

Análisis de redes sociales

¿Cómo identificar personas significativos en una red social y analizar su papel?

¿Qué características globales de la estructura de las redes interesa medir?

¿Es posible explicar analíticamente la estructura y formación de redes sociales?

Análisis de redes sociales

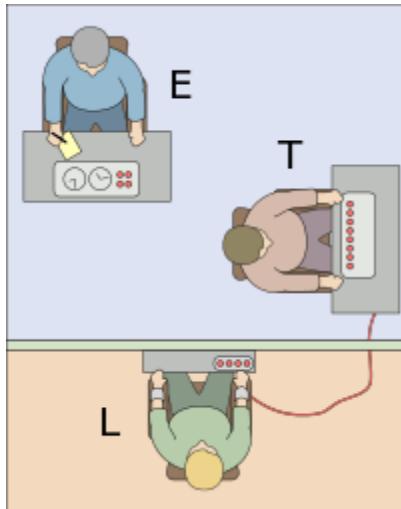
1. Introducción
2. Topologías de red: métricas y estadísticas
3. Subdivisión de redes
4. Modelos de formación de redes
5. Procesos de difusión

1. Introducción

“The” Milgram experiment (1963)



Stanley Milgram
(1933-1984)



¿Estamos todos conectados?

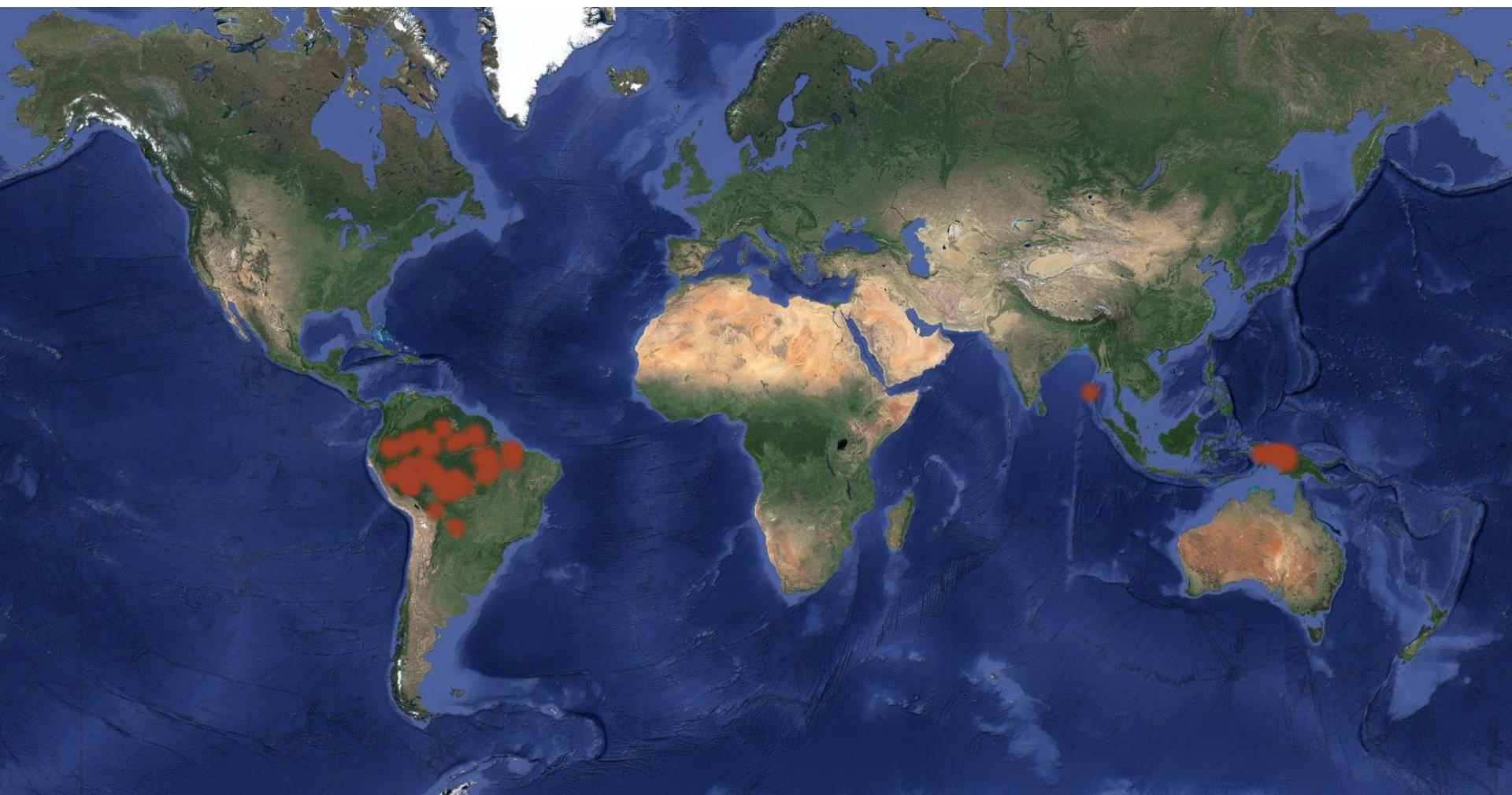
Sentinel Island



Acre, Brasil

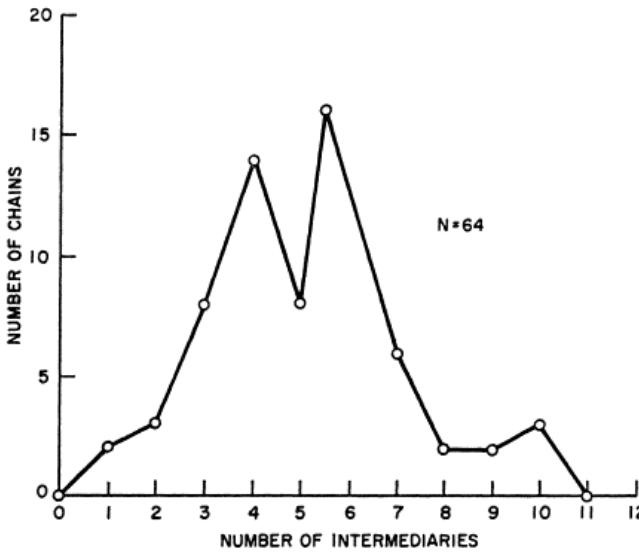


