actores que presentan un arco en una u otra dirección pero no en ambas. Una *díada mutua* incluye los arcos en las dos direcciones

MEDIDAS GLOBALES

Medidas de Centralización (1)

Las **medidas de Centralización** indican la variabilidad de los valores de centralidad entre los actores del grupo. Miden el grado en que un actor tiene una centralidad alta mientras que el resto la tiene baja

Existen distintas formas de construir una medida de este tipo, como la formulación de Freeman (1979):

$$C_A = \frac{\sum_{i=1}^g [C_A(n^*) - C_A(i)]}{\max \sum_{i=1}^g [C_A(n^*) - C_A(i)]}$$


Valor máximo en la red

- C_A es una medida de centralidad y $C_A(i)$ su valor para el actor i
- $C_A(n^*)$ es el valor máximo de esa medida para los g actores de la red ($C_A(n^*) = \max_j C_A(j)$)
- $\max \sum_{i=1}^g [C_A(n^*) - C_A(i)]$ es el máximo **teórico** del numerador

MEDIDAS GLOBALES

Medidas de Centralización (2)

El índice C_A se define en $[0,1]$. Toma valor 0 cuando todos los actores tienen exactamente el mismo valor de Centralidad ($C_A(n^*)$) y 1 cuando un actor domina totalmente (**eclipsa**) al resto

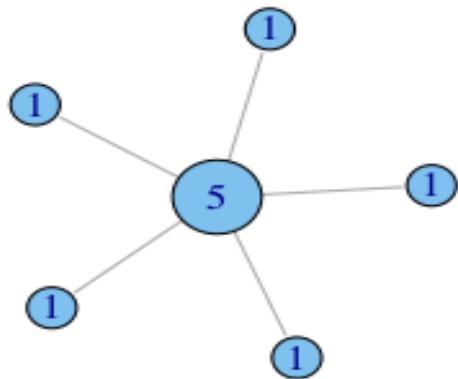
Un ejemplo sería la **medida de Centralización de grado C_D** donde el máximo teórico es $(g-1) \cdot (g-2)$:

$$C_D = \frac{\sum_{i=1}^g [C_D(n^*) - C_D(i)]}{(g-1) \cdot (g-2)}$$

C_D toma valor máximo 1 cuando un solo actor está conectado al resto que no tienen conexiones entre sí (grafo en estrella). El valor mínimo 0 corresponde a un grafo regular donde todos los nodos tienen el mismo grado

MEDIDAS GLOBALES

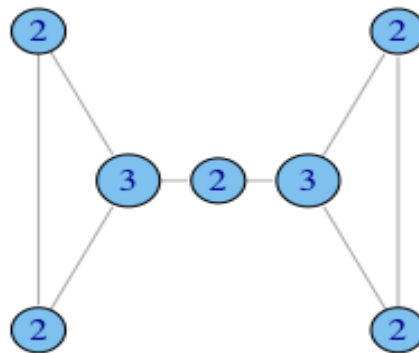
Ejemplos de Centralización de grado



$$C_D = 1.0$$



$$C_D = 0.167$$



$$C_D = 0.167$$

La **medida de Centralización de intermediación** C_B sería:

$$C_B = \frac{\sum_{i=1}^g [C'_B(n^*) - C'_B(i)]}{(g - 1)}$$

Se calcula a partir de la intermediación normalizada C'_B . Toma el valor máximo 1 en un grafo en estrella y el valor mínimo 0 en un grafo en el que todos los valores de intermediación sean iguales

La **medida de Centralización de cercanía C_c** sería:

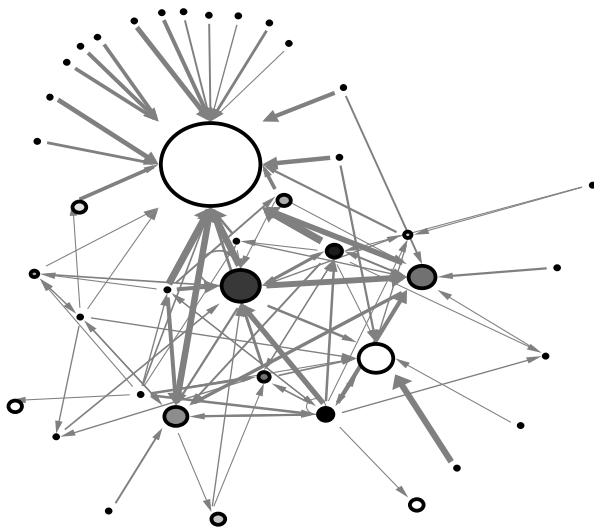
$$C_c = \frac{\sum_{i=1}^g [C_c(n^*) - C_c(i)]}{[(g-1) \cdot (g-2)] / (2g-3)}$$

C_c toma valor 1 en un grafo en estrella y el valor mínimo 0 en un grafo en el que todas las distancias sean iguales (p.ej. un grafo completo o circular)

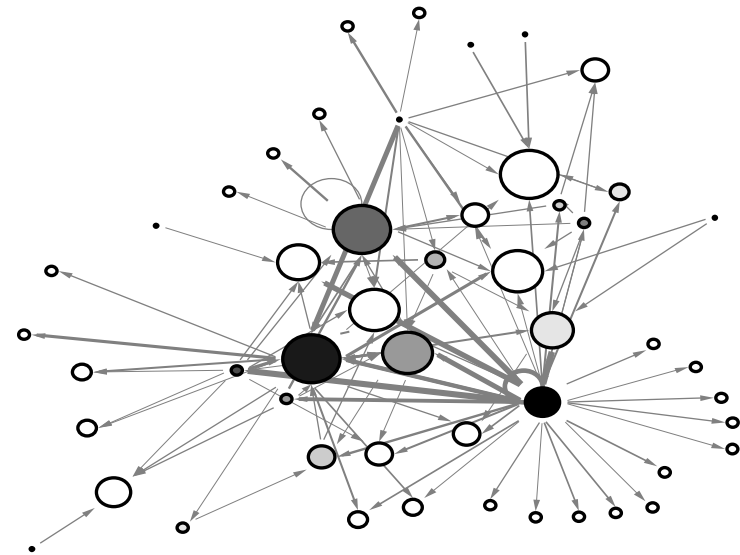
El problema de estas medidas es el máximo teórico, que no es calculable para algunas **Centralidades de prestigio** en redes dirigidas (aunque si para la Centralidad de grado)

Alternativamente, se puede usar **la varianza de los valores de Centralidad**

Ejemplo de redes de mercados financieros



Alta Centralización de soporte (in-degree): un país comprando mucho de muchos otros



Baja Centralización de soporte (in-degree): la compra está distribuida más uniformemente

