

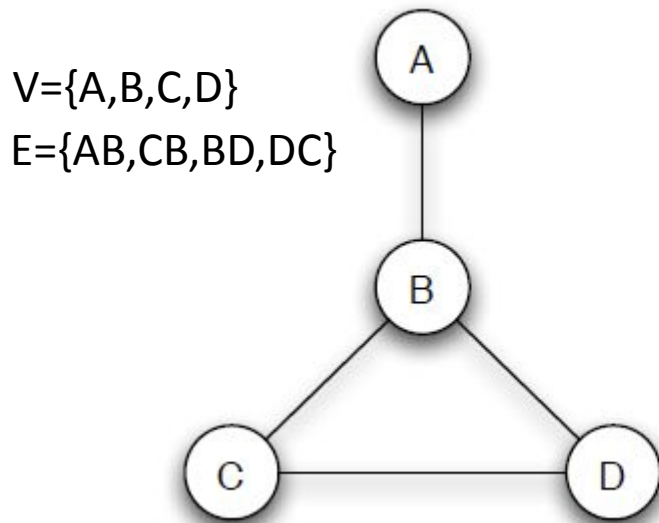
# **Análisis de Redes (Sociales)**

## **Caminos y Conectividad**

# Conceptos Básicos

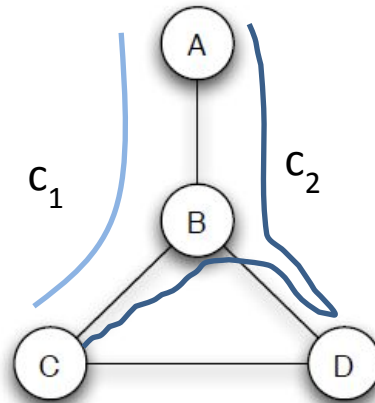
- Un grafo ( $G$ ) es un conjunto de vértices ( $V=\{v_i\}$ ) conectados mediante enlaces ( $E=\{e_j\}$ ).  $G=G(V,E)$

Grafo de 4 nodos y 4 enlaces



relaciones simétricas

# Caminos y ciclos

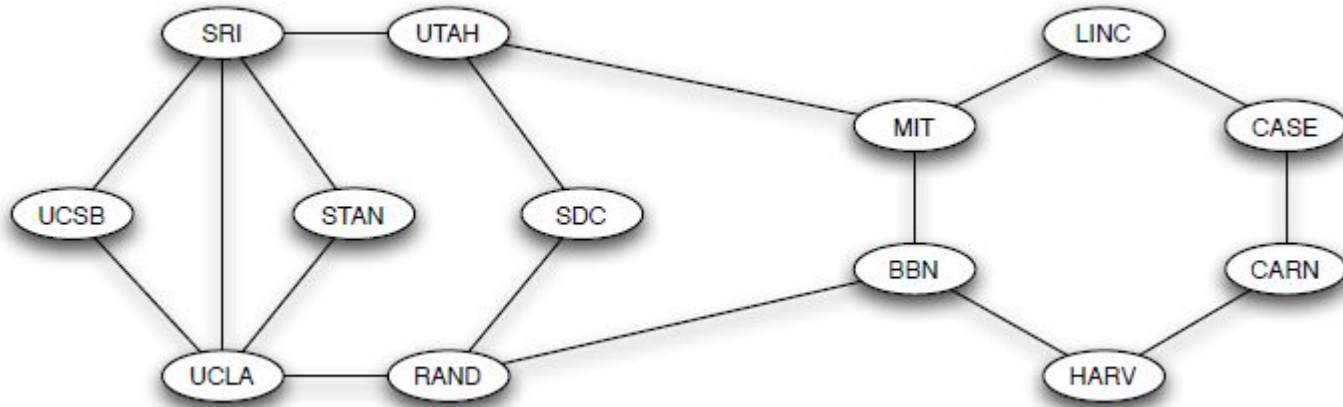


**camino:** secuencia de vértices tales que vértices consecutivos están conectados  
 $c_1 = \{A, B, C\}$ ,  $c_2 = \{A, B, D, B, C\}$

**camino simple:** es un camino que no repite vértices  
 $c_1$  es simple,  $c_2$  no lo es

**ciclo:** camino de más de dos vértices que no repite elementos, salvo el primero y último  
 $c_3 = \{B, C, D, B\}$  es un ciclo.

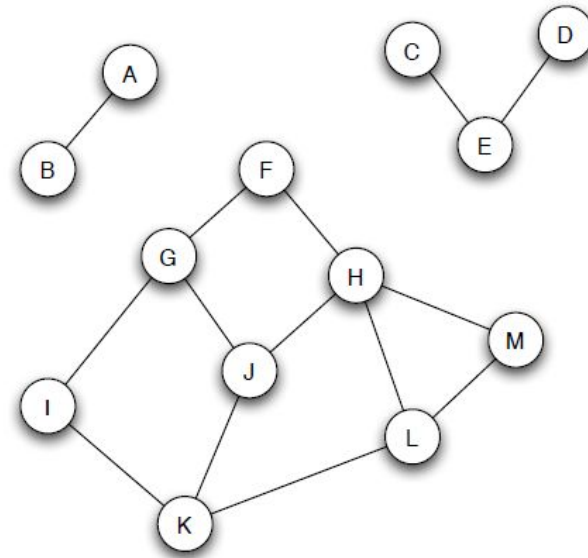
# Ciclos



- Arpanet (Internet 1970) - Red de comunicaciones que vinculaba 13 universidades
- **Por diseño** cada enlace pertenecía a un ciclo (robustez vía redundancia)
- Nota: En redes sociales suelen aparecer ciclos todo el tiempo: conoces a tal, que conoce a tal?

# Conectividad: componentes de un grafo

No siempre es posible establecer un camino entre cualquier par de nodos



**grafo conexo:** Un grafo  $G$  es conexo si existe un camino entre cualquier par de vértices del mismo.

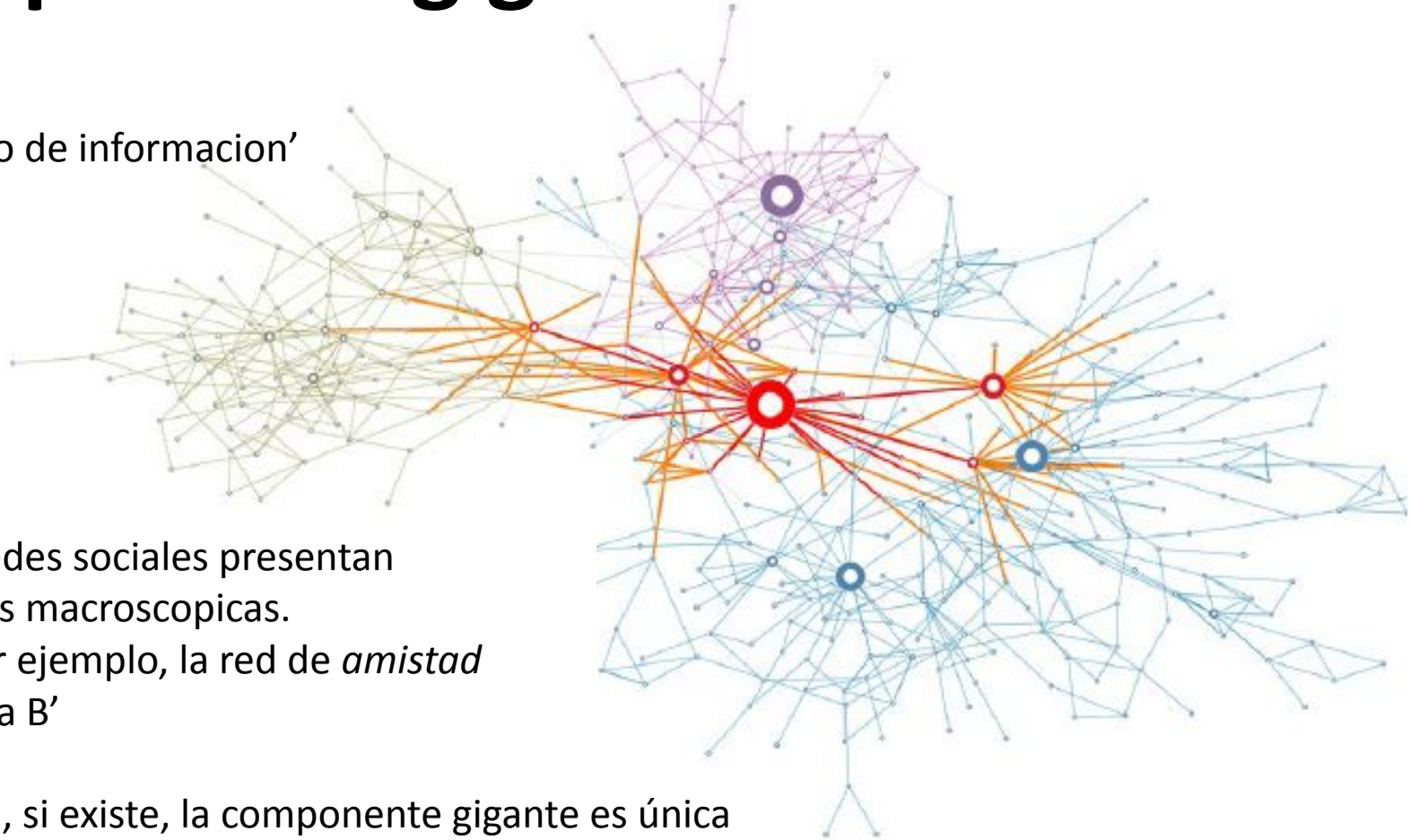
**componente:** subconjunto de vértices de un grafo que:

- I. definen un (sub)grafo conexo
- II. no están incluidos en ninguna otra estructura conexa

**componente gigante:** componente más grande de un grafo

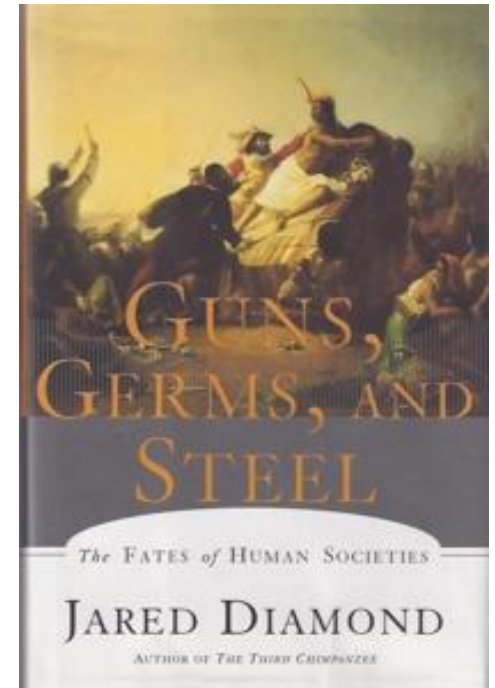
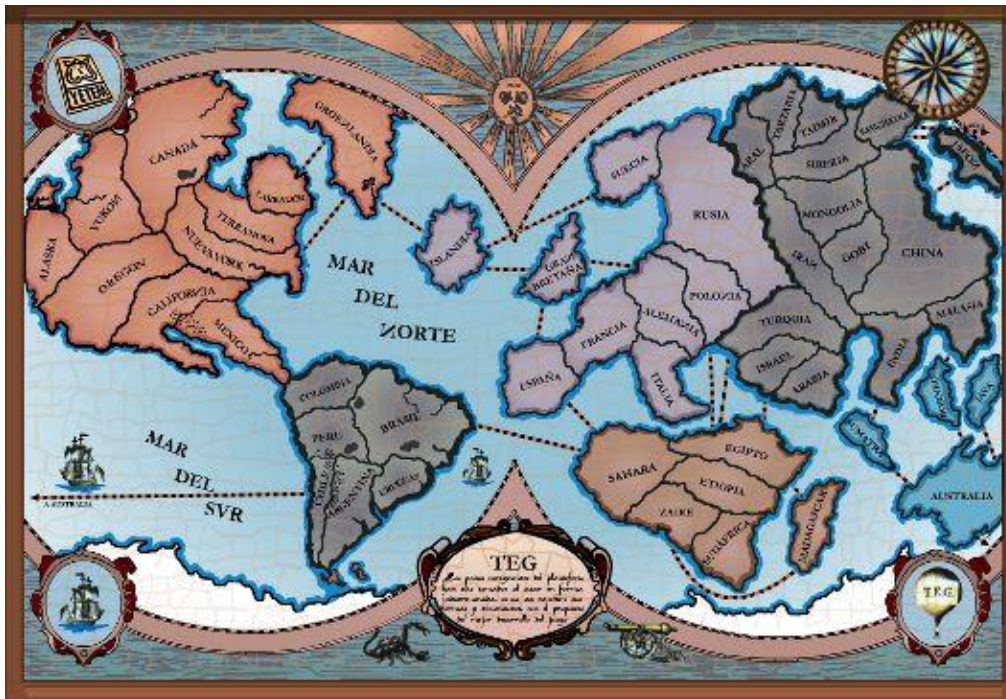
# Componente gigante

Red de 'flujo de informacion'



- Muchas redes sociales presentan estructuras macroscópicas.
- Pensar por ejemplo, la red de *amistad* 'A conoce a B'
- En general, si existe, la componente gigante es única
- Si hubiera más de una, la existencia de al menos un enlace entre objetos de componentes distintas tiende a aparecer con probabilidad  $\sim 1$  (bastaría que alguien de la CG1 'se haga amigo' de alguien de la CG2).