¿Estamos todos conectados?



Experimento small-world Milgram (1967)

¿Cuál es la distancia promedio entre dos personas al azar?

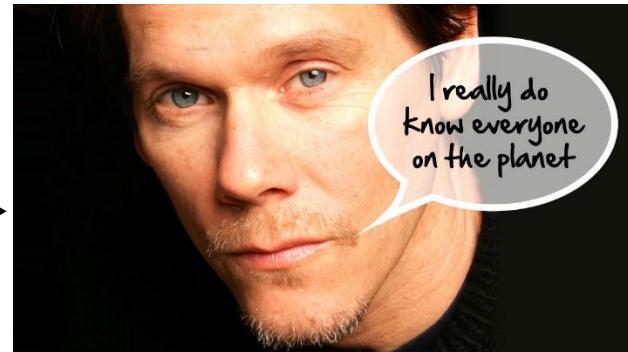


- ◆ Varias personas origen elegidas “al azar”, una persona destino
- ◆ Tarea: hacer llegar un paquete a la persona destino por medio de una cadena de amigos
 - Cada persona añade su nombre en una lista + notifica a Milgram
- ◆ 64/296 cadenas llegaron al destino
- ◆ Resultado: distancia promedio ~ 6.2 (nº arcos)

“Six degrees of separation” en la cultura popular



Erdős nr = 4.65 promedio
entre matemáticos en activo



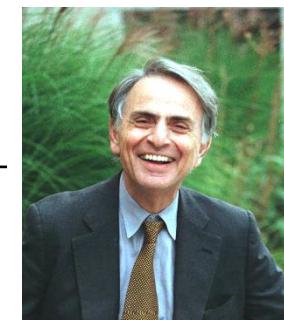
Bacon nr, avg ~ 3, máx = 10
oracleofbacon.org