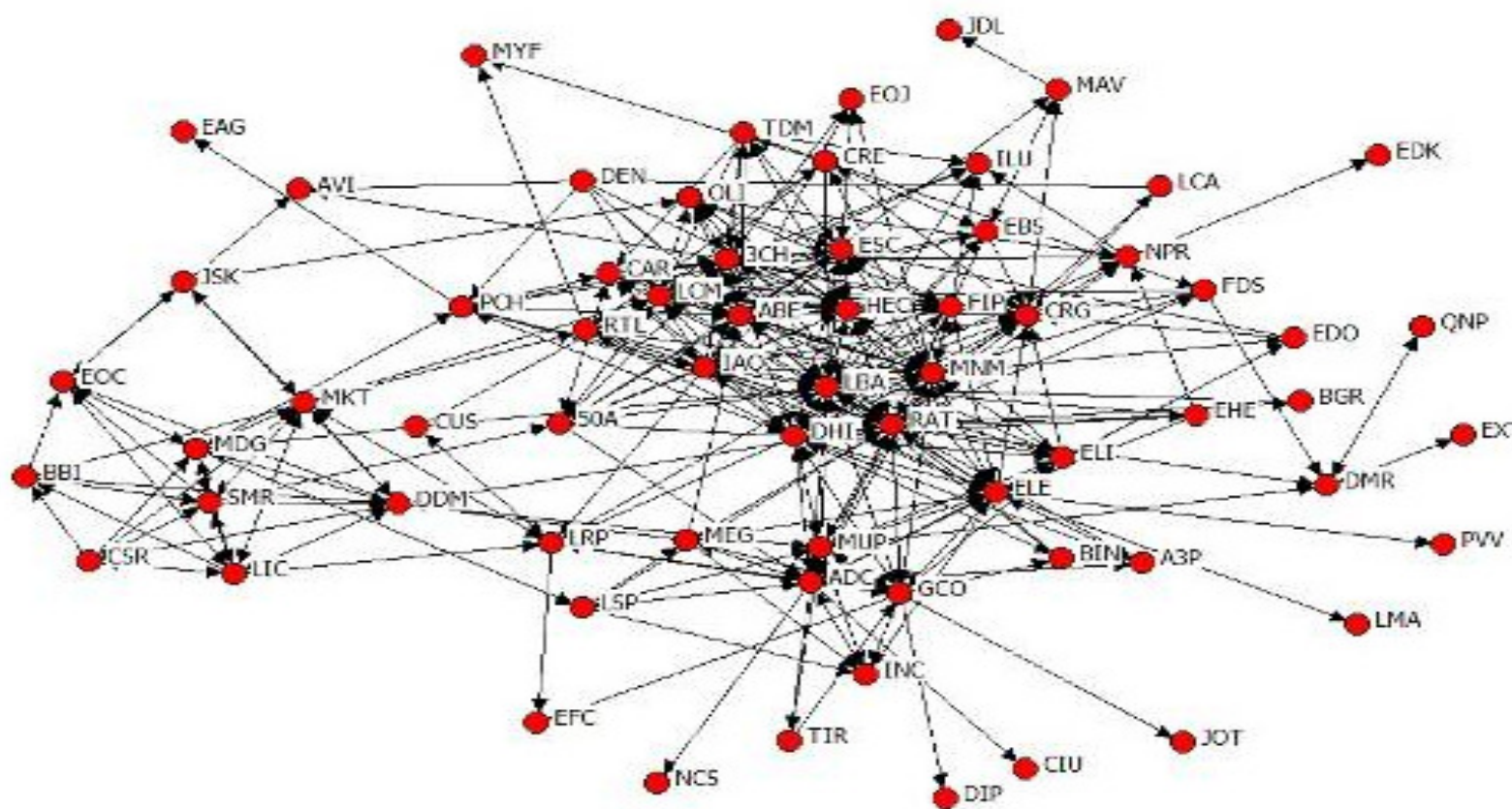
ALGUNAS APLICACIONES DEL ANÁLISIS DE REDES SOCIALES

S. Guerrero. Aplicación del análisis de redes sociales al estudio de la estructura y desarrollo de una red de blogs económicos políticos en Argentina. Redes 2008, 14(7)

Estudio de las relaciones entre 67 blogs económico-políticos en Argentina para determinar su prominencia

Los enlaces entre blogs vienen determinados por hiperenlaces en los blogrolls

- 67 nodos, 440 enlaces, $D=0.0995$, una sola componente débilmente conexa
- $\langle k \rangle = 6.57$, $\langle k^{\text{in}} \rangle = 6.13$, $\langle k^{\text{out}} \rangle = 5.88$



Red con los 67 blogs seleccionados representados como nodos y sus hiperenlaces en sus respectivos blogrolls presentes en 2007 como enlaces

APLICACIONES

Análisis de blogs económico-políticos en Argentina (3)

	Degree	Closeness	Betweenness
-----	-----	-----	-----
1 LCM	30.303	51.563	4.711
2 RAT	43.939	63.462	19.757
3 HEC	22.727	49.254	0.868
4 OLI	15.152	46.154	1.661
5 3CH	30.303	51.563	4.211
6 ABE	27.273	50.769	3.934
7 CRG	27.273	50.382	3.929
8 LBA	42.424	58.929	10.479
9 AVI	4.545	36.872	0.571
10 CRE	10.606	42.308	0.940
11 ILU	9.091	41.772	0.141
12 50A	16.667	48.175	2.212
13 MUP	19.697	51.969	4.766
14 CAR	18.182	46.809	0.311
15 ELE	30.303	52.800	8.796
16 MNM	36.364	55.932	6.101
17 DDM	16.667	45.833	6.610
18 GCO	24.242	50.000	9.219
19 NPR	12.121	44.898	3.256
20 ADC	25.758	51.163	8.398
21 ELI	18.182	48.889	0.774
22 DHI	36.364	56.897	6.279

23 DMR	7.576	41.509	6.042
24 BGR	3.030	41.250	0.000
25 LCA	4.545	41.250	0.054
26 PCH	15.152	47.482	4.355
27 FDS	9.091	44.295	0.824
28 TDM	13.636	40.994	0.211
29 ESC	25.758	47.482	1.468
30 EOJ	6.061	39.053	0.000
31 FIP	22.727	48.889	0.605
32 MAV	6.061	36.872	3.046
33 IAQ	27.273	51.563	2.113
34 LRP	13.636	46.809	3.619
35 RTL	13.636	48.175	3.123
36 MYF	3.030	33.333	0.000
37 EBS	12.121	45.205	1.093
38 SMR	12.121	38.824	0.863
39 TIR	4.545	36.667	0.000
40 INC	9.091	40.741	0.095
41 PVV	1.515	34.737	0.000
42 EDO	7.576	40.491	0.026
43 A3P	6.061	40.994	0.000
44 MEG	12.121	48.889	1.356
45 LMA	1.515	34.737	0.000
46 EAG	1.515	32.353	0.000
47 MDG	13.636	39.521	1.881
48 EOC	10.606	35.294	0.663
49 JSK	6.061	36.264	0.354
50 MKT	15.152	43.421	3.435
51 CSR	9.091	34.921	0.031
52 BBI	9.091	34.375	0.053
53 LIC	12.121	37.931	0.676
54 DIP	1.515	33.503	0.000
55 JOT	1.515	33.503	0.000
56 EFC	3.030	35.676	0.000
57 BIN	6.061	43.421	0.000
58 EDK	1.515	31.132	0.000
59 LSP	9.091	44.000	0.752
60 CIU	1.515	34.021	0.000
61 NCS	1.515	34.021	0.000
62 EHE	7.576	43.709	0.092
63 EXT	1.515	29.464	0.000
64 QNP	1.515	29.464	0.000
65 DEN	10.606	43.709	0.826
66 JDL	1.515	27.049	0.000
67 CUS	6.061	42.581	0.343

Se puede ver que, mas allá de las covariaciones entre éstas medidas, en algunos casos el grado (degree) y la intermediación (betweenness) siguen tendencias distintas. Es posible que haya blogs que cumplan funciones de broker (intermediarios) entre diferentes sectores de la red, pese a que no se destaquen por la cantidad de enlaces que envían o reciben. Los casos mas notables son los siguientes que están bastante mejor situados en el ranking según los valores de intermediación que en el ranking según grado: GCO 3 y 13, DDM 6 y 20 y DMR 9 y 42.