# Componentes gigantes en sistemas sociales



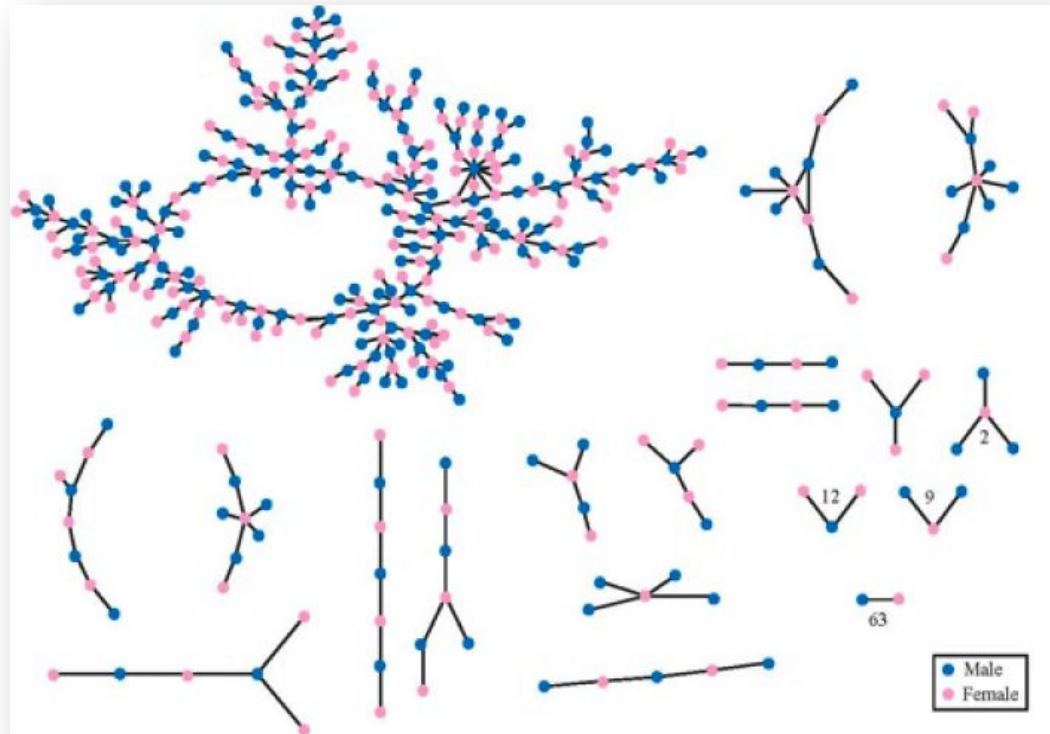
- El 'merging' de dos componentes gigantes puede tener consecuencias enormes.
- Pensar el choque de culturas/tecnologías/sistemas\_inmunitarios que hasta 1500 evolucionaron de forma casi aislada Europa-América

# La red del amorrrr

## Estructura de amor adolescente



- En entrevistas individuales se pidió a jóvenes de una escuela secundaria identificar hasta 3 personas con las que hayan mantenido relaciones románticas y/o no-románticas (sexuales) en los últimos 18 meses



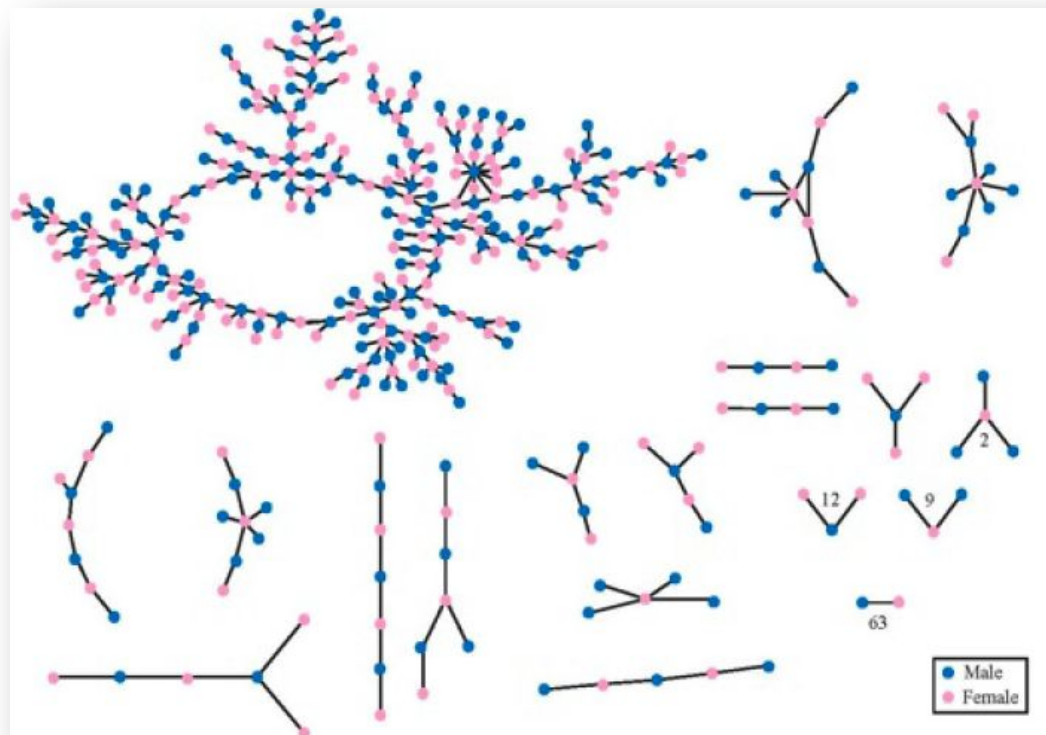


# La red del amorrrr

## Estructura de amor adolescente



- La red es *practicamente* **bipartita** (nodos azules se enlazan con rosas)
- Hay componentes diádicas y trios
- Muy pocas de tamaño intermedio
- Existe una **componente gigante** (52% de chic@s)
- La **distribución de grado** de los nodos similar a la de una red aleatoria (media bien definida)

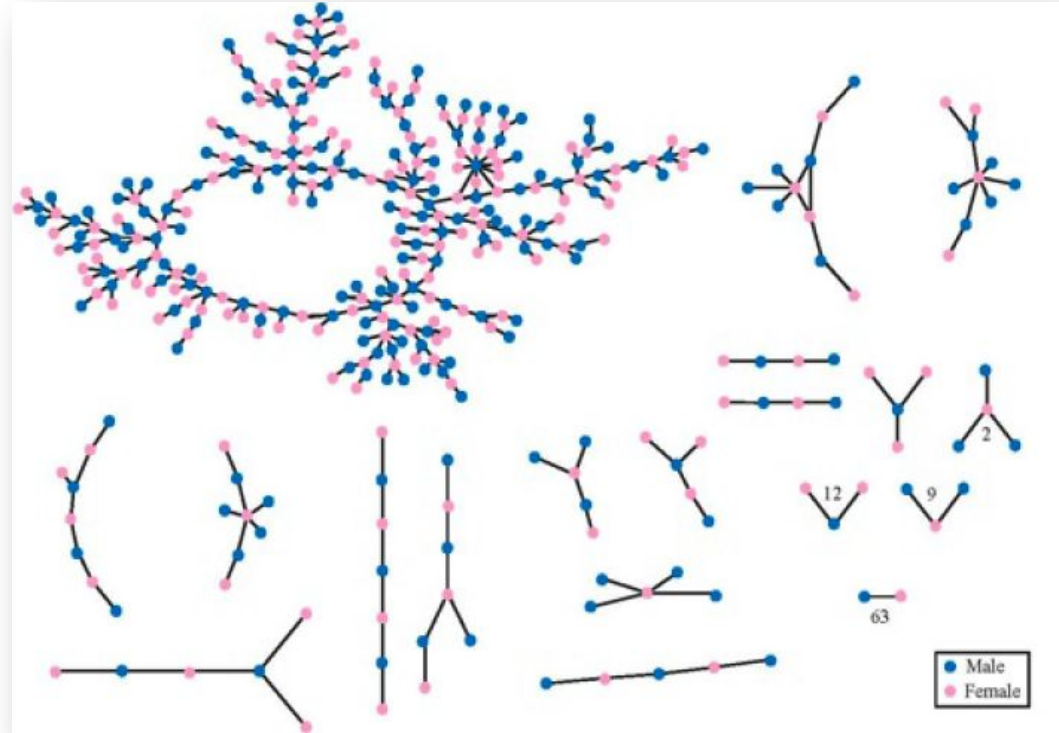


# La red del amorrr

## Estructura de amor adolescente

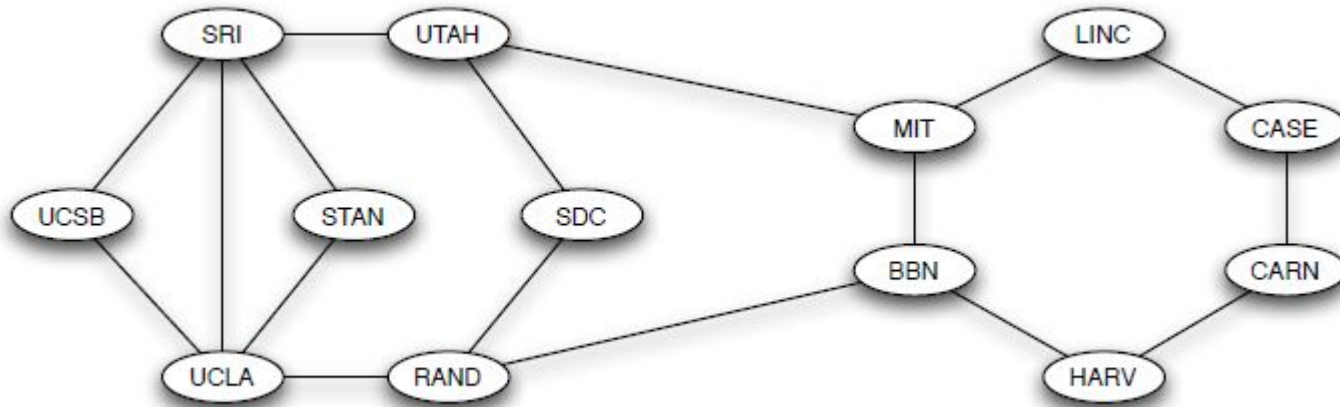


- Ausencia de **ciclos** (solo hay un gde en CG), numerosas ramas cortas, estructuras tipo arbol: a medida que avanzo encuentro nodos nuevos
- Estructuras relevante para analizar riesgos de enfermedades de transmisión sexual



"...These structures reflect relationships that may be long over, and they link individuals together in chains far too long to be the subject of even the most intense gossip and scrutiny. Nevertheless, they are real: like social facts, they are invisible yet consequential macrostructures that arise as the product of individual agency."

# Longitud y Distancia



**longitud** de un camino  $C$ : número de pasos que contiene desde su comienzo hasta su final.

**distancia** entre un par de vértices de un grafo: longitud del camino más corto que los une.

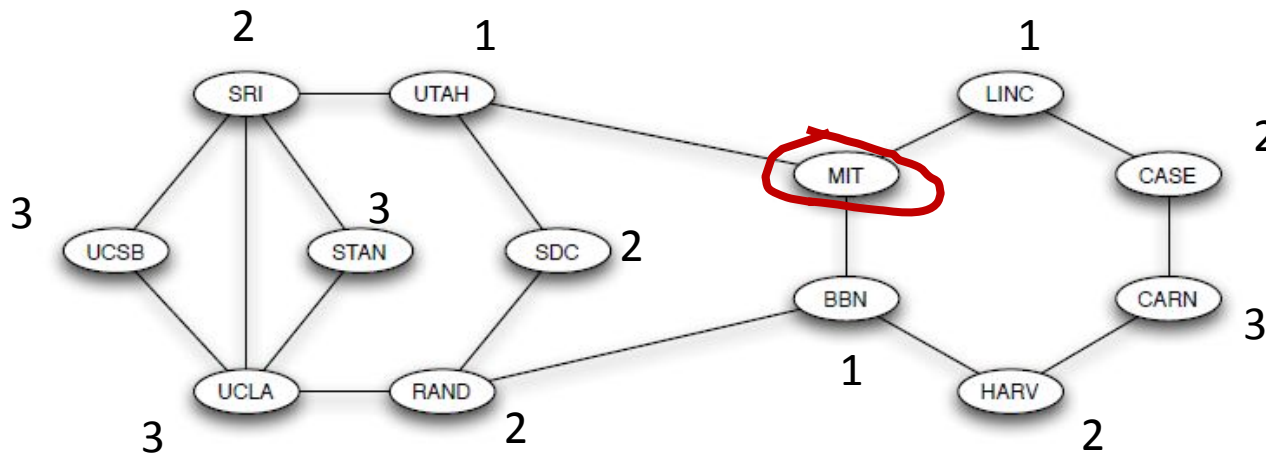
$$d(\text{UCLA}, \text{UTAH})=2$$

$$d(\text{CARN}, \text{UCSB}) =5$$

# Que tan lejos?

- Búsqueda en anchura (Breadth-first search)

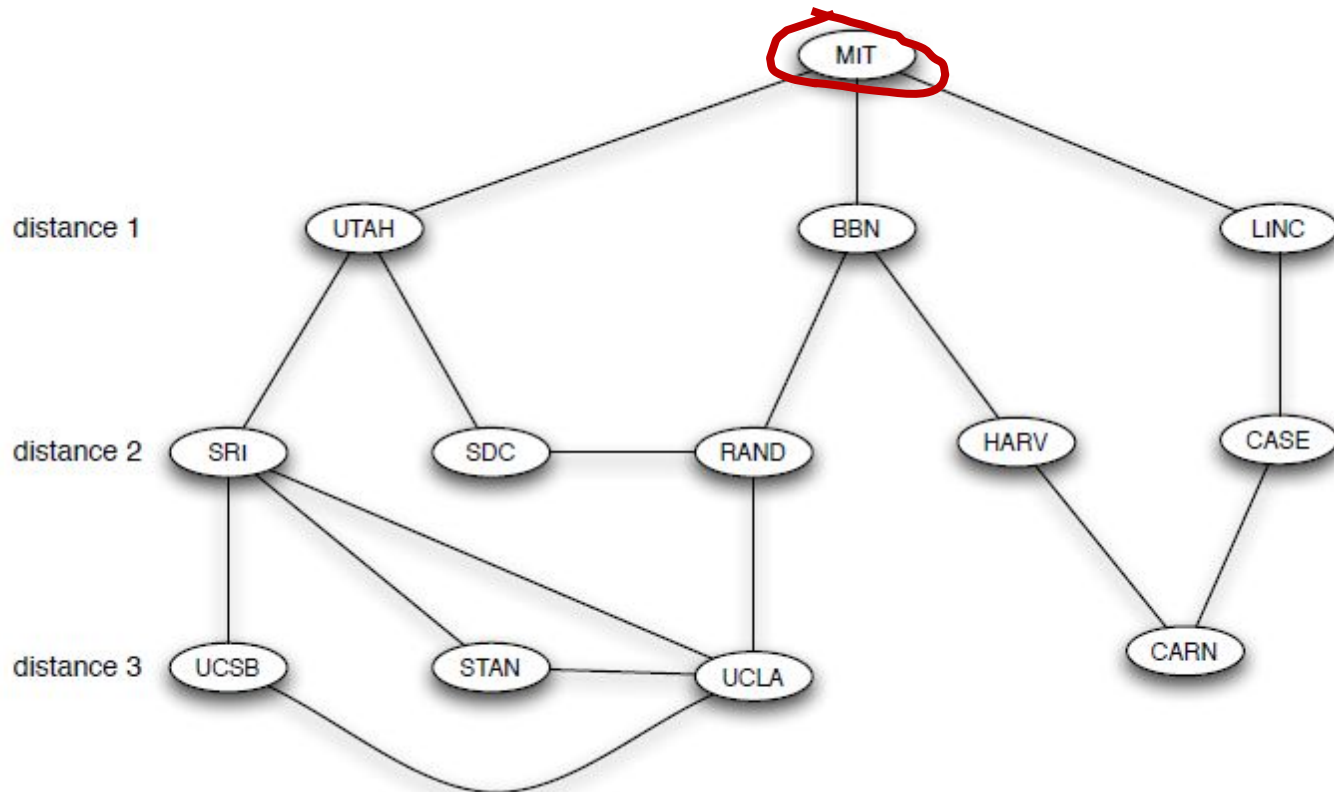
Manera eficiente y sistemática de estimar distancias a un nodo



- I. A vecinos directos de MIT,  $N(MIT)$ , le corresponde  $d=1$
- II. A vecinos directos de vecinos directos de MIT,  $N(N(MIT))$  aun no etiquetados le corresponde  $d=2$
- III.  $d(MIT, N(N(N(MIT)))) = 3$
- IV. etc..