4



2

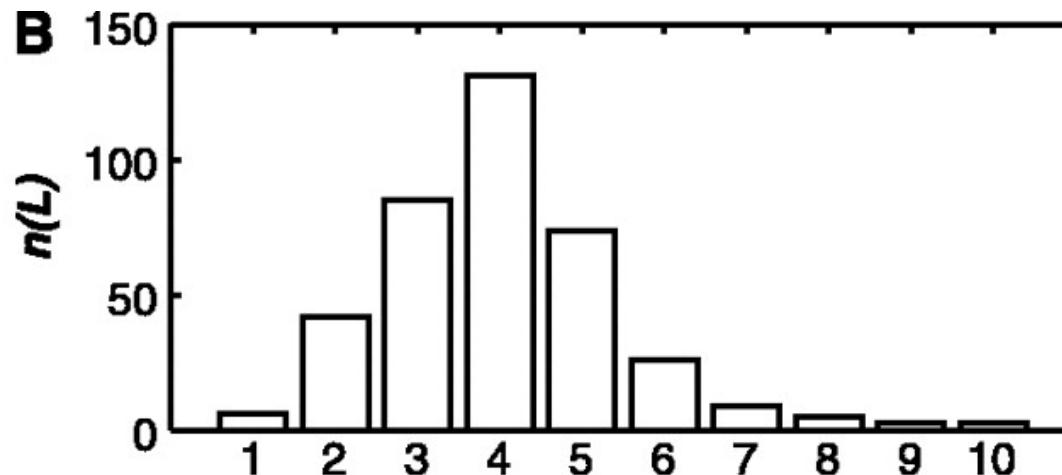
4



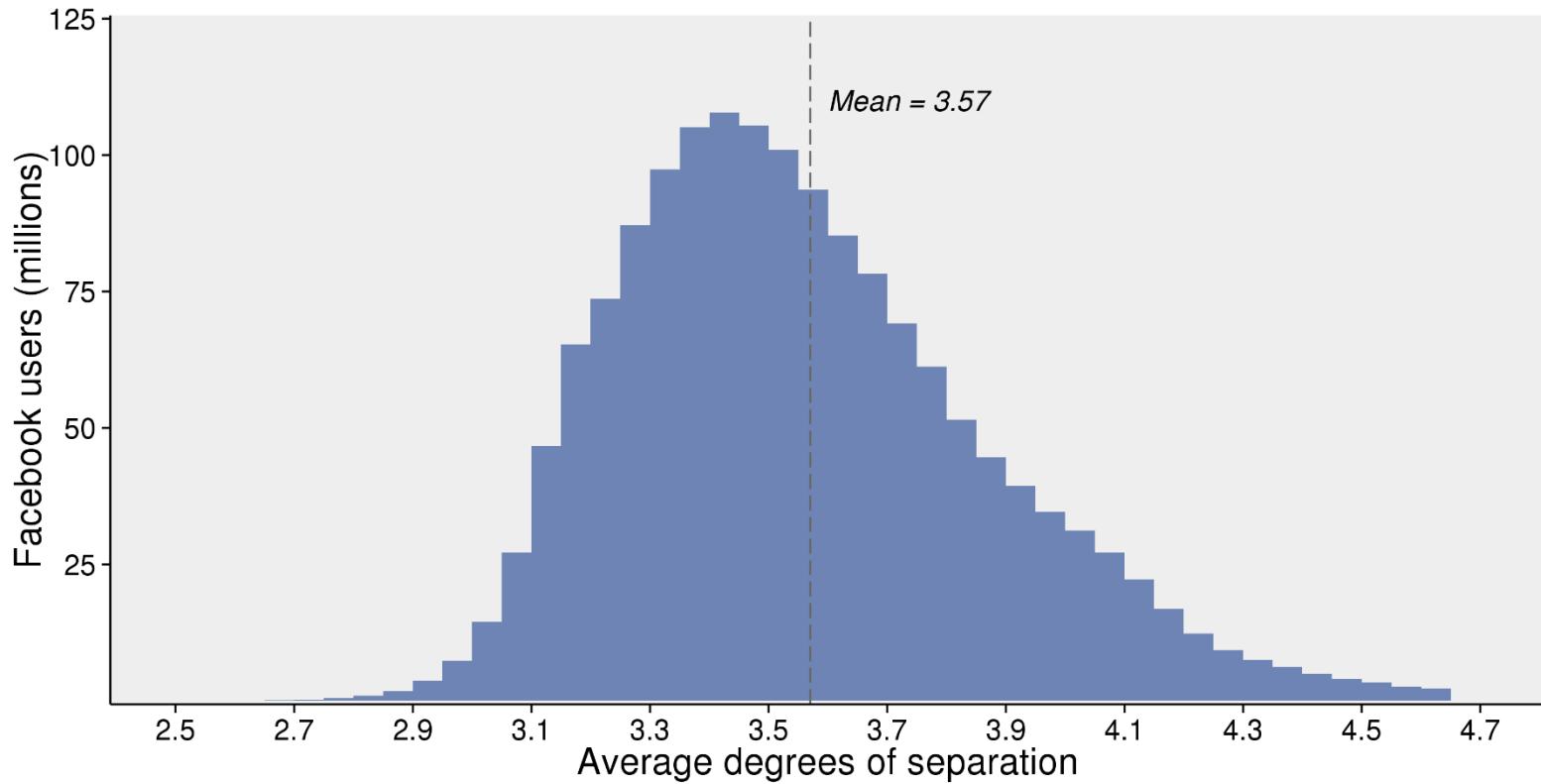
2

Experimentos posteriores

- ◆ P.e. Dodds, Muhamad & Watts 2003 repitieron el experimento a gran escala
- ◆ 18 personas objetivo en 13 países diferentes puntos del globo
- ◆ 24.163 cadenas de email, 384 llegan al destino
- ◆ Distancia promedio $\sim 4\text{-}6$



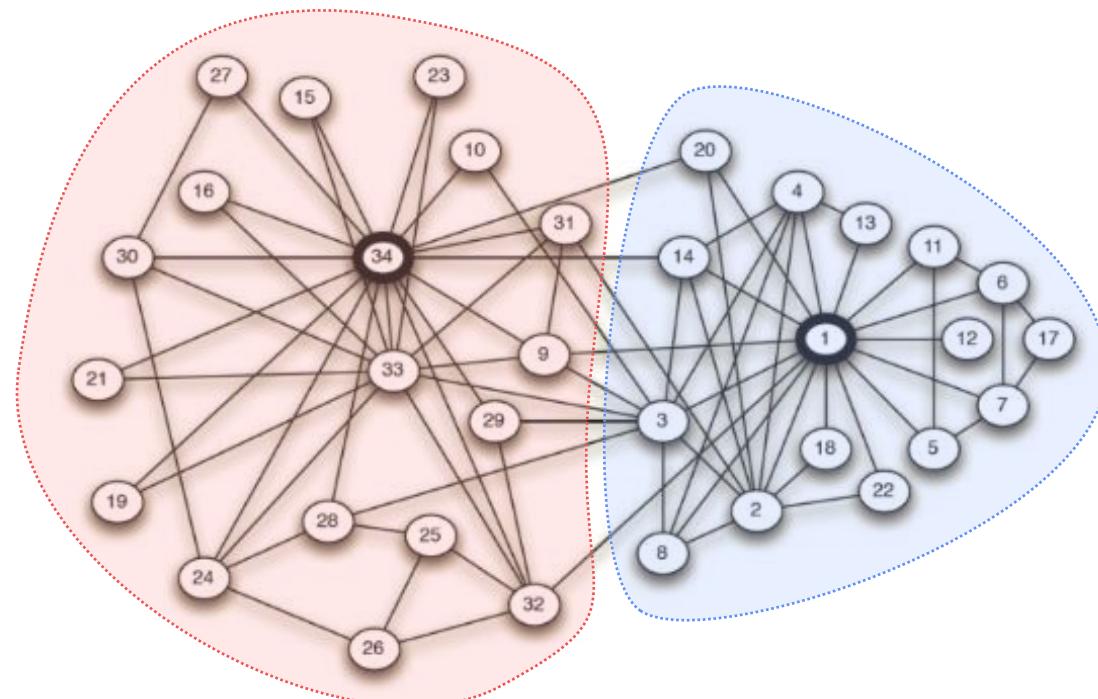
Small world en Facebook

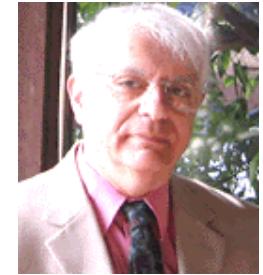


<https://research.facebook.com/blog/three-and-a-half-degrees-of-separation> (Feb 2016)

Zachary's karate club

- ◆ W. W. Zachary, 1970
- ◆ 34 miembros de un club de karate
- ◆ 78 relaciones de amistad fuera de las clases
- ◆ Ruptura de pequeños grupos: maximum flow – minimum cut
Ford–Fulkerson algorithm from 1 to 34