U.H. Karkada, L.A. Adamic LA, J.M. Kahn, T.J. Iwashyna. Limiting the spread of highly resistant hospital-acquired microorganisms via critical care transfers: a simulation study. Intensive Care Med. 2011 37(10): 1633-40

PURPOSE:

Hospital-acquired infections with highly resistant organisms are an important problem among critically ill patients. Control of these organisms has largely focused within individual hospitals. **We examine the extent to which transfers of critically ill patients could be a vector for the wide spread of highly resistant organisms, and compare the efficiency of different approaches to targeting infection control resources**

METHODS:

We analyzed the **network of interhospital transfers of intensive care unit patients** in 2005 US Medicare data and 2004-2006 Pennsylvania all-payer data. **We simulated the spread of highly resistant hospital-acquired infections by randomly choosing a single hospital to develop a highly resistant organism and following the spread of infection or colonization throughout the network under varying strategies of infection control and varying levels of infectivity**

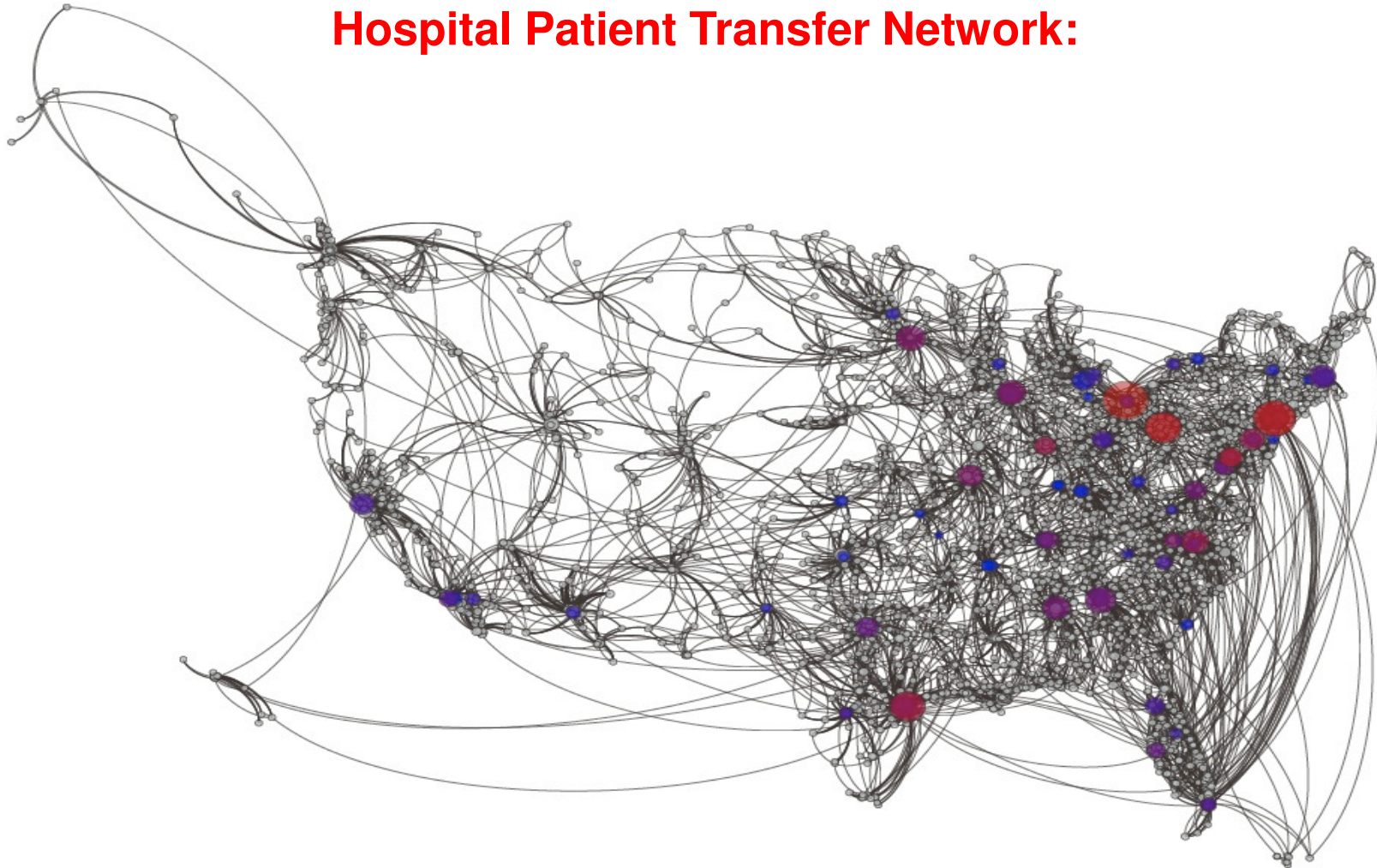
RESULTS:

Critical care transfers could spread a highly resistant organism between any two US hospitals in a median of 3 years. Hospitals varied substantially in their importance to limiting potential spread. **Targeting resources to a small subset of hospitals on the basis of their position in the transfer network was 16 times more efficient than distributing infection control resources uniformly.** Within any set of targeted hospitals, the best strategy for infection control heavily concentrated resources at a few particularly important hospitals, regardless of level of infectivity