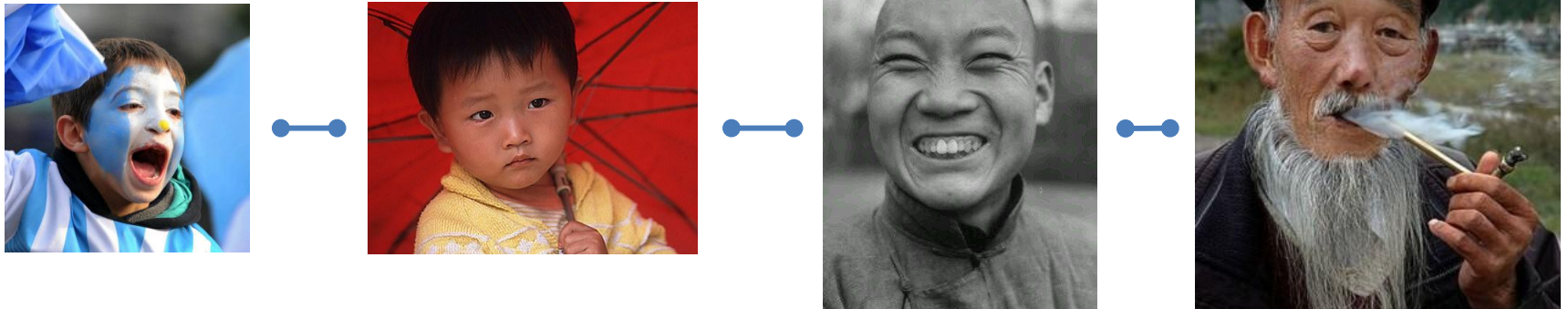
# BFS

- Jerarquía inducida por el BFS



# Mundo Pequeño

- Redes sociales (A conoce a B)
  - presentan una **componente gigante** (existen caminos)
  - suelen existir caminos inesperadamente cortos



$$d(\text{[Boy]} \text{ [Elderly Man]}) = 3$$

# Mundo Pequeño



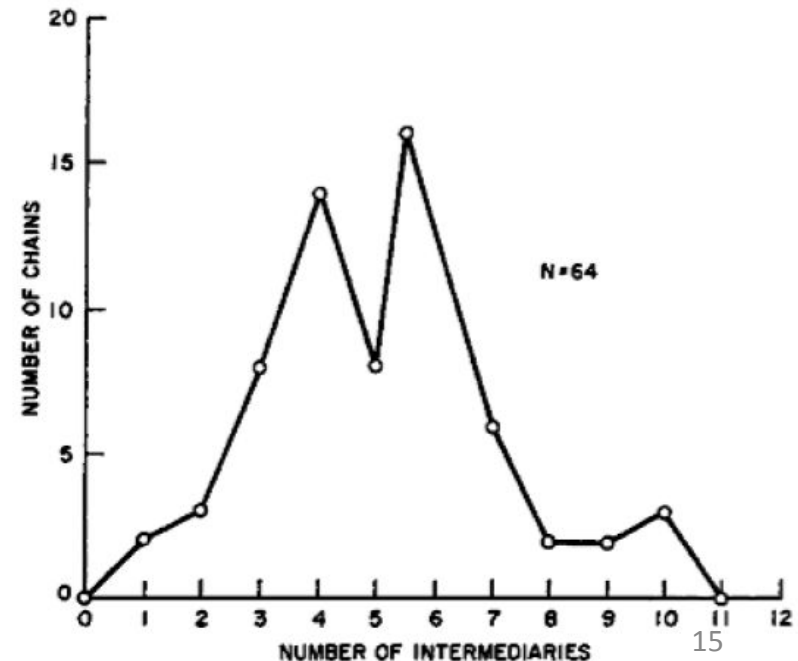
Hipótesis: estamos conectados **globalmente** por medio de **cadena cortas** de conocidos

Experimento:

- I. Selección aleatoria de 296 personas de la costa oeste.
  - II. Tarea: hacer llegar una carta a una dada persona (corredor de bolsa de dirección conocida en Boston)
  - III. Sólo podían entregar la carta a alguien que conocieran, y encomendarle II y III
- ✓ 64/296 cadenas llegaron a destino
  - ✓ Mediana de intermediarios: 6

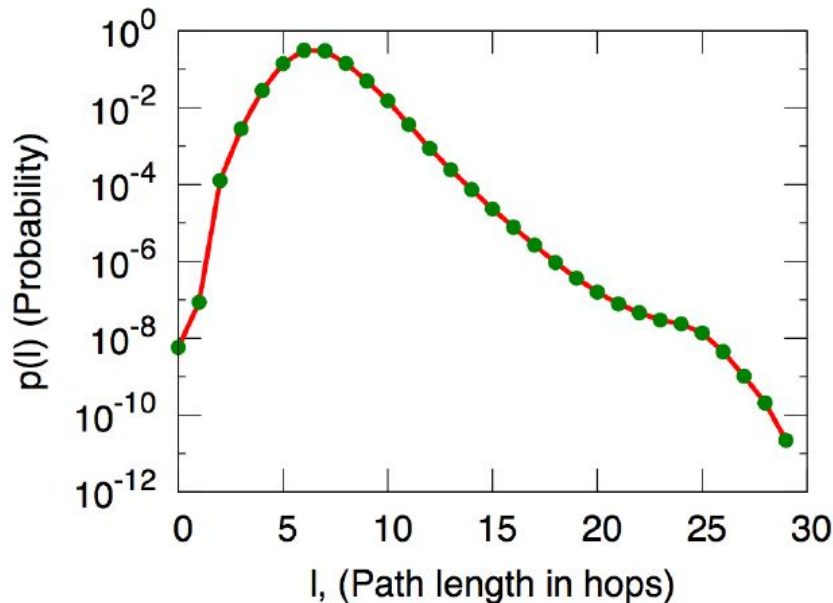
Notas:

- Conclusión sobre prop de la red usando estimación via *trazadores*
- No se desprende del estudio que entre cualquier par de puntos de la red hay 6 grados de separacion
- 6 es mucho o poco? Invitarías a un grado-6 a tu cumpleaños?



# Milgram reloaded

- Analisis de 240.000.000 de cuentas de Microsoft Instant Messenger activas, durante un mes.
- Enlaces entre usuarios que intercambian mensajes



- ✓ Sampling: BFS para 1000 usuarios
- ✓ Estimación un poco más global, aunque aún sesgada (personas tecnoalfabetizadas)
- ✓ Existencia de componente gigante que engloba a la mayor parte de la red
- ✓  $\langle d \rangle \sim 6.6$  (!)





# THE ORACLE OF BACON

- Welcome
- Credits
- How it Works
- Contact Us
- Other stuff »

© 1999-2021 by Patrick Reynolds. All rights reserved.

Data from Wikipedia, used under CC BY-SA 3.0.

## How the Oracle of Bacon Works

Every couple of weeks the Oracle [downloads every English-language article](#) from [Wikipedia](#). Using an open-source [script](#), we produce a [JSON file](#) with 128,000 films and 358,000 actors and actresses.

There is a database service running at all times that stores the database file in memory. It is [written in C](#) for maximum CPU and memory efficiency. The service handles three different types of requests:

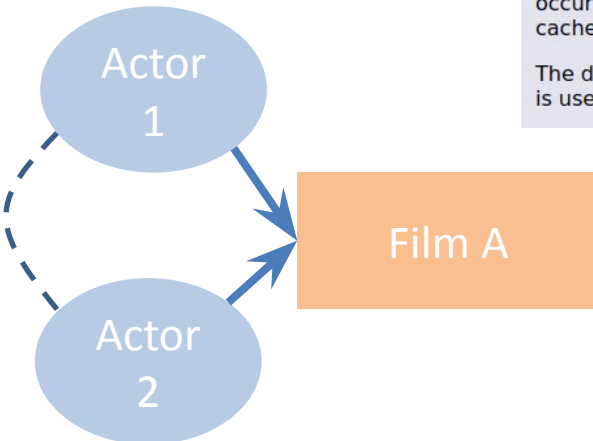
- [Find the link from Actor A to Actor B.](#)
- [How good a "center" is a given actor?](#)
- [Who are all the people with an Actor A number of N?](#)

There are several PHP pages -- one for each of the above types of queries -- that run on the Oracle of Bacon web server, which all connect to the database service using TCP.

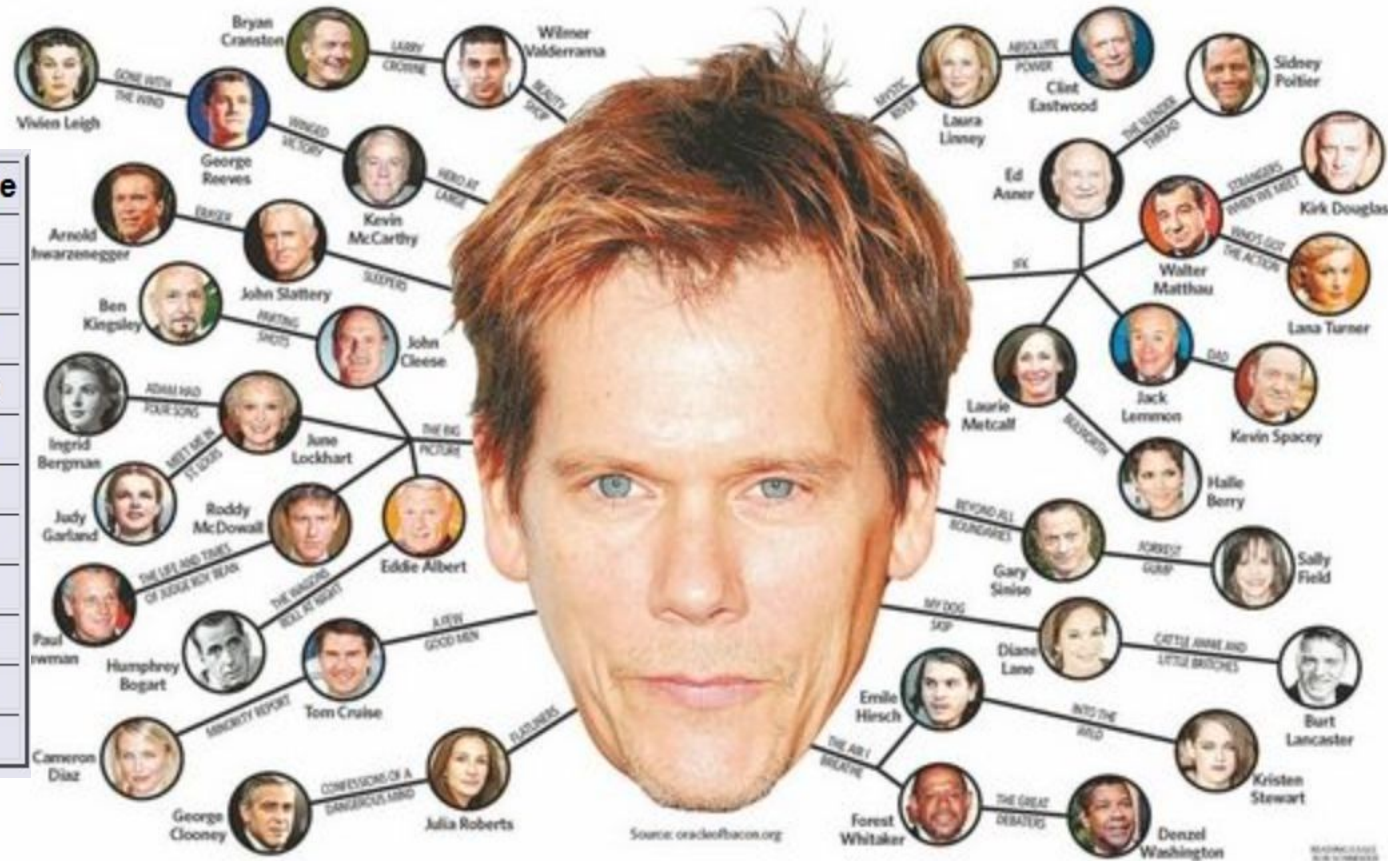
The database service uses a breadth-first search (BFS) to find the shortest path between pairs of actors. If you want to dig further into how shortest-path algorithms work, I recommend the textbook [by Cormen, Leiserson, Rivest, and Stein](#) as an excellent place to start. Other algorithms textbooks are likely to cover the subject as well, if *Introduction to Algorithms* isn't available. You may also look at materials that I wrote to explain graph algorithms (including BFS) to Duke undergraduate CS students [here](#).

Whenever the Oracle answers a query, the results are cached so that future requests to link to the same actor will occur more quickly. About 95% of all queries can be served instantly from the result cache. The current contents of the cache (i.e., which actors can be linked quickly) can be found [here](#).

The database server runs on an [Amazon EC2](#) node with Ubuntu Linux. It consumes 500MB of RAM, about half of which is used for the [results cache](#).



# Mundo Pequeño



Source: oracleofbacon.org

REPRODUCED WITH PERMISSION

Kevin Bacon Number	# of People
0	1
1	3150
2	373876
3	1340703
4	340756
5	28820
6	3383
7	451
8	52
9	8
10	1

<d><sub>IMDB</sub> 2.9

# Mundo Pequeño



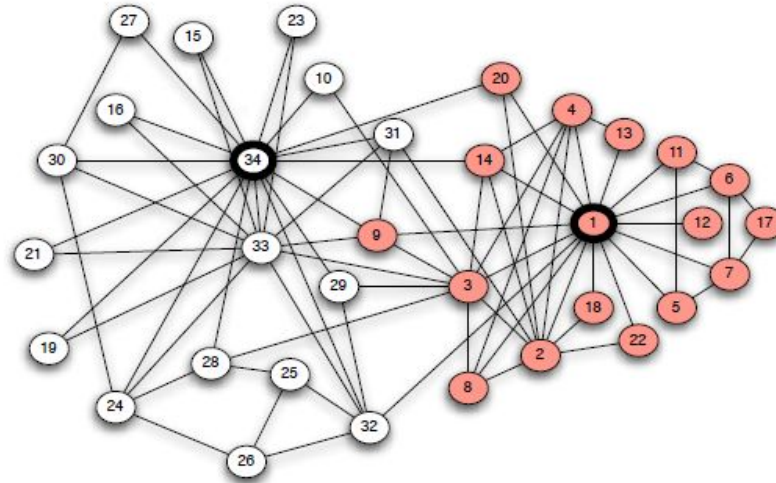
# Camino y conectividad

## Hasta ahora

- Grafo:  $G(V,E)$
- camino, camino simple, ciclo
- componentes, grafo conexo, componente gigante
- longitud de camino, distancia entre nodos
- caminos *sorpresivamente* cortos en redes sociales:  
*mundo pequeño*



# Lazos fuertes / lazos débiles



La estructura de conectividad de las redes sociales es altamente no trivial.

- El entorno de un dado vértice es diferente al de otro.
- Puedo inferir ciertos tipo de *roles* a partir del cableado

Pero hay información adicional que aún no tuvimos en cuenta: Los vínculos interpersonales pueden ser de distinta intensidad. En un escenario simplificado: puede haber **enlaces fuertes y débiles**.  $G(V,E,f:E \times \{w,s\})$

¿Cómo correlaciona/condiciona esta distinción a la estructura de la red?

# Buscando Trabajo



Mark Granoveter (1974) “Getting a Job. A study of contacts and careers”

Entrevista a personas que recientemente cambiaron empleo: ¿cómo obtuvieron información de la oportunidad laboral?

Observación:

- Las personas reciben información sobre nuevos trabajos vía contactos personales
- Esos contactos eran **mayormente** calificados como “conocidos” (weak-ties) y no como “amistades” (strong-ties)

¿Por qué la recomendación de gente que debería conocer más y mejor nuestros intereses no funcionaba mejor??