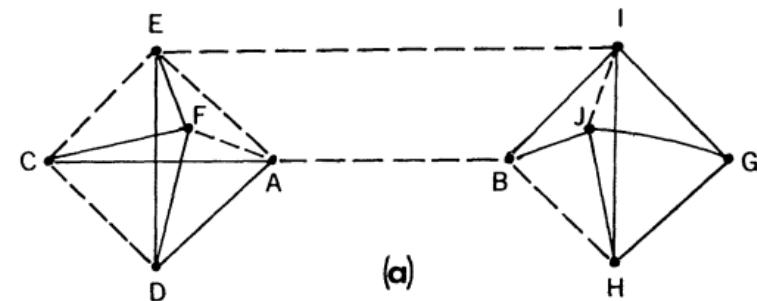




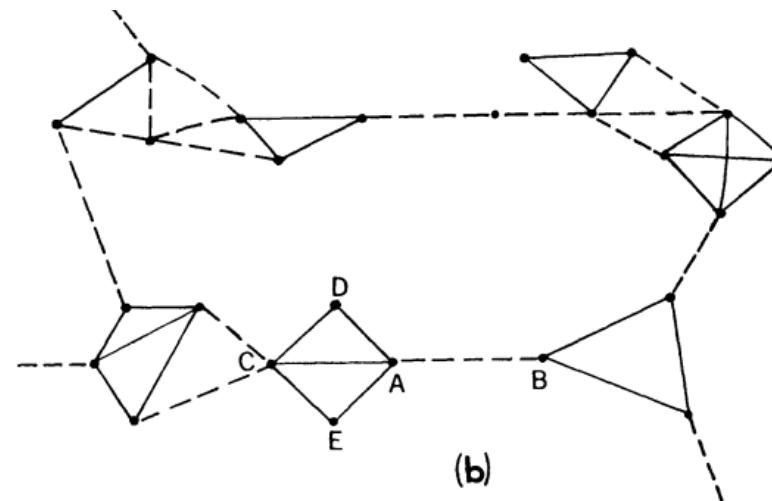
The strength of weak ties

◆ M. S. Granovetter, 1973

Mark S. Granovetter
(1943-)



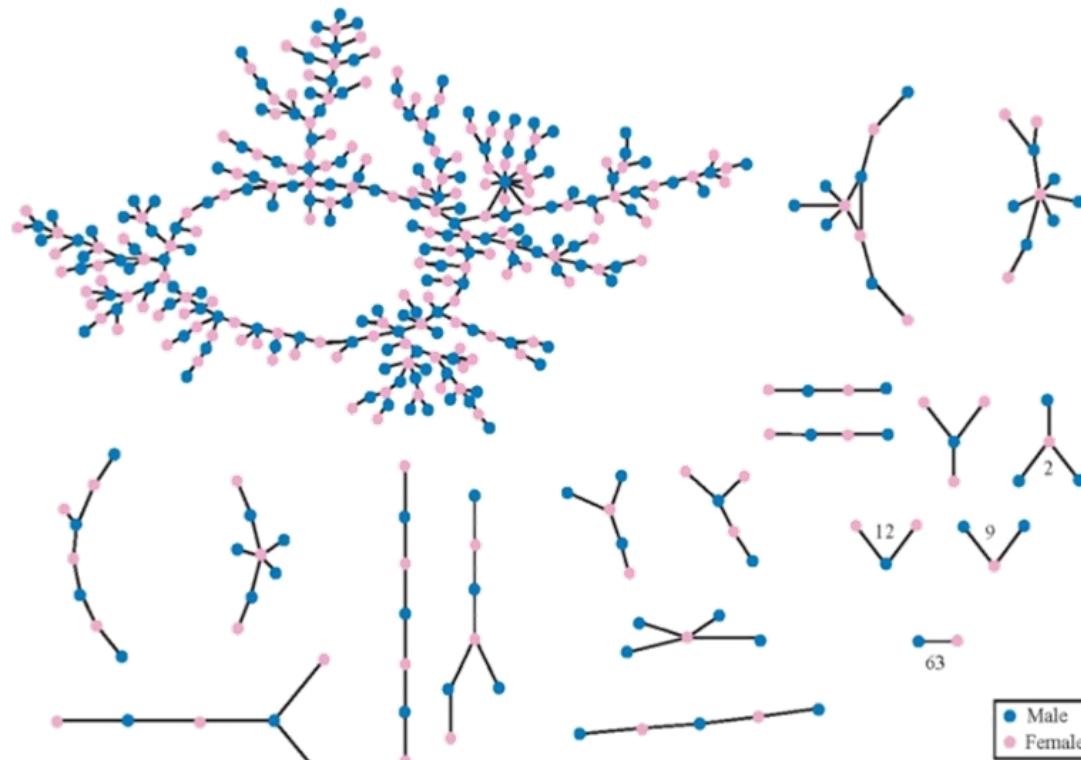
(a)



(b)