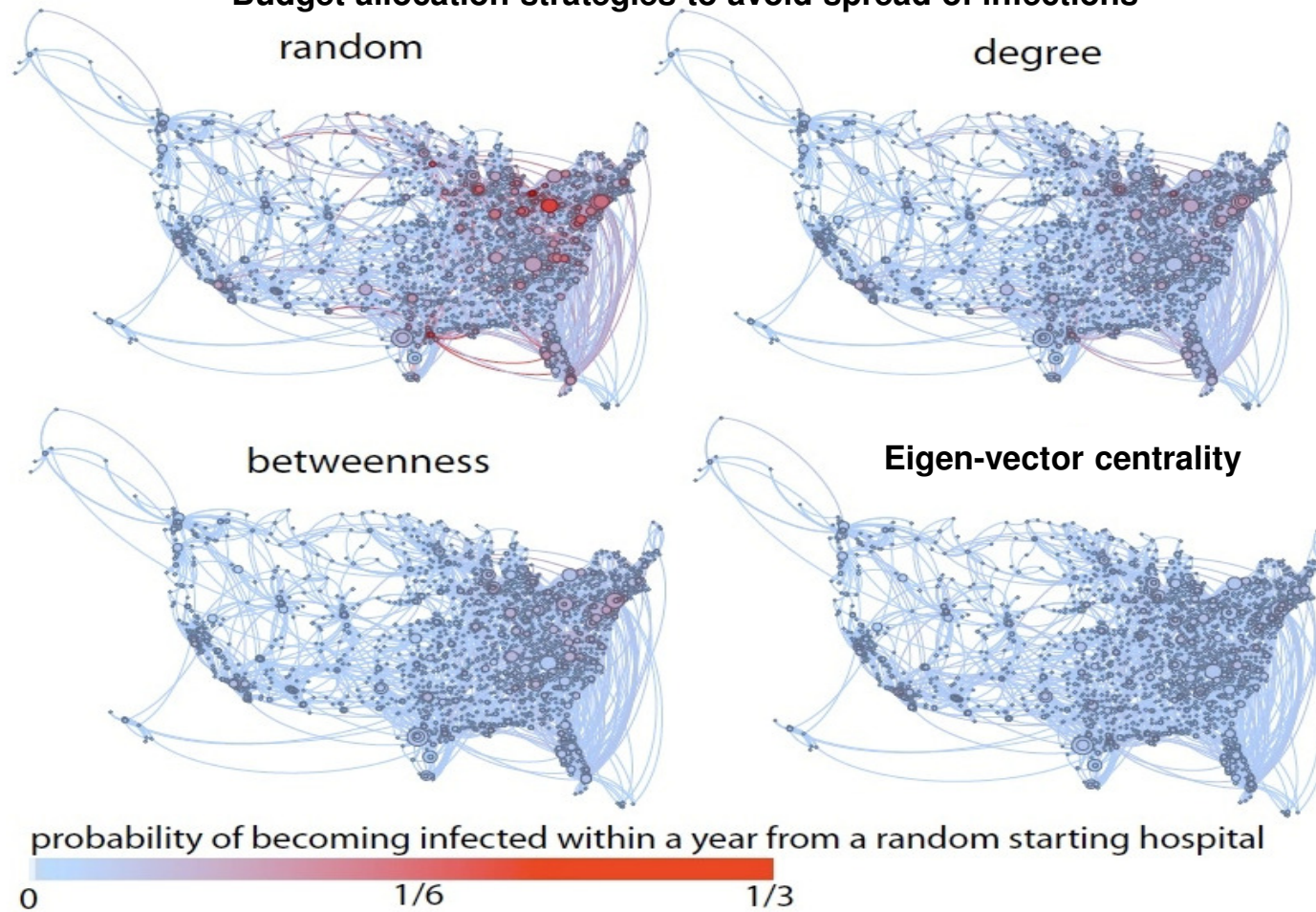
CONCLUSIONS:

Critical care transfers provide a plausible vector for widespread dissemination of highly resistant hospital-acquired microorganisms. Infection control efforts can be made more efficient by selectively targeting hospitals most important for transmission

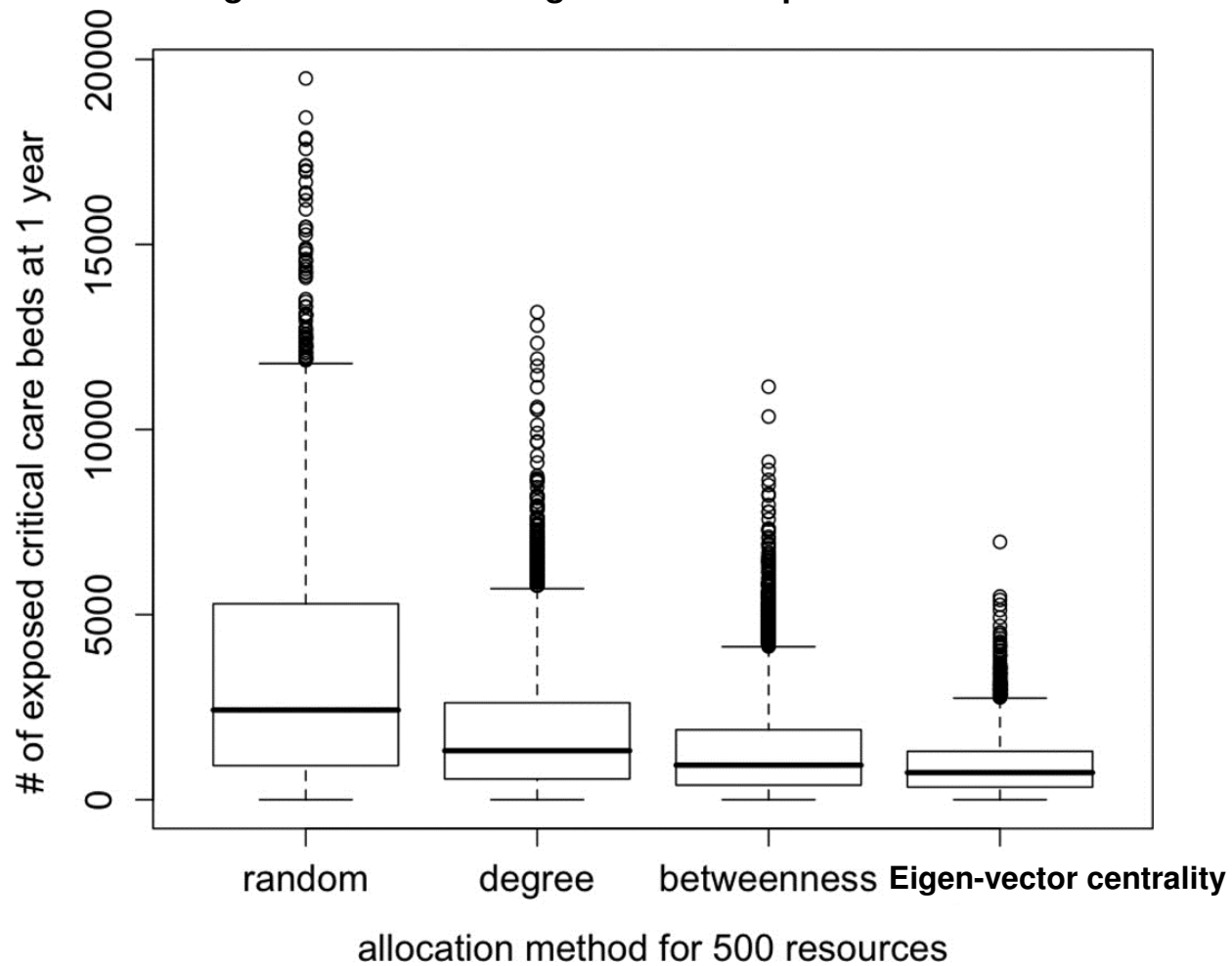
Hospital Patient Transfer Network:



Infection prevention strategies in a hospital patient transfer network
Budget allocation strategies to avoid spread of infections