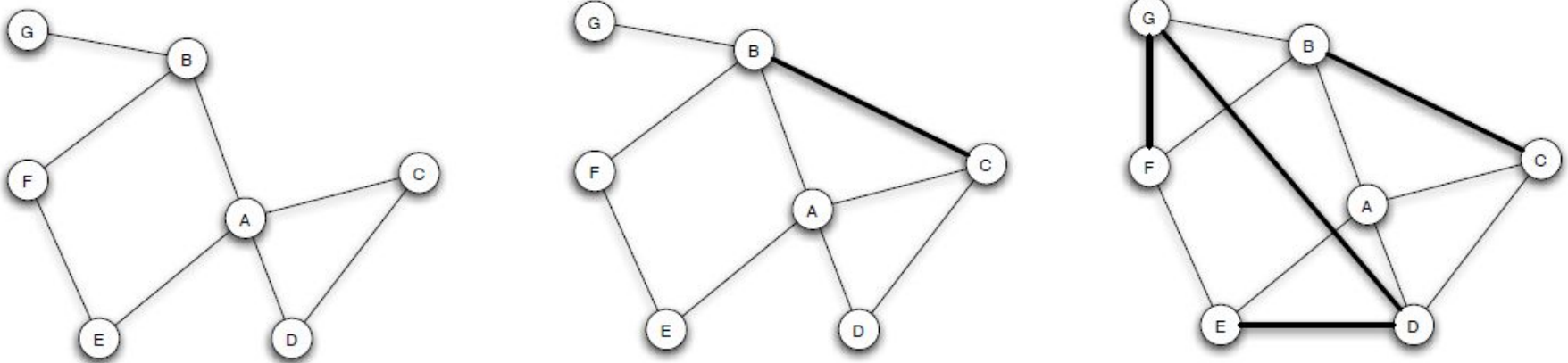
estructura de la red y vínculos locales *mejor (más efectivo)* que conocimiento intensivo de la persona

# veamos un poco....

- Cuales son los mecanismos por los que una red social podría conformarse?

## Principio de **clausura transitiva**:

Si dos personas tienen una amistad en común, aumenta la probabilidad de que se vuelvan amigos entre sí

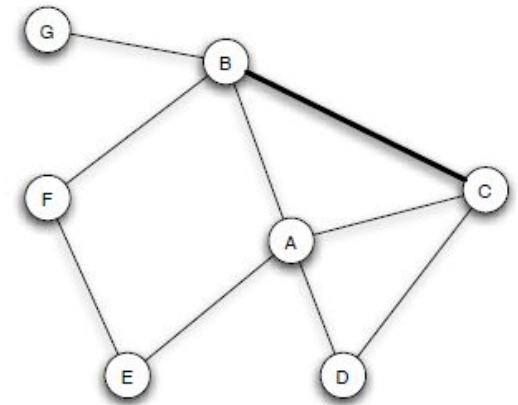




# Clausura transitiva

## Por qué funcionaría?

- 1) Si A y B son amigos y A y C son amigos se amplían las probabilidades de que B y C interactuen y se relacionen
- 2) El hecho de que B y C sean amigos de la misma persona, y ambos lo sepan, genera cierta confianza implícita
- 3) habla de cierta similaridad implícita en compartir la amistad de una misma persona.
- 4) Conveniencia de A de generar el lazo de amistad entre B y C (evitar stress)

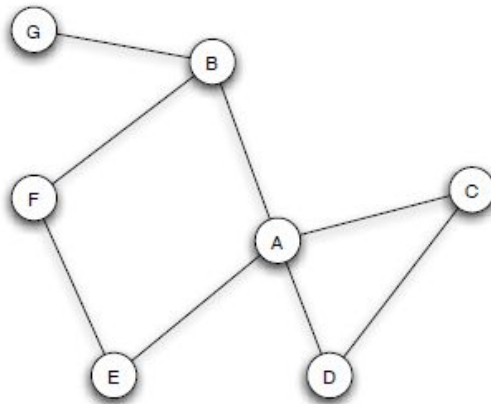


# Son amigos, mis amigos?

Cuantificando la clausura transitiva:

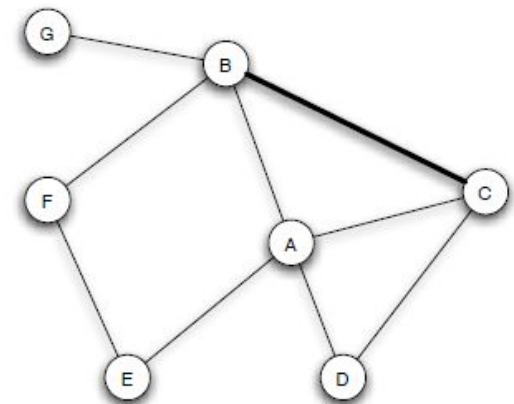
**coeficiente de *clustering*** de nodo A: fracción de vecinos de A, conectados entre sí

$$C(A) = \frac{\#triangulos\ con\ vertice\ en\ A}{\#triangulos\ posibles\ con\ vertice\ en\ A} = \frac{1}{6}$$



**coeficiente de *clustering*** de grafo G:  $C(G) = \frac{1}{N} \sum_{i=1}^N C(v_i) \sim 0.31$

# Clausura transitiva



## Suicide and Friendships Among American Adolescents

| Peter S. Bearman, PhD, and James Moody, PhD

Although suicide rates for most subpopulations tend to be stable over time, the suicide rate among adolescents has risen dramatically in recent years. Suicide is now the third leading cause of death among adolescents and young adults aged 15 to 24 years.<sup>1</sup> Although only 1 in 200 suicide attempts results in death, more than one third of all suicide attempts result in injuries serious enough to require professional treatment. Four percent of American adolescents reported at least 1 attempted suicide in the past year, and 13% of adolescents had seriously considered suicide at least once in the past year.<sup>2</sup> The broader impact of suicide on adolescents is substan-

*Objectives.* We investigated male and female adolescent

*Methods.* We analyzed friendship patterns in the longitudinal Survey of Adolescent and suicidal ideation and suicide with suicidality.

*Results.* Having had a friend with suicidal ideation and attempted suicide was likely to have suicidal thoughts. Among adolescents, friendship patterns were stochastic, with few consistent

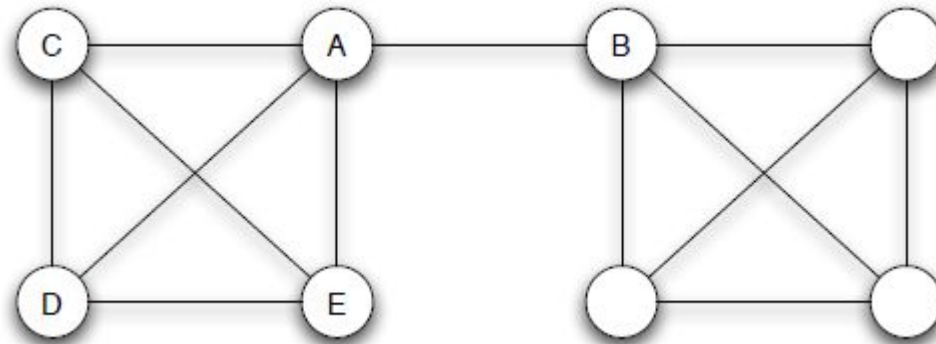
*Conclusions.* The friendship patterns of male adolescents' suicidal ideation and attempted suicide in which friendship patterns in which (2004;94:89–95)

Our findings revealed significant social network effects on suicidal ideation for adolescent females, after we controlled for previously identified correlates of suicidal ideation. Adolescent females who are isolated from the adolescent community or whose social relationships are intransitive and likely dissonant<sup>18</sup> are at greater risk for suicidal thoughts than are girls who are embedded in cohesive friendship groups. Health care providers can assess this risk for girls through simple screens that allow adolescents to describe the relational world in which they are embedded. Adolescent males are more impervious to social context than are adolescent females.

# La fortaleza de los lázos débiles



De vuelta al trabajo...por qué las mejores recomendaciones vienen de *conocidos* y no de *amigos*?



A se vincula con cuatro personas:  $N(A)=\{C,D,E,B\}$

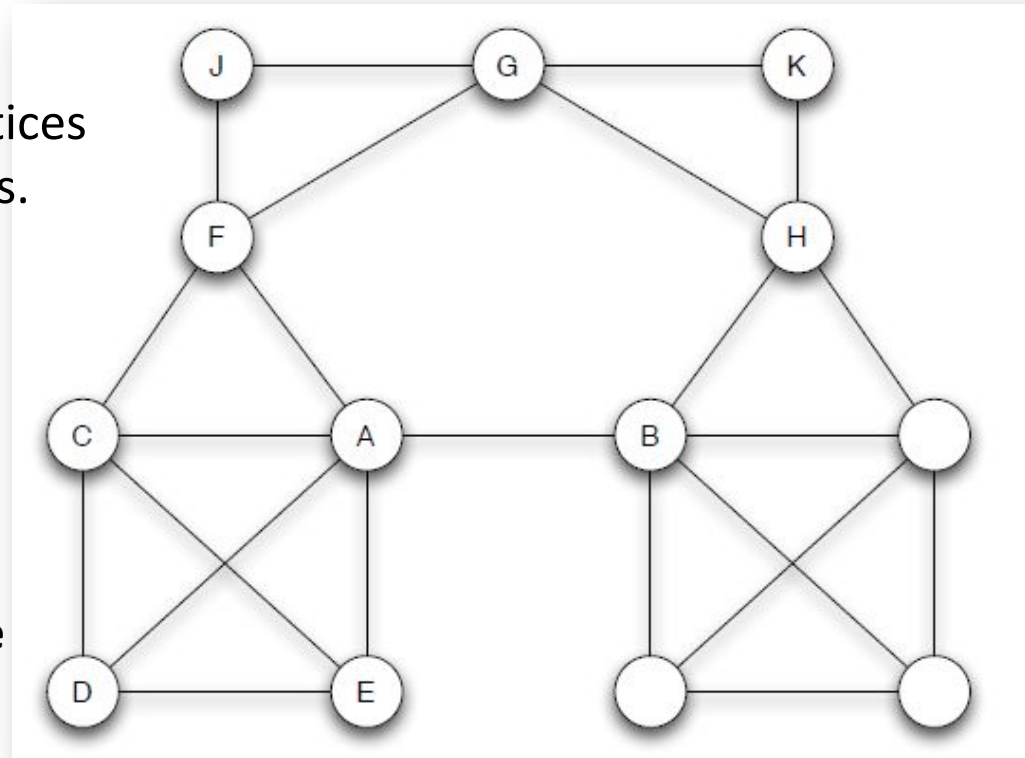
El enlace  $e_{AB}$  es *estructuralmente inusual*. Formalmente es un: **enlace puente**: enlace que si es removido provoca que sus vértices queden en componentes diferentes del grafo.

# Tendiendo puentes

Los enlaces **punto** son raros de encontrar en la práctica en redes sociales reales

- **punto local**: enlace cuyos vértices no comparten primeros vecinos.
- **alcance** de un **punto local** : distancia a la que quedan sus vértices cuando el mismo es removido

Por definición el **alcance** de un **punto local** es estrictamente mayor que 2



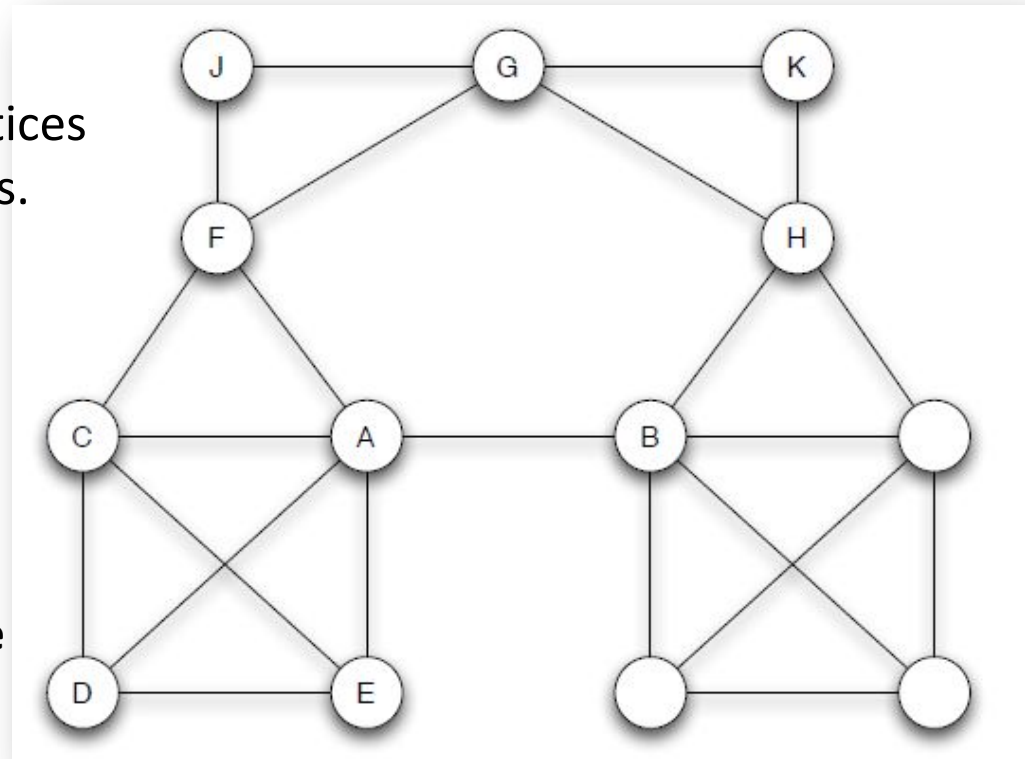
$e_{AB}$  es un **punto local** de alcance 4

# Tendiendo puentes

Los enlaces **punto** son raros de encontrar en la práctica en redes sociales reales

- **punto local**: enlace cuyos vértices no comparten primeros vecinos.
- **alcance** de un **punto local** : distancia a la que quedan sus vértices cuando el mismo es removido

Por definición el **alcance** de un **punto local** es estrictamente mayor que 2



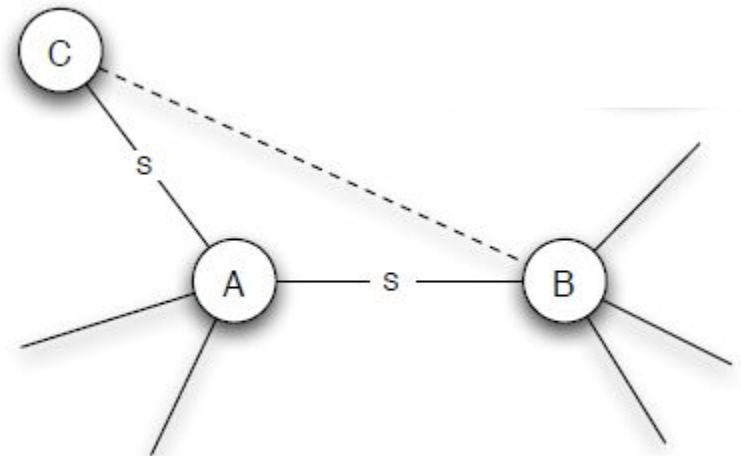
$e_{AB}$  acerca el mundo de  $A$ , al mundo de  $B$ . Facilita el acceso de  $A$  a nueva información

# Clausura Transitiva Fuerte (CTF)

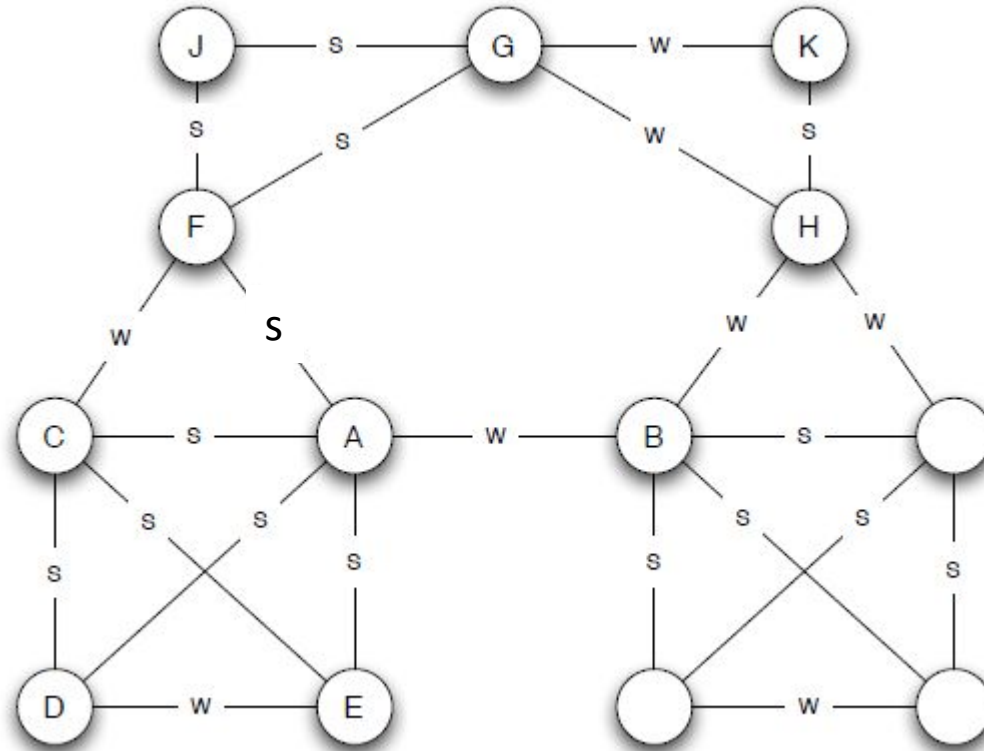
- I. Distinción de vínculos personales: **fuertes, débiles.**
- II. Noción estructural

El concepto **CTF** vincula las dos nociones:

- El nodo **A** viola la propiedad **CTF** si **A** presenta vínculos fuertes con dos de sus vecinos, pero no existe un enlace entre ellos
- El nodo **A** satisface **CTF** si no la viola.