Budget allocation strategies to avoid spread of infections

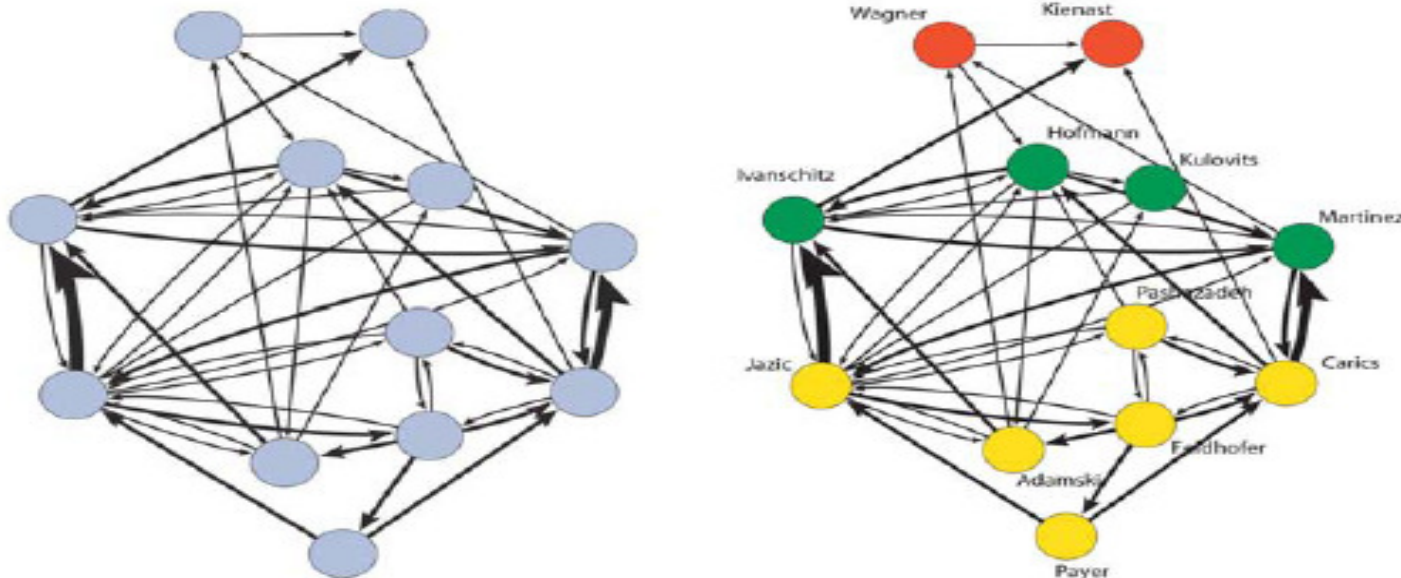


APLICACIONES

Análisis de juego en equipos de fútbol (1)

Let's look at an example on the soccer field.

The Rapid Vienna network consists of 12 persons (11 players on the field and one substitute), and we observed who passed the ball to whom during the course of a match. The resulting graph consists of a quantity of players (nodes) and a quantity of passes (arcs). To the left is the graph that depicts Rapid's passing game during the last 15 minutes of a soccer match between Rapid Vienna and Sturm Graz on December 7, 2003. As soon as we add additional information such as the players' names and their positions (red = attack, green = midfield, yellow = defense), we have produced a network. Networks are graphs with additional information about nodes and / or arcs.



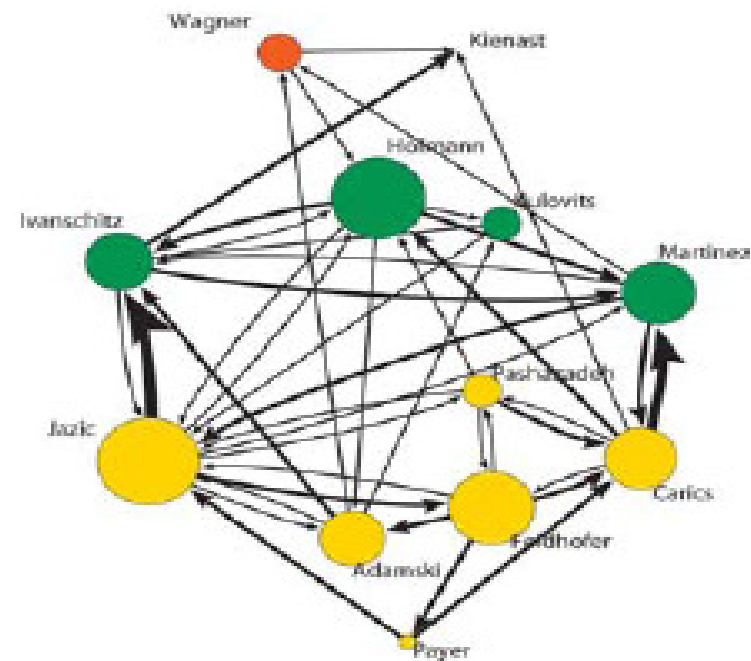
Rapid's passing game during the last 15 minutes of a soccer match between Rapid Vienna and Sturm Graz on December 7, 2003 (data by Harald Katzmaier and Helmut Neundlinger). Left: graph, Right: network

APLICACIONES

Análisis de juego en equipos de fútbol (2)

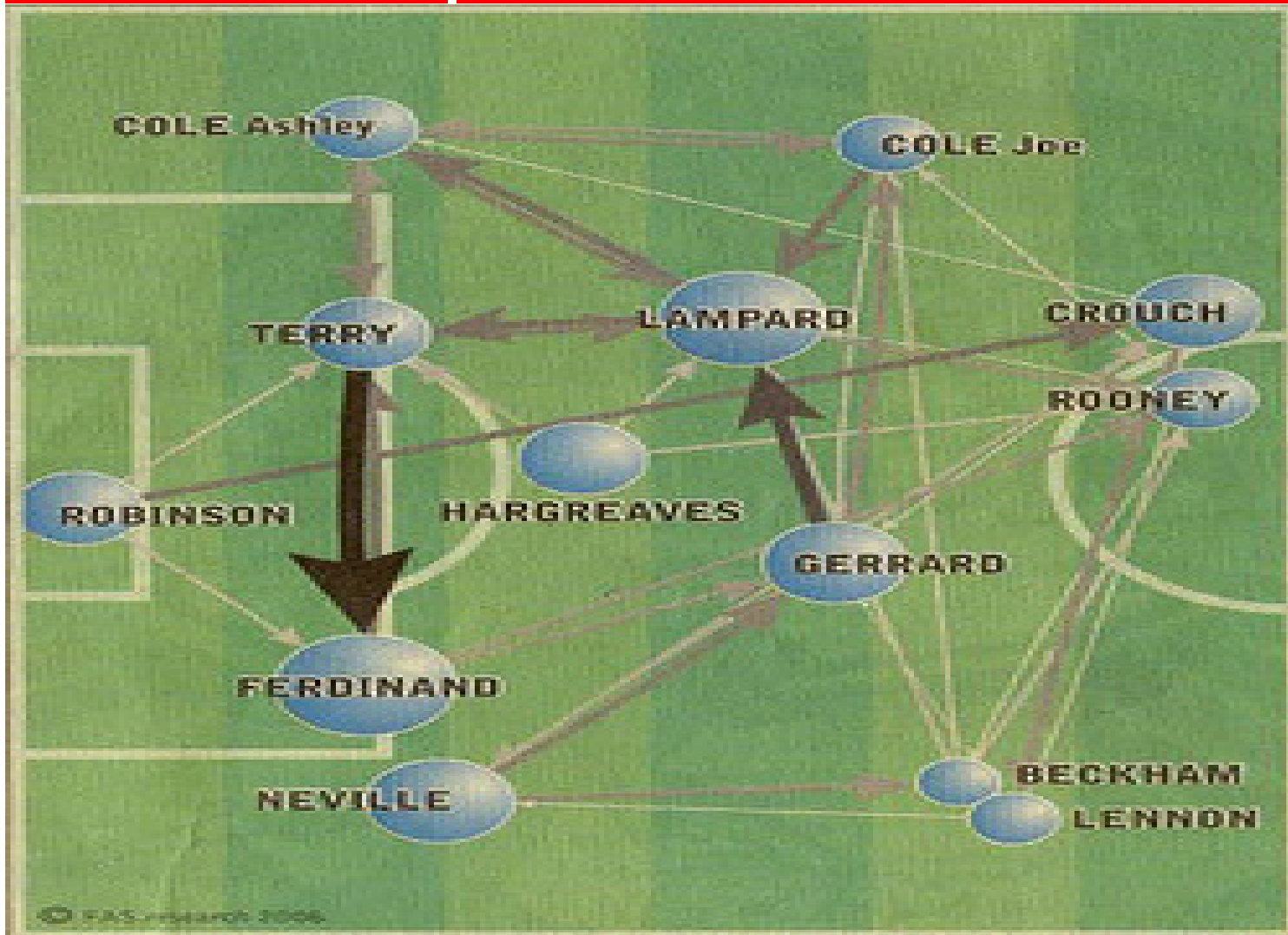
Posibles factores de análisis:

- ¿Qué jugador ha iniciado más pases (grado ponderado de salida)? **Jazic**
- ¿Qué jugador ha recibido más pases (grado ponderado de entrada)? **Jazic**
- ¿Quién ha controlado el juego del Rapid (centralidad)? **Jazic** y **Hoffman**
- ¿Qué jugadores han estado implicados en jugadas con el mayor número de pases (camino)? **Jazic**, **Hofmann**, **Feldhofer**, **Martinez** y **Carics**
- ¿Quién ha jugado con quién y quién no (análisis de los enlaces)? **Ni un solo pase de Ivanschitz a Wagner**
- ¿Qué grupos de jugadores han compuesto la columna vertebral del equipo (análisis de triadas)? **Por ejemplo, Feldhofer-Carics-Pashazadeh**
- ¿Qué jugadores han tenido un rol similar (análisis de enlaces)? **Por ejemplo, Ivanschitz / Martinez**



APLICACIONES

Análisis de juego en equipos de fútbol (3)



Chelsea FC:

Robinson	1
Cole Ashley	2
Terry	3
Ferdinand	4
Neville	5
Cole Joe	6
Lampard	7
Heargraves	8
Gerrard	9
Beckham	10
Lennon	11
Crouch	12
Rooney	13

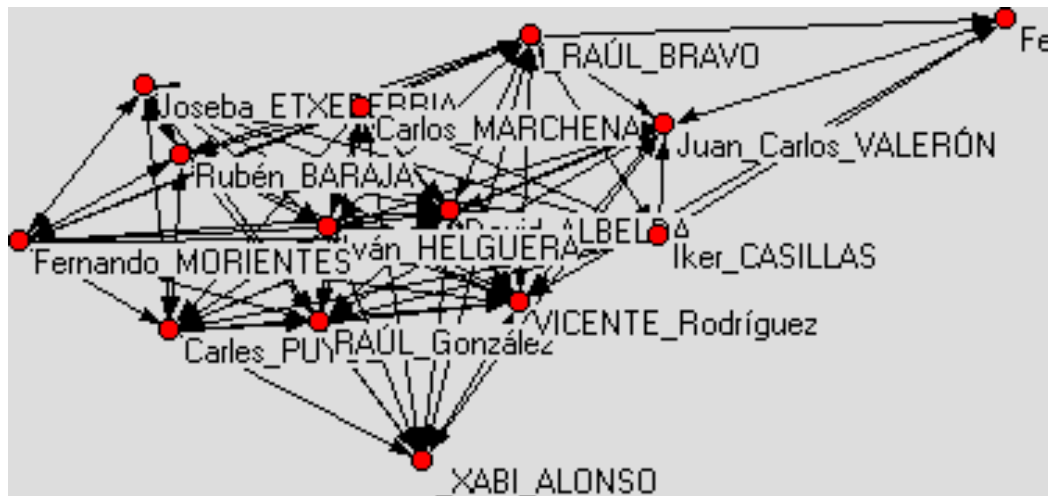
APLICACIONES

Análisis de juego en equipos de fútbol (4)

J.J. Merelo. Redes contra redes: el fútbol es así. <http://atalaya.blogalia.com/historias/19642>

En la lista Redes hemos propuesto analizar las redes que se forman en los partidos de la Eurocopa de Portugal 2004. Con un programilla me bajo las páginas de estadísticas y les paso un programilla en Perl que extrae las estadísticas de pases

Los nodos son futbolistas y los enlaces, pases. Vamos a ver cómo fue la red del España 1 – Rusia 0



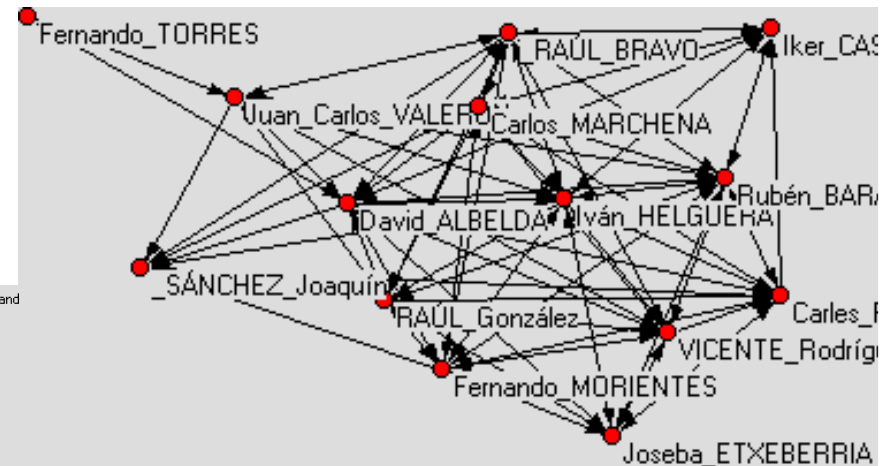
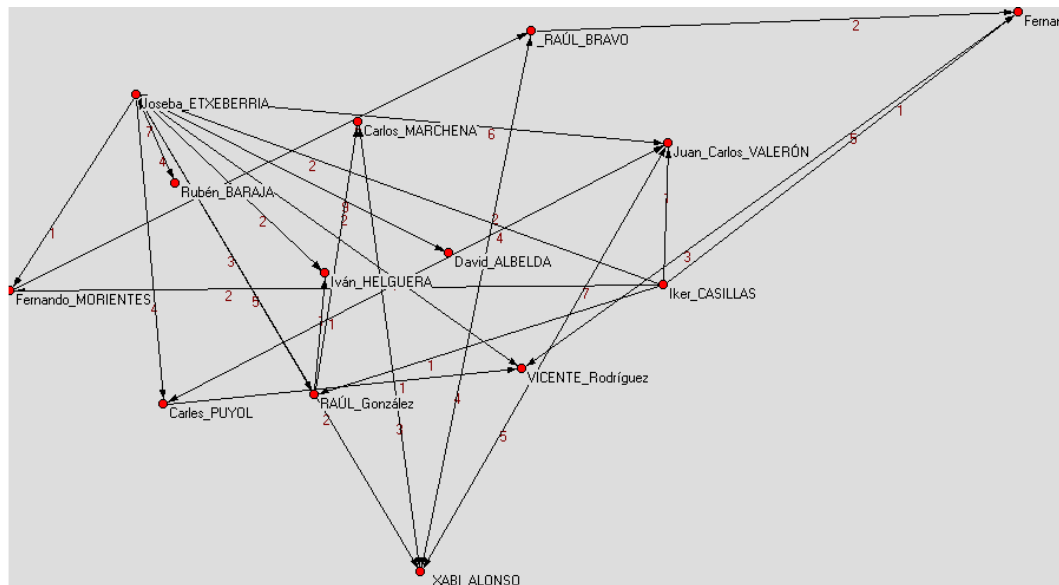
Una red donde, curiosamente, el jugador que tiene más centralidad es Iker Casillas, cuando debería ser un medio como Baraja