# Clausura Transitiva Fuerte (CTF)

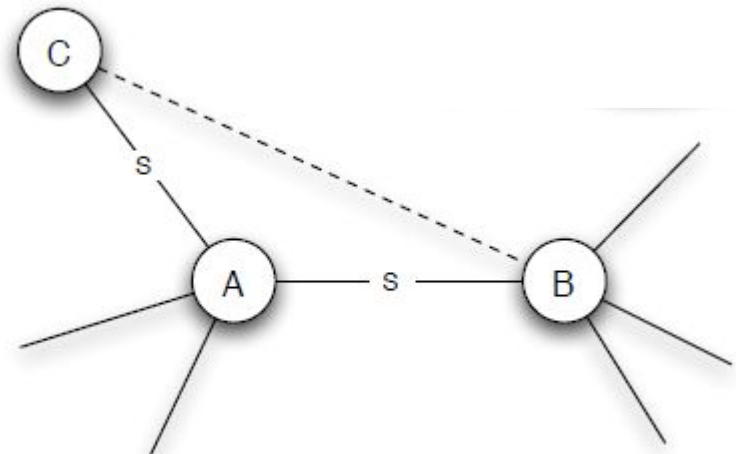
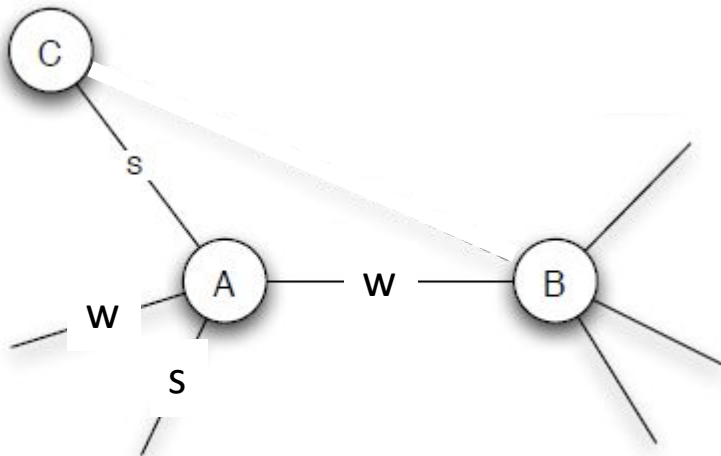


**A viola CTF:**  $e_{AF}$ ,  $e_{AE}$  son fuertes, pero  $e_{EF} \notin E$



# CTF, puentes locales y lazos débiles

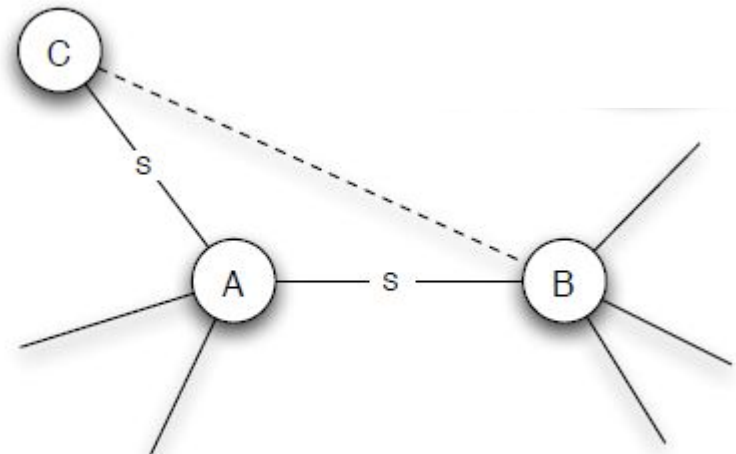
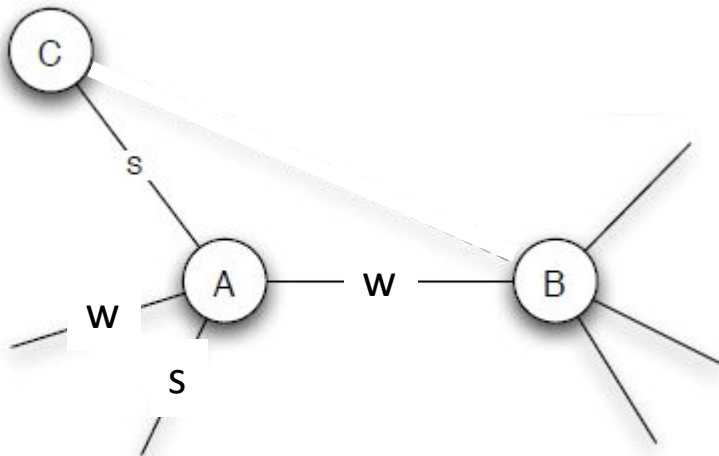
Si un nodo **A** satisface la propiedad de **CTF** y presenta al menos dos **enlaces fuertes**, entonces cualquier **puente local** en el que esté involucrado es necesariamente un enlace **débil**



$e_{AB}$  no puede ser un enlace fuerte y un puente local  
(de otra manera debería existir un enlace CB, y dejaría de ser puente)

# CTF, puentes locales y lazos débiles

En criollo: En concordancia con nuestra hipótesis de trabajo original (sobre como opera la **clausura transitiva**) un **puente local** tiene que ser **débil** porque si no lo fuera aparecerían eventualmente enlaces en la vecindad que harían que dicho **puente local** deje de serlo



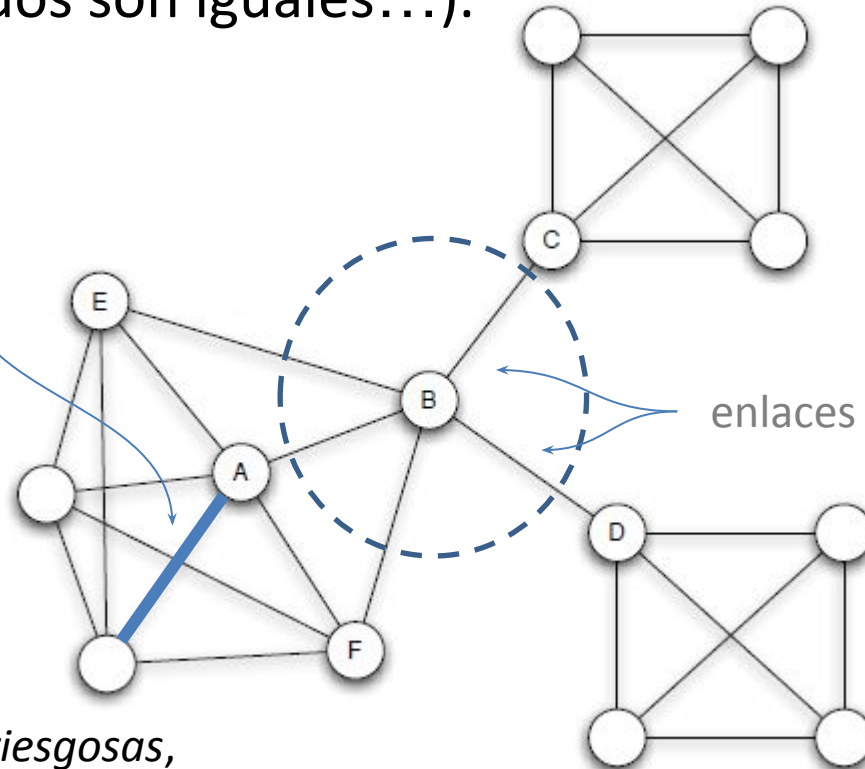
$e_{AB}$  no puede ser un enlace fuerte y un puente local

# Roles topológicos

- El acceso a **enlaces** que conectan grupos no suele estar uniformemente distribuido a lo largo de los **nodos** de la red (no todos los nodos son iguales...).

enlace de alto *embedding*  
#vecinos comunes de  
sus nodos extremos

enlaces de bajo *embedding*



Caracterización de nodos:

**A** se conecta con gente mediante enlaces de alto embedding (confiabilidad)

Las relaciones de **B** son mas *riesgosas*, pero tienen ventajas respecto a flujo de info

← Tweet



Steven Strogatz  
@stevenstrogatz

LinkedIn Ran Social Experiments on 20 Million Users Over Five Years (and provided evidence for Granovetter's theory of "the strength of weak ties", which says you're more apt to find a job through an acquaintance than through a close friend).

The New York Times

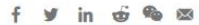
### LinkedIn Ran Social Experiments On 20 Million Users Over Five Years

nytimes.com



12:28 AM · Sep 26, 2022 · Twitter Web App

REPORT | SOCIAL NETWORKS



## A causal test of the strength of weak ties

KARTHIK RAJAKUMAR, GUILLAUME SAINT-JACQUES, JAVOR BOJINOV, ERIK BRYNJOLFSSON, AND SINAN ARAL [Authors Info & Affiliations](#)

SCIENCE · 15 Sep 2022 · Vol 377, Issue 6612 · pp. 1304-1310 · DOI:10.1126/science.abi4476

4,009 1



### RELATED PERSPECTIVE

Weak ties, failed tries, and success

### The influence of weak associations

The strength of weak ties is an influential social-scientific theory that stresses the importance of weak associations (e.g., acquaintance versus close friendship) in influencing the transmission of information through social networks. However, causal tests of this paradoxical theory have proved difficult. Rajkumar *et al.* address the question using multiple large-scale, randomized experiments conducted on LinkedIn's "People You May Know" algorithm, which recommends connections to users (see the Perspective by Wang and Uzzi). The experiments showed that weak ties increase job transmissions, but only to a point, after which there are diminish-

# Es válida la hipótesis de CT?

- Onnela (2007): análisis de comunicaciones de red móvil de  $4.6 \cdot 10^6$  usuarios por 18 meses.

PNAS PNAS PNAS

## Structure and tie communication n

J.-P. Onnela<sup>\*,†</sup>, J. Saramäki<sup>\*</sup>, J. Hyvö

<sup>\*</sup>Laboratory of Computational Engineering, Helsinki University of Technology, Espoo, Finland; <sup>†</sup>Clarendon Laboratory, Oxford University, Oxford University of Notre Dame, South Bend, IN 46556; Boston, MA 02115; <sup>‡</sup>John F. Kennedy School of Government, Harvard University, Cambridge, MA 02138; <sup>§</sup>Theoretical Physics, Budapest University of Techn

Edited by H. Eugene Stanley, Boston University, I

Electronic databases, from phone to e-mail detailed records of human communication avenues to map and explore the structure of communication networks. Here we examine the c of millions of mobile phone users, allowi

communication network. We observe a coupling between interaction strengths and the network's local structure, with the counterintuitive consequence that social networks are robust to the removal of the strong ties but fall apart after a phase transition if the weak ties are removed. We show that this coupling significantly slows the diffusion process, resulting in dynamic trapping of information in communities and find that, when it comes to information diffusion, weak and strong ties are both simultaneously ineffective.

complex systems | complex networks | diffusion and spreading | phase transition | social systems

Here we examine the communication patterns of millions of mobile phone users, allowing us to simultaneously study the local and the global structure of a society-wide communication network. We observe a coupling between interaction strengths and the network's local structure, with the counterintuitive consequence that social networks are robust to the removal of the strong ties but fall apart after a phase transition if the weak ties are removed

nodes and  $\sim 10^6$  to  $10^7$  links, the vast majority (>90%) of these nodes belonging to a single connected cluster [giant component (GC)]. Given the very large number of users and communication events in the database, we find that the statistical characteristics of the network and the GC are largely saturated, observing little difference between a two- or a three-month-long sample. Note that the MCG captures only a subset of all interactions between individuals, a detailed mapping of which would require face-to-face, e-mail, and land line communications as well. Yet, although mobile phone data capture just a slice of communication among people, research on media multiplexity suggests that the use of one medium for communication between two people implies

# Es válida la hipótesis de CT?

- Onnela (2007): análisis de comunicaciones de red móvil de  $4.6 \cdot 10^6$  usuarios por 18 meses.
- 2 nodos se conectan si intercambiaron una comunicación en ambos sentidos durante el período.
- El peso de un enlace dado por la recurrencia de las llamadas
- Estructura de la red: proxy de la red social subyacente
- **Componente gigante** del 84% de la masa de usuarios

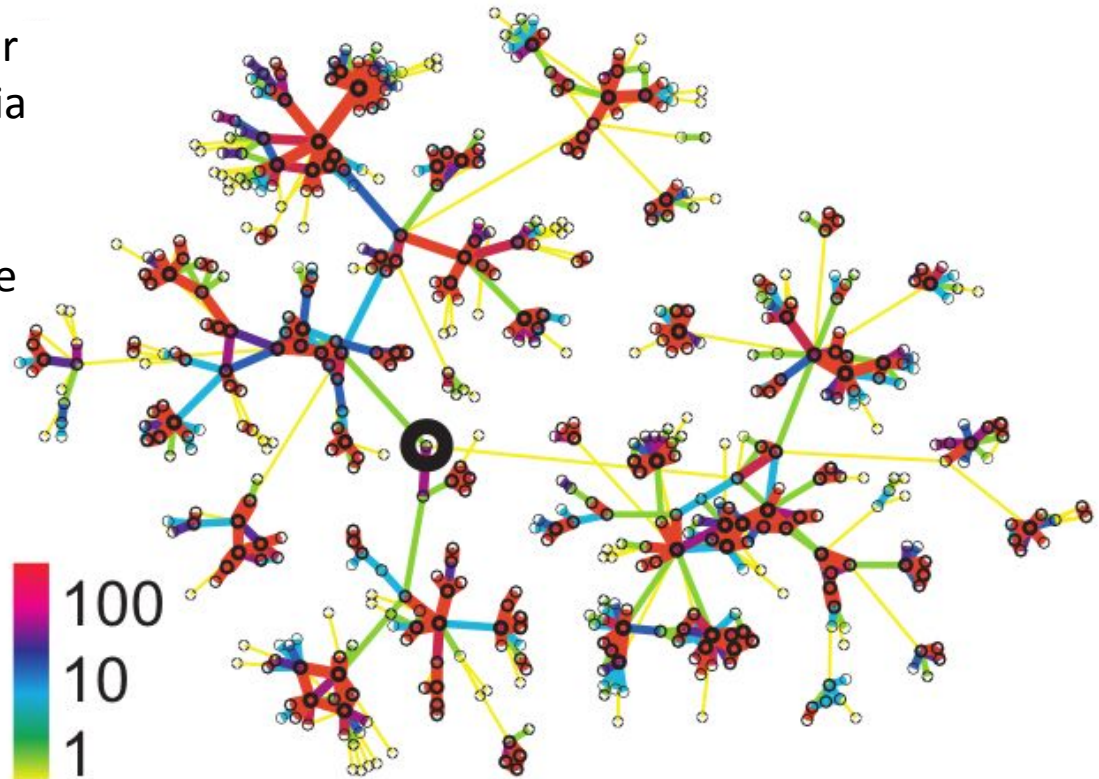
# Red ego

análisis de comunicaciones de red móvil de  $4.6 \cdot 10^6$  usuarios por 18 meses

- Red *ego* de un usuario elegido al azar
- Se incluyen nodos hasta una distancia de 6 pasos.
- En colores se representa el *campo* de intensidades sobre enlaces

$$G = G(V, E, \underbrace{f_E: E \rightarrow \mathbb{R}})$$

Intensidad o peso de un enlace



# Intensidad de enlace <-> Estructura conectividad

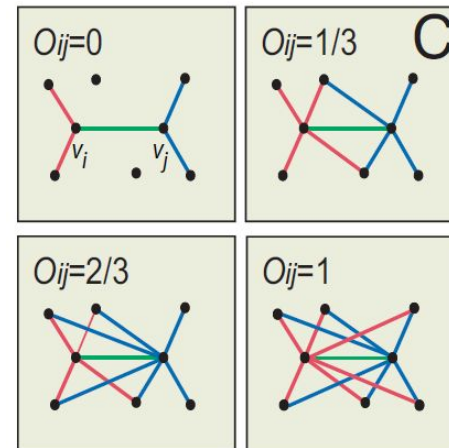
$$G = G(V, E, f_E: E \rightarrow \mathbb{R})$$

## Intensidad de un enlace

Número de llamadas  
intercambiadas

$$f_E \leftrightarrow E$$

**Overlap:** medida continua de la  
noción de **punto local**



$$O_{ij} = \frac{N(i) \cap N(j)}{N(i) \setminus j \cup N(j) \setminus i}$$



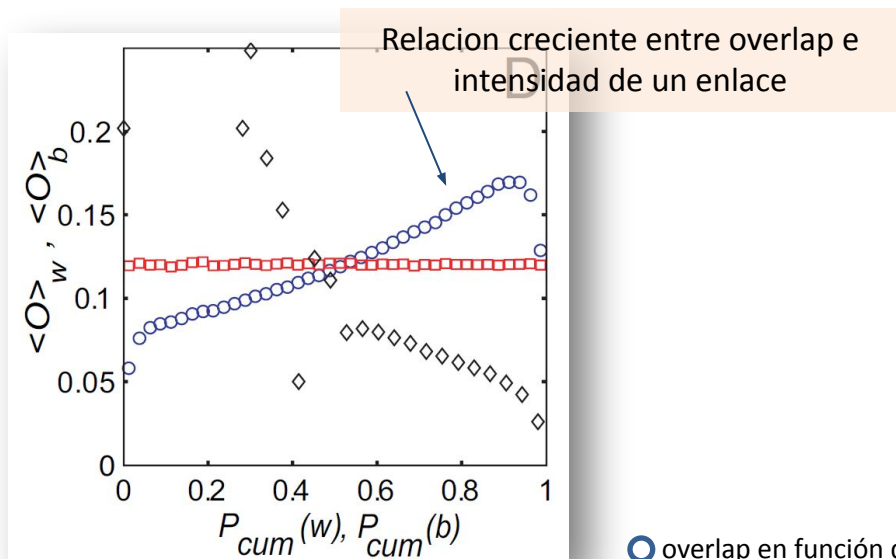
# Intensidad de enlace <-> Estructura conectividad

$$G = G(V, E, f_E: E \rightarrow \mathbb{R})$$

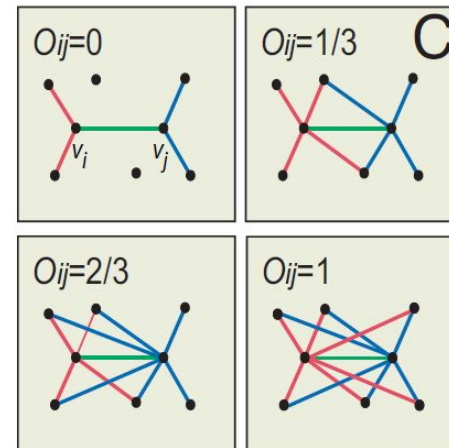
## Intensidad de un enlace

Número de llamadas  
intercambiadas

$$f_E \leftrightarrow E$$



**Overlap:** medida continua de la noción de **punto local**



$$O_{ij} = \frac{N(i) \cap N(j)}{N(i) \setminus j \cup N(j) \setminus i}$$

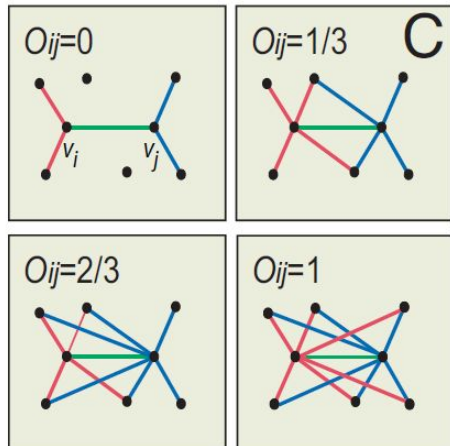
- overlap en función de fracción de intensidad acumulada
- control random
- ◇ Overlap en función de betweenness

# Intensidad de enlace <-> Estructura conectividad

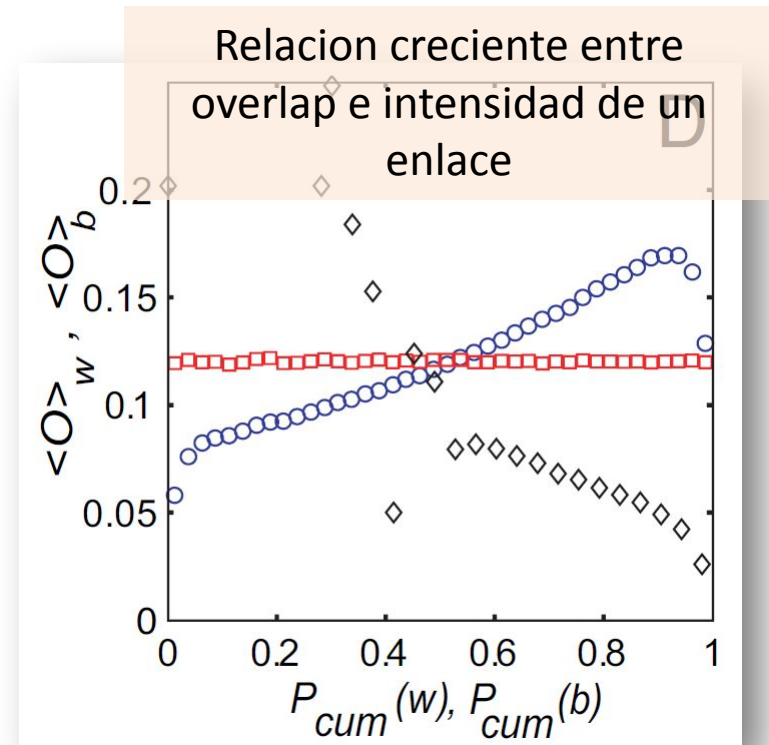
## Intensidad de un enlace

Numero de llamadas intercambiadas

**Overlap:** medida continua de la noción de **punto local**

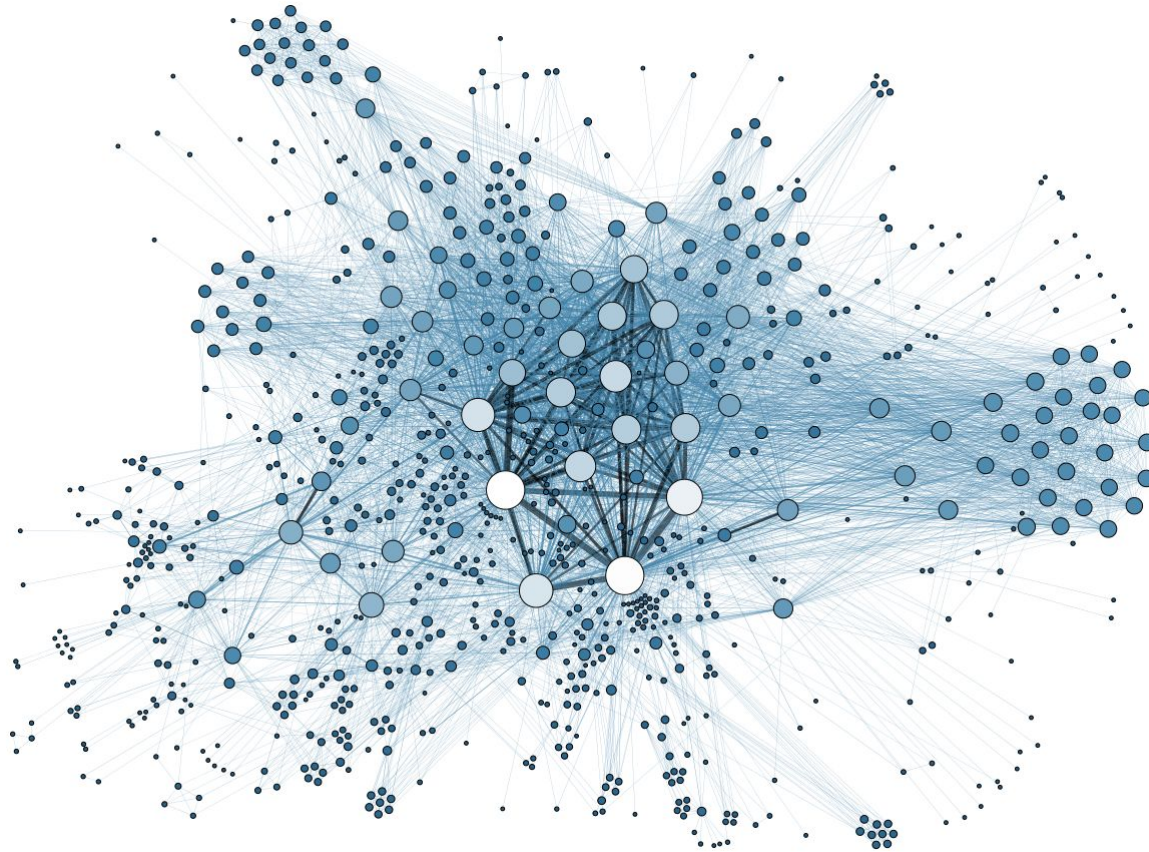


$$O_{ij} = \frac{N(i) \cap N(j)}{N(i) \setminus j \cup N(j) \setminus i}$$



- overlap en función de fracción de intensidad acumulada
- control random
- ◇ Overlap en función de betweenness

# Desarmando la **componente gigante**

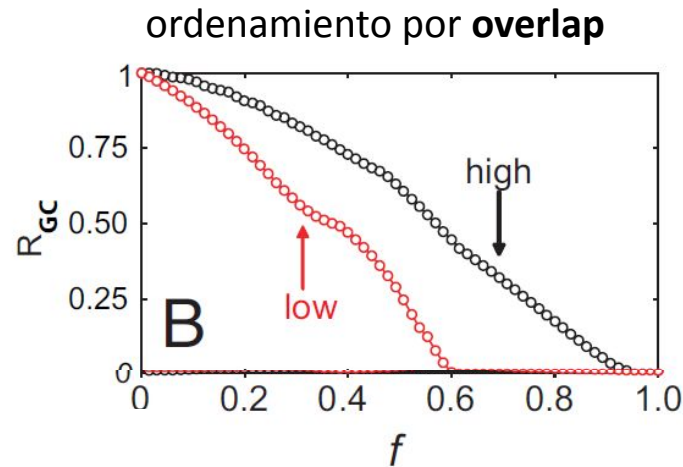
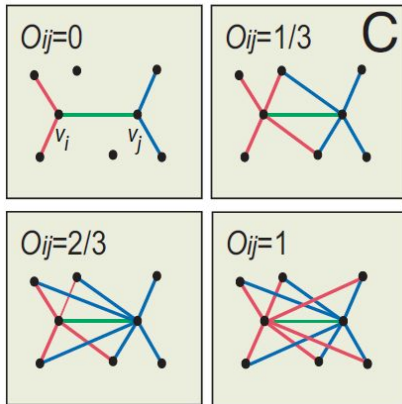


Análisis de estabilidad de la CG frente a remoción ordenada de enlaces

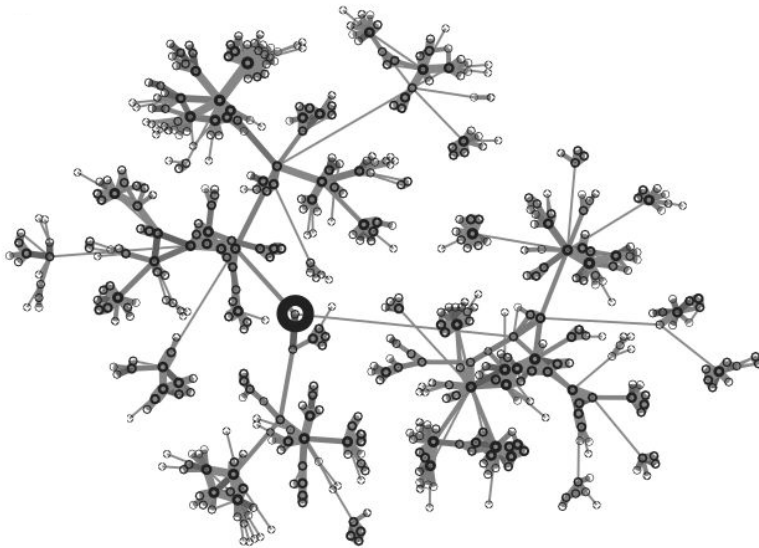
$$f_E \leftrightarrow E$$

Redes Complejas

# Desarmando la componente gigante



$R_{GC}$  tamaño relativo de la CG  
 $f$  fracción de enlaces removidos  
 • orden ascendente  
 • orden descendente