APLICACIONES

Análisis de juego en equipos de fútbol (5)

La situación no cambió mucho en el segundo partido (Grecia 1 – España 1) salvo que, en este caso, Albelda, Baraja y Helguera organizaron un poco más el juego. Y Fernando Torres a su bola, claro

Casi el 90% de los pases fueron los mismos. Esta es la diferencia de las dos redes (sin Joaquín porque es un nodo nuevo)



Es curioso ver también que la "autoridad" de la red es Vicente, un extremo. Lo lógico sería que las autoridades fueran los delanteros, pero Morientes y Raúl se hallan ahí perdidos, en la maraña de la red

APLICACIONES

Aprendizaje Colaborativo por Ordenador (1)

Aprendizaje Colaborativo por Ordenador: énfasis en las relaciones entre los actores en un curso on-line en BSCW

Relación entre el profesor y los alumnos en un curso, así como entre los propios alumnos de distintos grupos

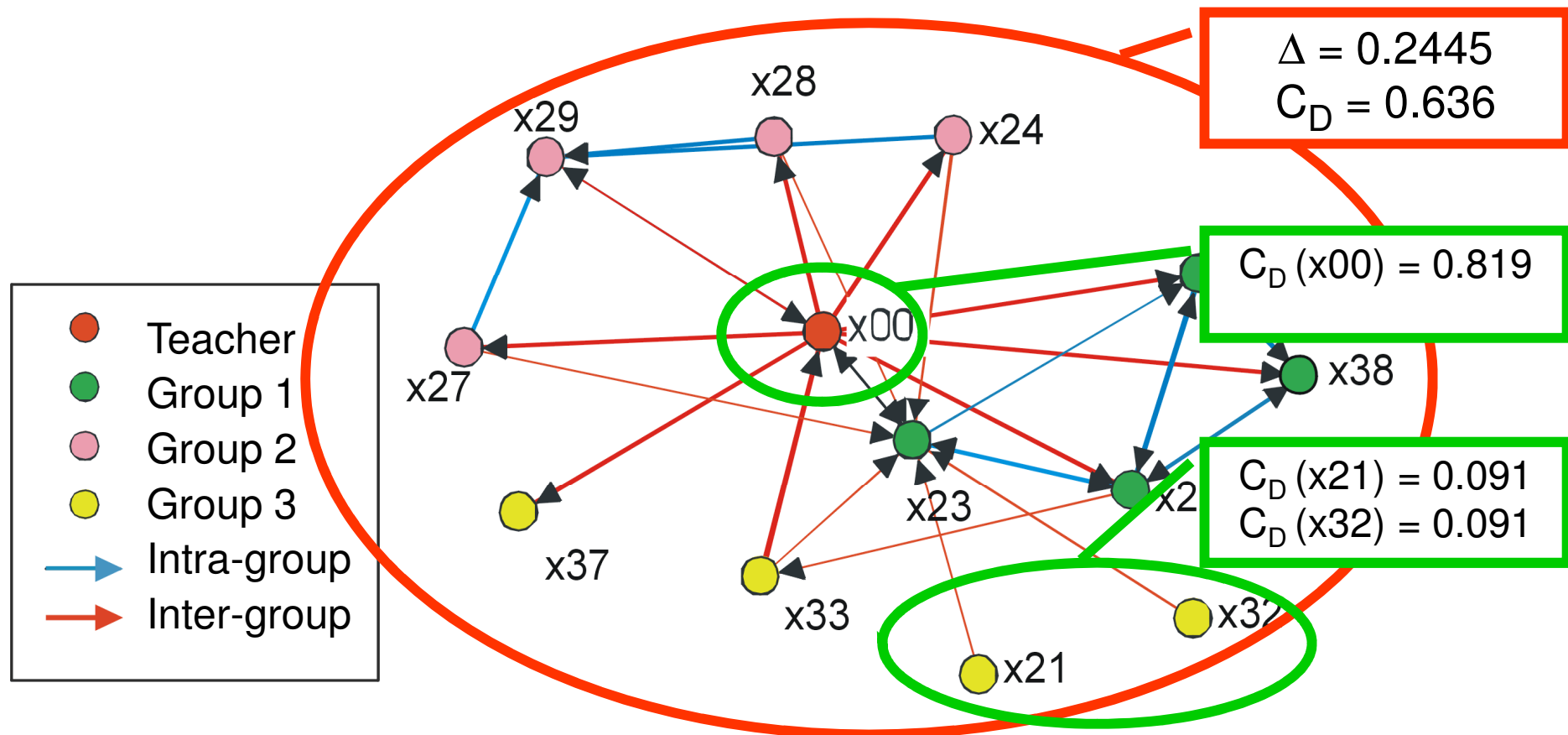
Red Social con distintas variantes: trimodal egocéntrica (tres grupos de alumnos-profesor) y trimodal, unimodal completa (todos los alumnos con todos) y trimodal completa (miembros de un grupo con los de otros)

Pregunta global: ¿Cómo ayudar a los profesores a monitorizar aspectos colaborativos de aprendizaje mediante la tecnología?

Análisis con dos medidas globales (Densidad de la red Δ y Centralización de grado C_D) y dos medidas locales (Centralidades de grado y de cercanía)

APLICACIONES

Aprendizaje Colaborativo por Ordenador (2)



Análisis de influencia de los actores en las discusiones del curso Workshop on interaction analysis approaches (CSCL 2009)