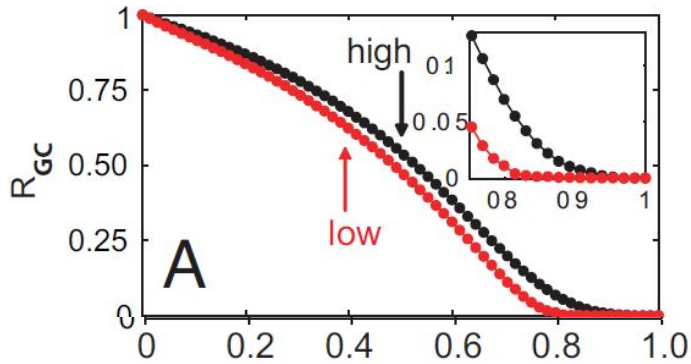


La **CG** se contrae:

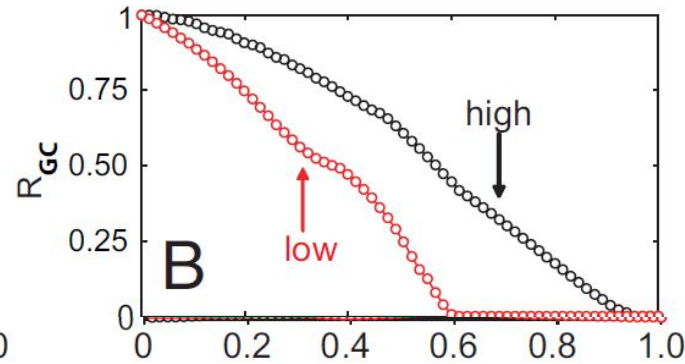
- de manera **gradual**, si se eliminan primero los **enlaces de mayor overlap**
- de manera **abrupta**, si se eliminan primero los **enlaces de menor overlap**

# Desarmando la componente gigante

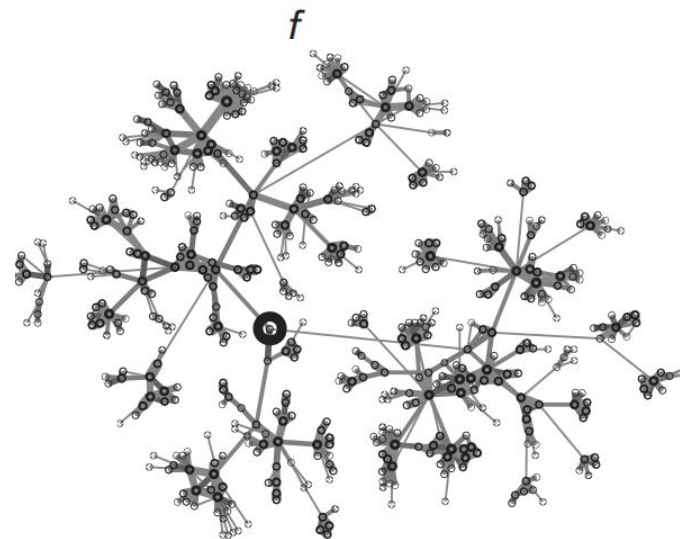
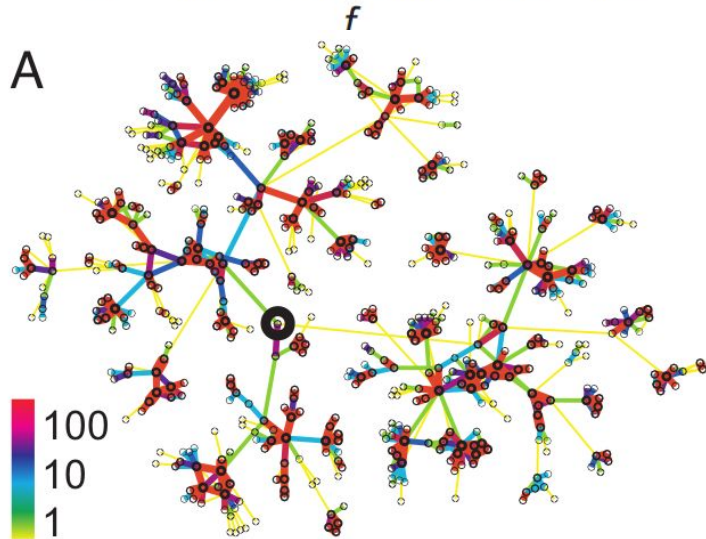
ordenamiento por **intensidad**



ordenamiento por **overlap**

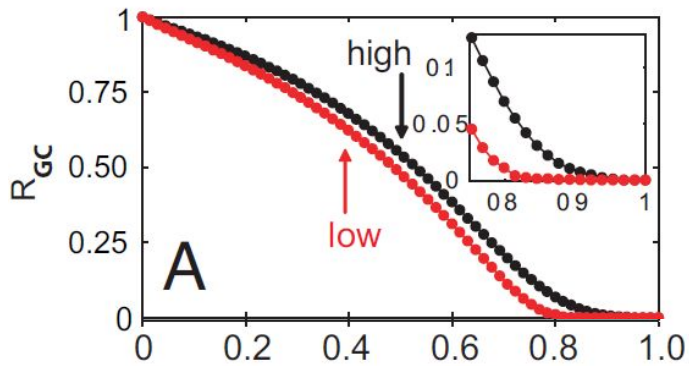


$R_{GC}$  tamaño relativo de la CG  
 $f$  fracción de enlaces removidos  
 • orden ascendente  
 • orden descendente

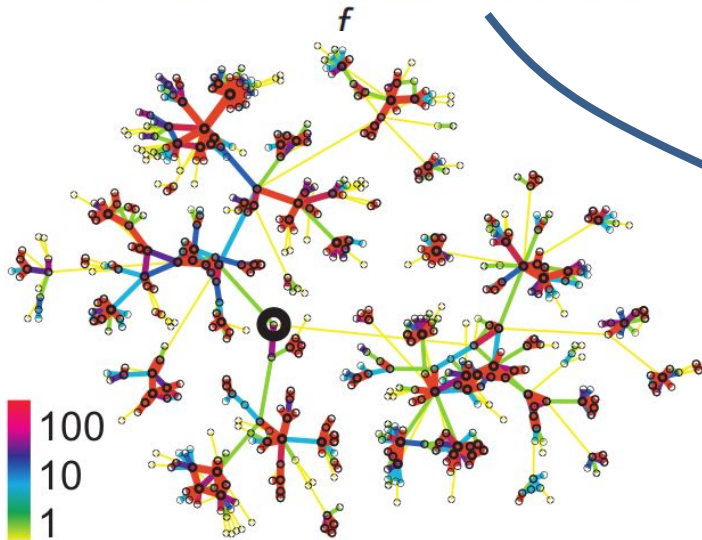


# Desarmando la componente gigante

ordenamiento por **intensidad**



- $R_{GC}$  tamaño relativo de la CG  
 $f$  fracción de enlaces removidos
- orden ascendente
  - orden descendente



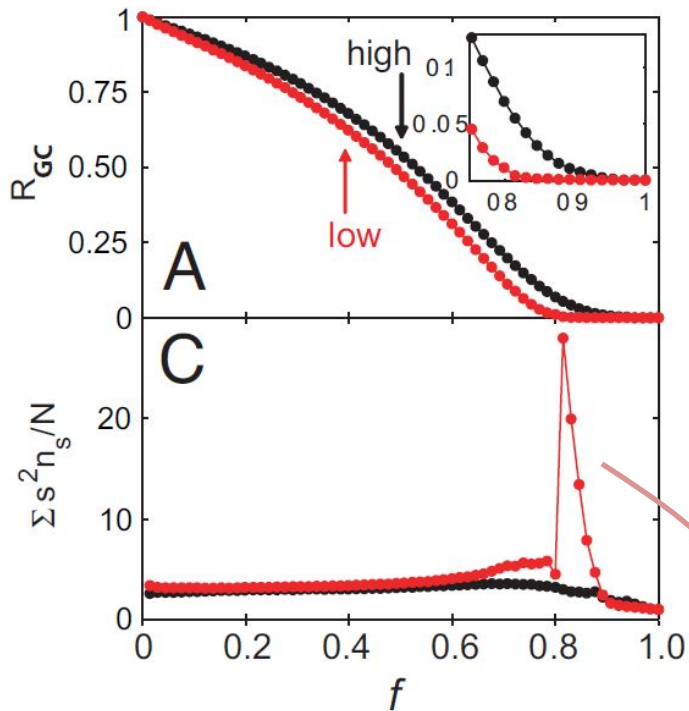
La **CG** se contrae:

- de manera **gradual**, si se eliminan primero los **enlaces** más **intensos**
- de manera **abrupta**, si se eliminan primero los **enlaces** más **débiles**

Para entender/caracterizar la transición abrupta....teoría de percolación

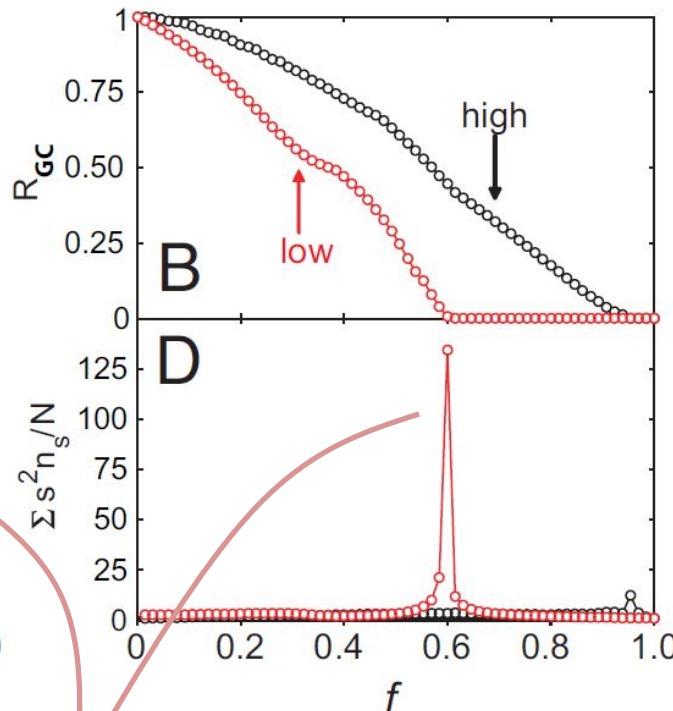
# Desarmando la componente gigante

ordenamiento por **intensidad**



Signatura de *criticalidad*  
Transición de fase de 2do orden

ordenamiento por **overlap**



$R_{GC}$  tamaño relativo de la CG  
 $f$  fracción de enlaces removidos

- orden ascendente
- orden descendente

Momento de orden 2 de la distribución de tamaños

$$\frac{1}{N} \sum s^2 n_s$$

$s$  : tamaños de componentes  
 $n_s$  : prob de ver una componen de tamaño  $s$

# Componentes: tamaños y probabilidades

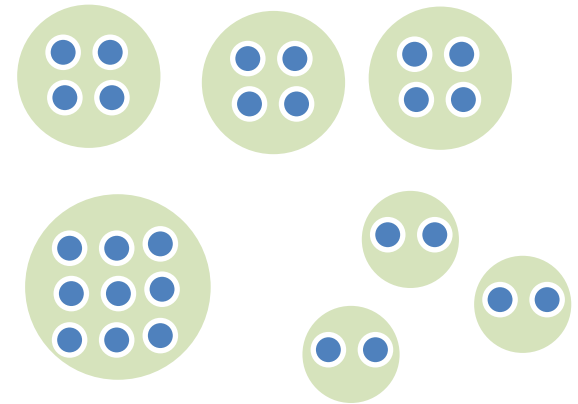
$r_s$  : Numero de componentes de tamaño  $s$

Prob de que 1 nodo tomado al azar pertenezca a una componente de tamaño  $s$

$$p_s = \frac{r_s \cdot s}{N} \quad \left( p_2 = \frac{6}{27} \right)$$

Prob de que una componente tenga tamaño  $s$ ,  $\pi_s$ , tiene que ser proporcional a  $p_s$

$$\pi_s = \frac{\frac{r_s \cdot s}{N}}{\sum_s \frac{r_s \cdot s}{N}} = \frac{r_s \cdot s}{1} = \frac{r_s \cdot s}{N} = p_s$$



Tamaño medio de un cluster eligiendo un nodo al azar

$$\bar{s} = \sum_s \pi_s \cdot s = \sum_s \frac{r_s \cdot s}{N} \cdot s = \frac{1}{N} \sum_s s^2 r_s$$

No me interesa caracterizar esta cantidad para una realizacion particular sino en un sentido estadístico ...



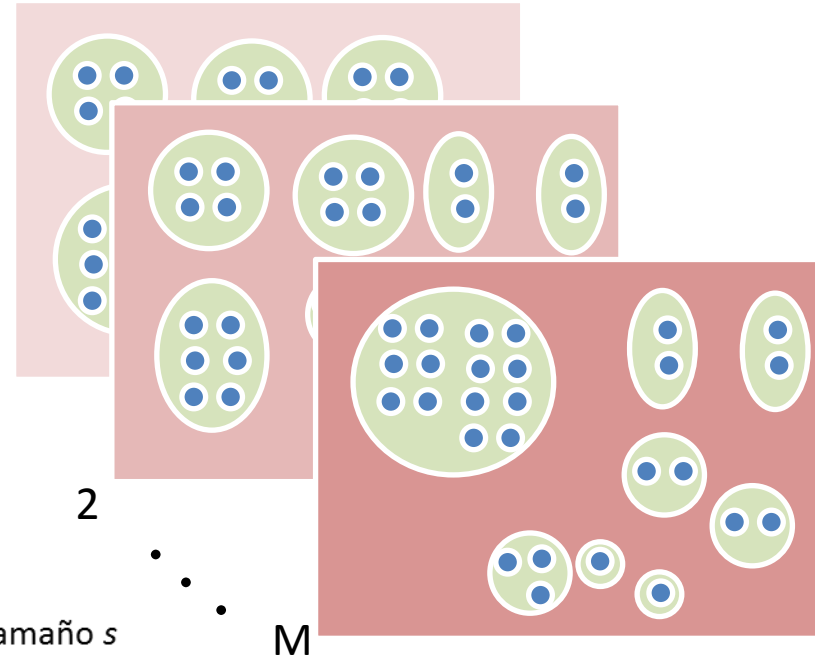
# Componentes: tamaños y probabilidades

$$\bar{S} = \sum_s \pi_s \cdot s = \frac{1}{N} \sum s^2 r_s$$

$$\langle S \rangle = \frac{1}{M} \sum_{i=1}^M \left[ \frac{1}{N} \sum_{s=1}^{s_{max}} s^2 r_s^i \right] = \frac{1}{M \cdot N} \sum_{s=1}^{s_{max}} \sum_{i=1}^M s^2 r_s^i$$

$$= \frac{1}{N} \sum_{s=1}^{s_{max}} s^2 \frac{\sum_{i=1}^M r_s^i}{M}$$

$n_s$  ← Prob de ver una comp de tamaño  $s$  en las  $M$  realizaciones ( $n_s$ )



$$\langle S \rangle = \frac{1}{N} \sum_{s=1}^{s_{max}} s^2 n_s$$

# Componentes: tamaños y probabilidades

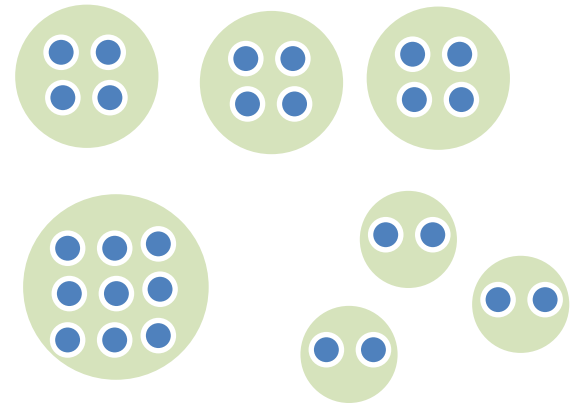
Prob de que 1 nodo, tomado al azar, pertenezca a una componente de tamaño  $s$