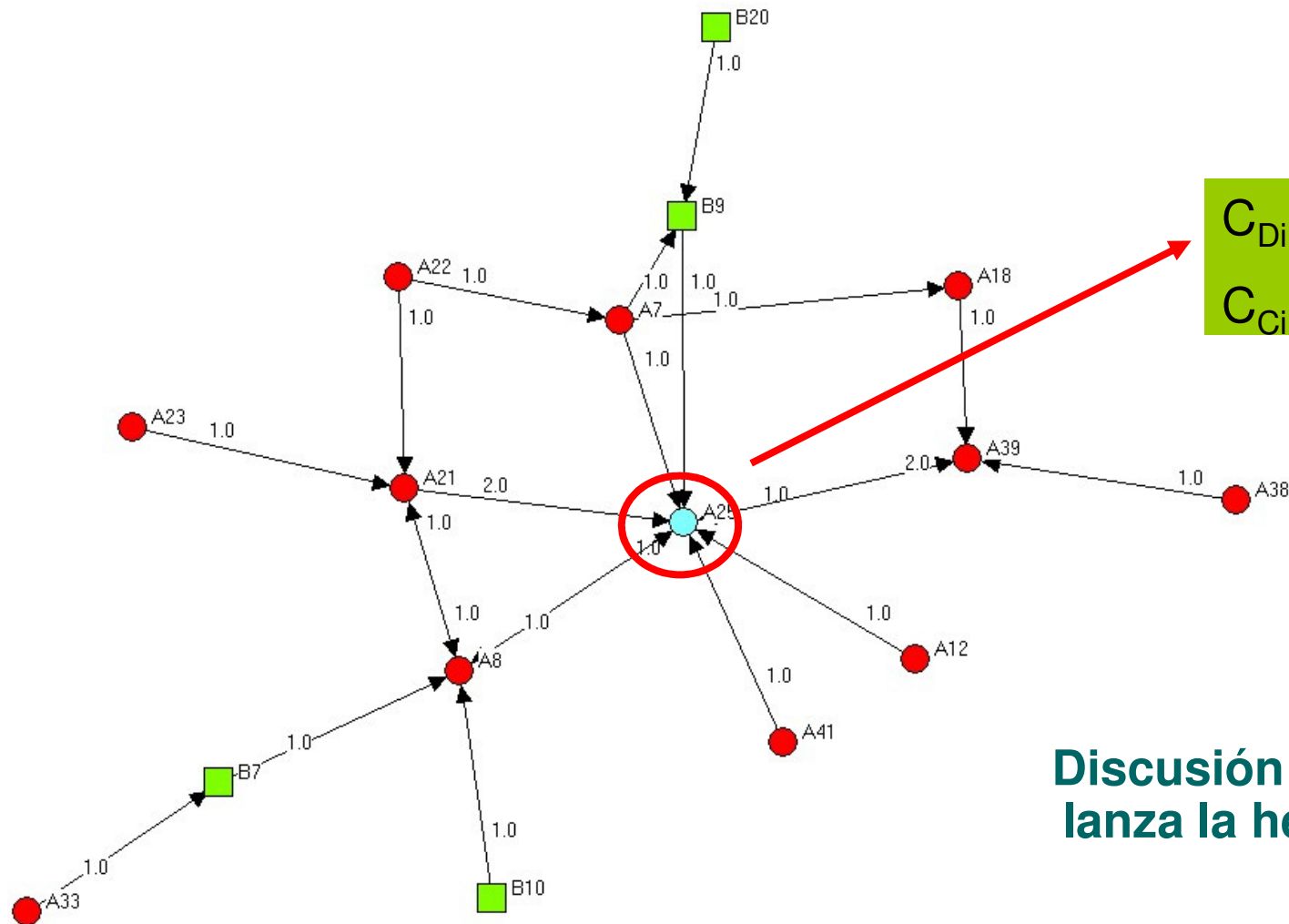
ID	Title	Thread_ID	Parent_ID	date	Student Name	School
21	Solutions of it is Energy Conservation?	3	0	2007-04-30	A41	A
22	I think it is too late for us to solve the global warming	3	21	2007-05-19	A5	A



```

<SESSION id="Thread_3" date=30.04.2007>
<ACTION>
  <ACT.TIMESTAMP>19.05.2007 00:00:00</ACT.TIMESTAMP>
  <ACT.SOURCE ref="A5" />
  <ACT.DISC>
    <ACT.DIR type="Debate">
      <ACT.DIR.DEST ref="A41" />
    </ACT.DIR>
  </ACT.DISC>
</ACTION>
</SESSION>

```



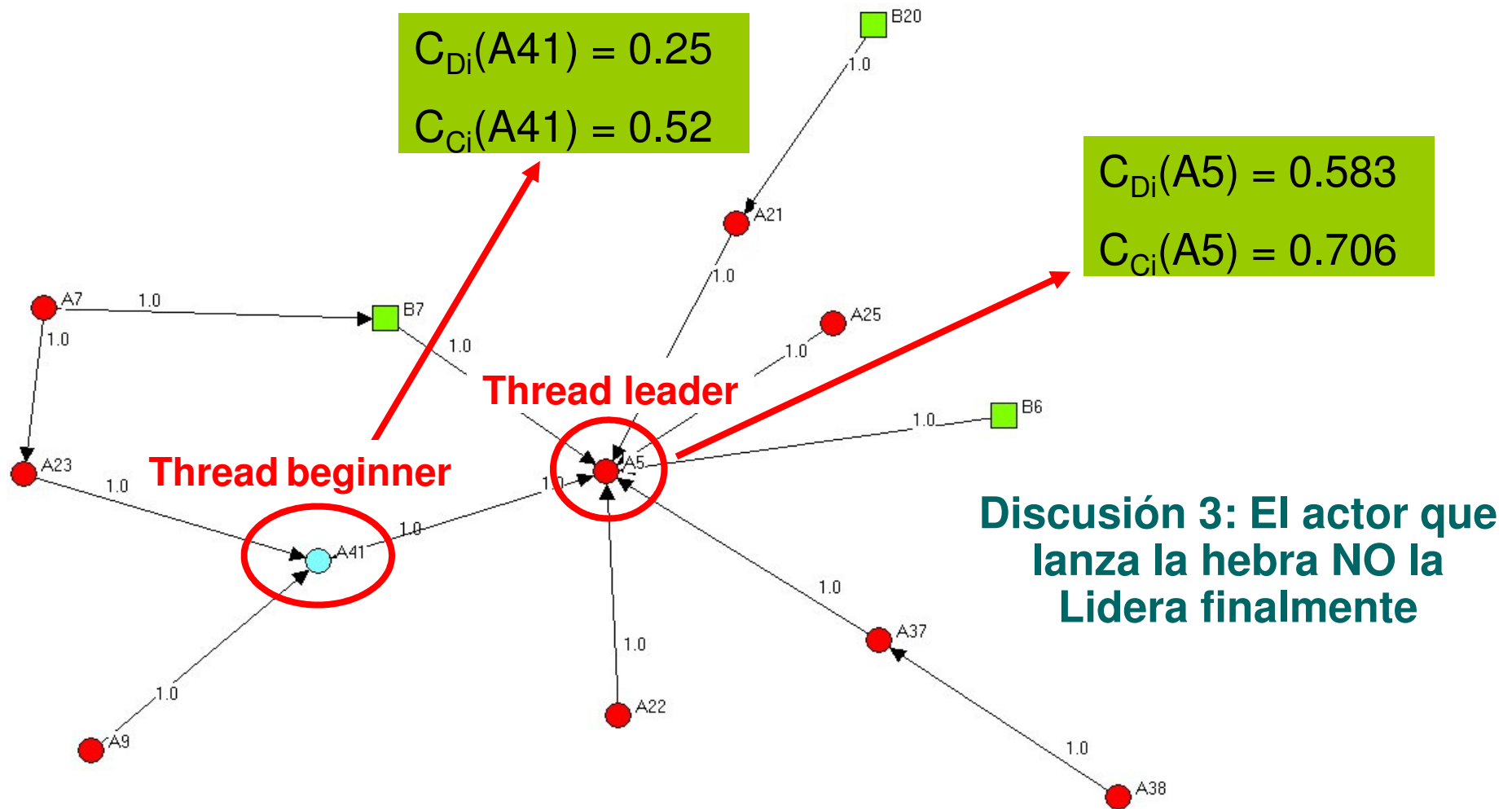
$$C_{Di}(A25) = 0.467$$

$$C_{Ci}(A25) = 0.625$$

Discusión 6: El actor que lanza la hebra es el líder

APLICACIONES

Aprendizaje Colaborativo por Ordenador (5)



APLICACIONES

Aprendizaje Colaborativo por Ordenador (6)

Non-participative

Isolated