$r_s$  : Numero de componentes de tamaño  $s$

$$p_s = \frac{r_s \cdot s}{N} \quad \left( p_2 = \frac{6}{27} \right)$$

Tamaño medio de un cluster eligiendo un nodo al azar

$$\bar{s} = \sum_s \frac{r_s \cdot s}{N} \cdot s = \frac{1}{N} \sum_s s^2 r_s$$



Interesa caracterizar esta cantidad en un sentido estadístico (e.g. desarmo a la CG con misma receta pero ligeras variaciones, debido a empates de los pesos por ejemplo)

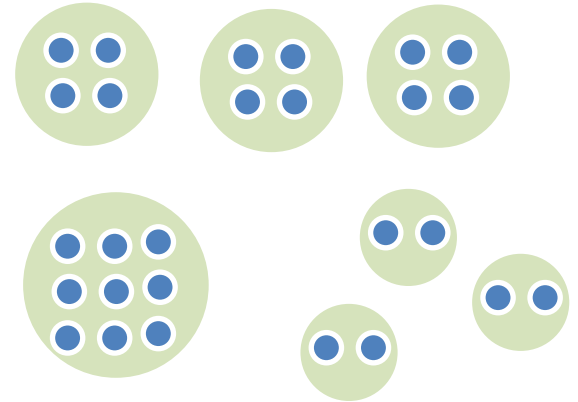
# Componentes: tamaños y probabilidades

$r_s$  : Numero de componentes de tamaño  $s$

Prob de que 1 nodo tomado al azar pertenezca a una componente de tamaño  $s$

$$p_s = \frac{r_s \cdot s}{N} \quad \left( p_2 = \frac{6}{27} \right)$$

$$p_s = \frac{r_s s}{N} = \pi_s \leftarrow \text{Prob de que una componente tenga tamaño } s, \pi_s, \text{ tiene que ser proporcional a } p_s$$



Tamaño medio de un cluster eligiendo un nodo al azar

$$\bar{s} = \sum_s \pi_s \cdot s = \sum_s \frac{r_s \cdot s}{N} \cdot s = \frac{1}{N} \sum_s s^2 r_s$$

Interesa caracterizar esta cantidad en un sentido estadístico ...

# Componentes: tamaños y probabilidades

$$\bar{S} = \frac{1}{N} \sum s^2 r_s$$

propiedad del  
ensemble estadístico

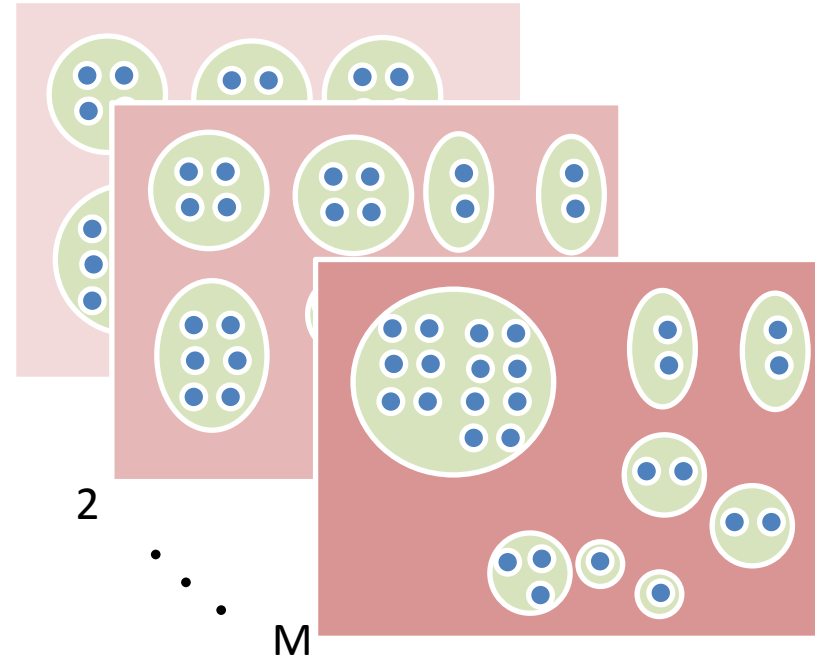
$$\langle S \rangle = \frac{1}{M} \sum_{i=1}^M \left[ \frac{1}{N} \sum_{s=1}^{s_{max}} s^2 r_s^i \right] = \frac{1}{M \cdot N} \sum_{s=1}^{s_{max}} \sum_{i=1}^M s^2 r_s^i$$

$$= \frac{1}{N} \sum_{s=1}^{s_{max}} s^2 \frac{\sum_{i=1}^M r_s^i}{M}$$

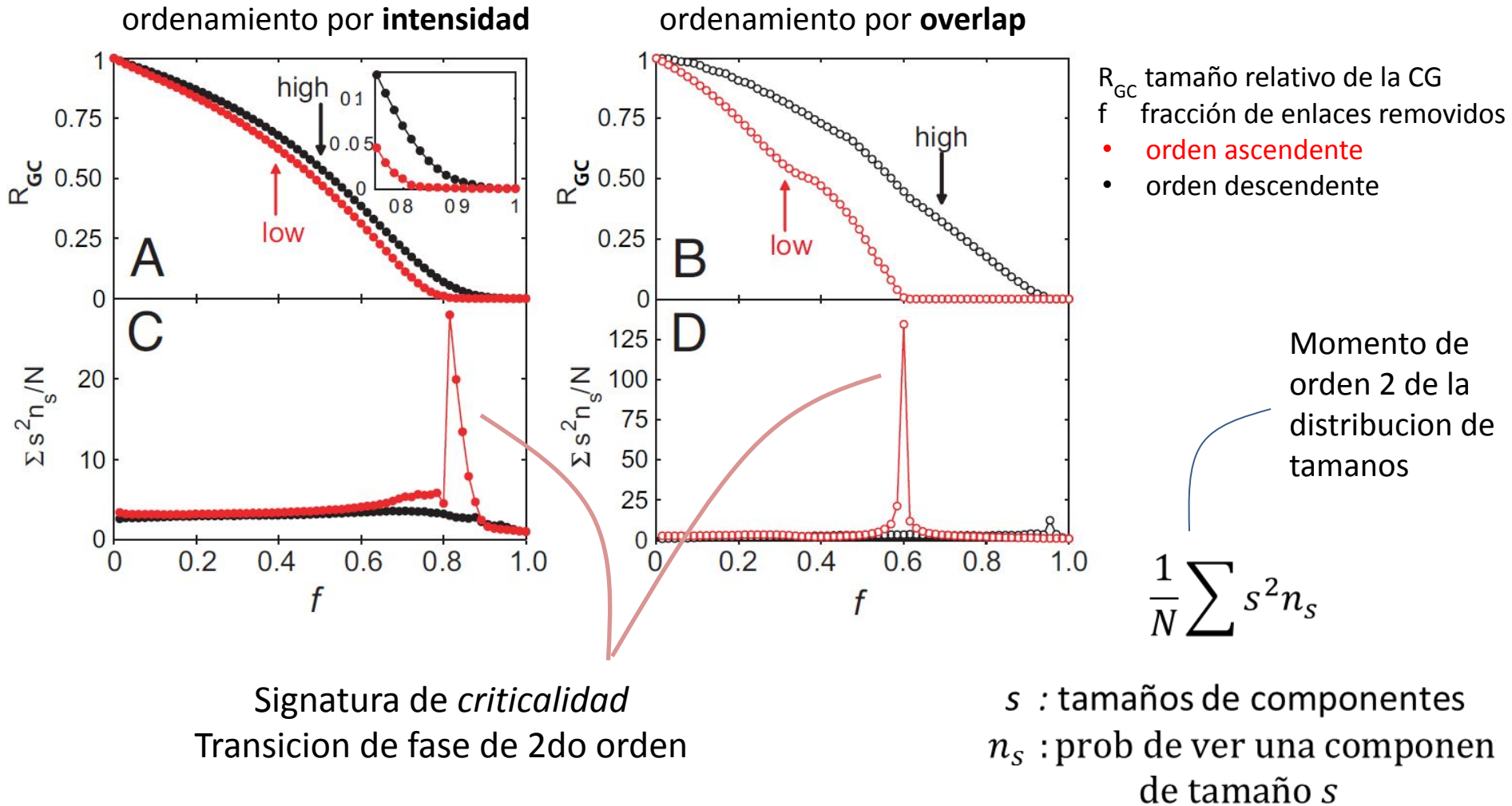
Prob de ver una comp de tamaño  $s$   
en las  $M$  realizaciones ( $n_s$ )

$$\langle S \rangle = \frac{1}{N} \sum_{s=1}^{s_{max}} s^2 n_s$$

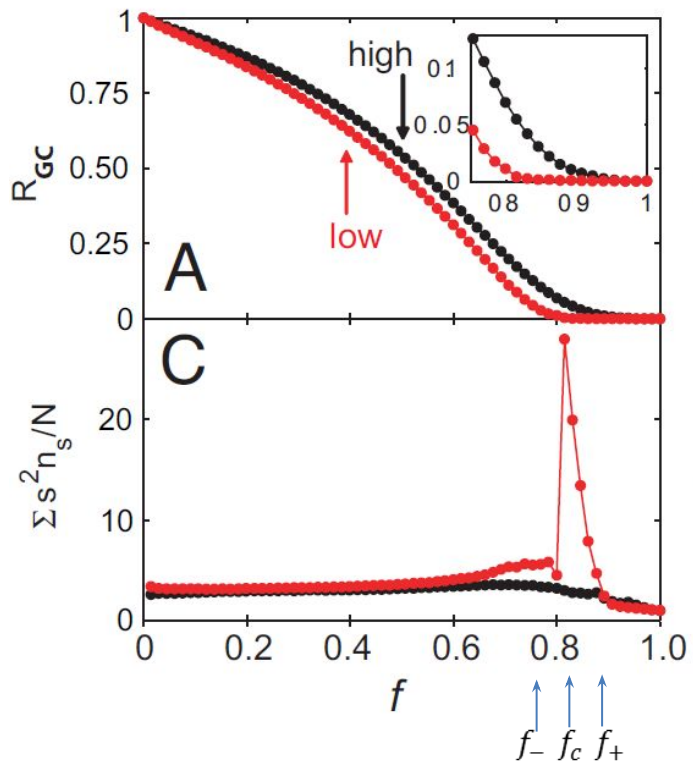
El tamaño medio de clusters  
es el segundo momento de la  
distribución de clusters



# Desarmando la componente gigante



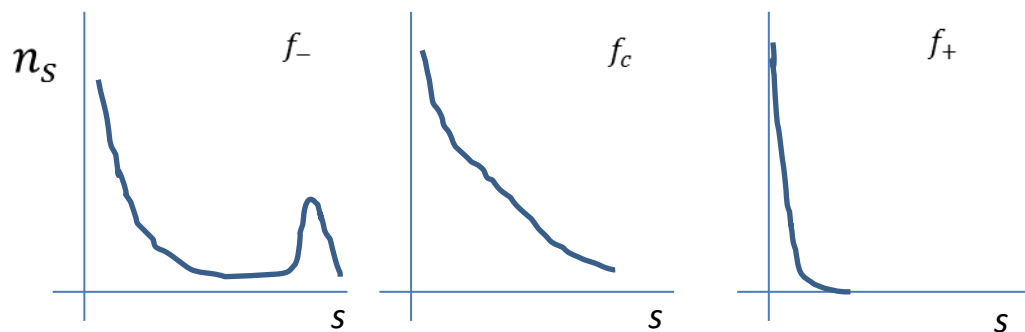
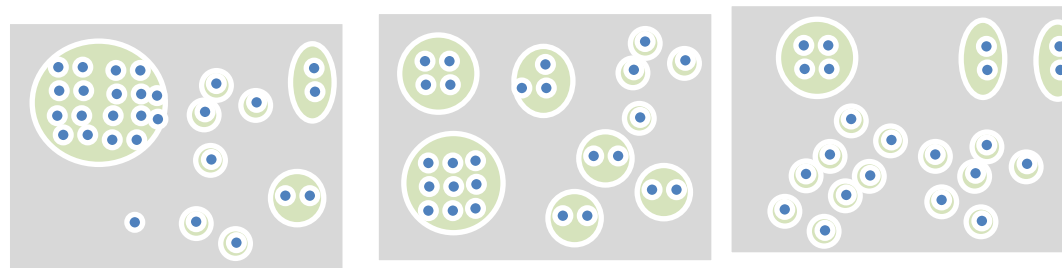
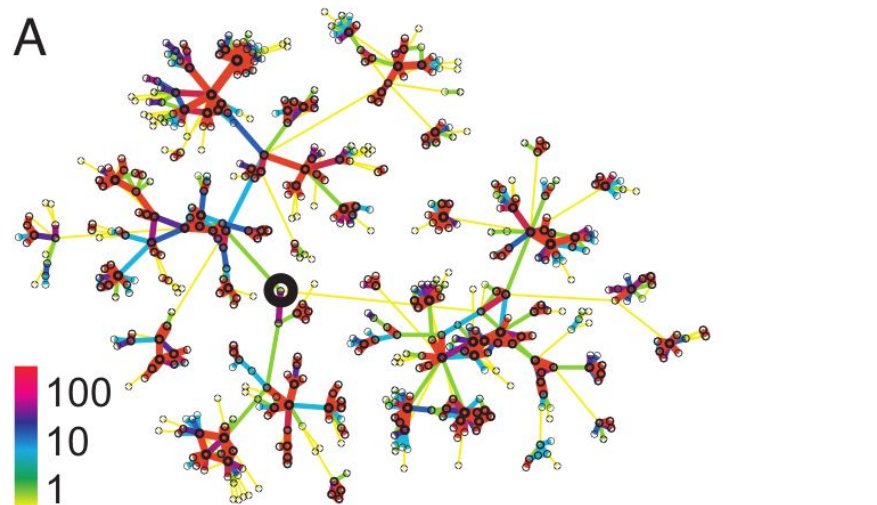
# La transición



Que significa que  $\langle S \rangle$  diverja?

$$\langle S \rangle = \frac{1}{N} \sum_{s=1}^{s_{max}} s^2 n_s$$

el 2do momento de la distribución de tamaños diverge



$$n_s \sim s^{-a}$$

# Referencias

- *Networks Crowds and Market*, Easley & Kleinberg Chap 2
- “Chains of affection: The structure of adolescent romantic and sexual networks”, Bearman *et al* (2004) *Am. Jour. Soc.*
- “The small world problem” , Stanley Milgram (1967) *Psychology Today*
- “Planetary-scale views of instant-messaging network”, Leskovec (2007)
- <https://oracleofbacon.org>
  
- “Getting a Job. A study of contacts and careers” Mark Grannoveter (1974)
- “Suicide and Friendships Among American Adolescents”, Bearman & Moody (2004) *Am. J. Public Health*
- “Structure and tie strengths in mobile communication networks”, Onnela *et al* (2007) *PNAS*