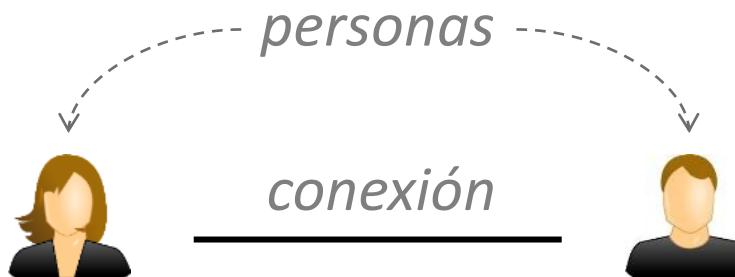
Estudio “Jefferson High”

- ◆ P. Bearman, J. Moody and K. Stovel, 1993
- ◆ 573 estudiantes ◆ Grado promedio 1.66
- ◆ ASP ~ 16, heterofilia, ciclos largos, bajo coeficiente de clustering
- ◆ Enfermedades de transmisión sexual



¿Qué es una red social?

La pieza básica:

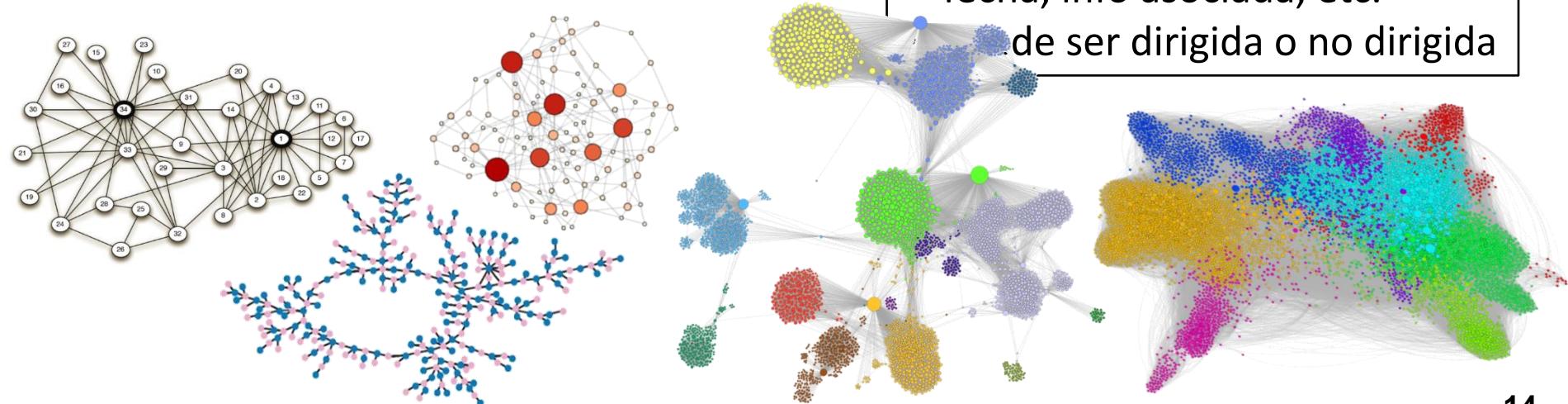


En principio personas, pero se puede generalizar a organizaciones, países, etc.

- Puede reflejar diferentes hechos: amistad, colaboración, comunicación, compartir, etc.
- Puede tener estructura: tipo, fecha, info asociada, etc.

de ser dirigida o no dirigida

Formación de estructuras complejas



Redes sociales

- ◆ Son un caso particular de los fenómenos de conectividad (*netwok science*)
 - Mucho en común con otras redes: de comunicación (transporte, Internet, etc.), relaciones y cadenas comerciales (países, empresas, organizaciones...), redes biológicas (cadena trófica, neuronas, interacción celular, etc.), interacciones químicas, grafos léxicos, la Web, etc.
- ◆ Las redes sociales pueden ser de muy diverso tipo
 - Amistad, colaboración, afiliación, parentesco, cita bibliográfica, interacción (email, teléfono, retweet, etc.)
 - Las redes online son un subconjunto visible de las redes completas
- ◆ Un campo multidisciplinar
 - Sociólogos, antropólogos, economistas, matemáticos, físicos, informáticos...

Breve perspectiva histórica

- ◆ Se empieza a hablar de redes sociales a finales del XIX
 - ◆ Primeros estudios en los 30 (sociogramas)
 - ◆ Formalización matemática en los 50
 - ◆ Amplio desarrollo de teorías y métodos en los 80
 - ◆ Trabajo de los físicos en los 90 (importantes avances teóricos)
 - ◆ Redes online en los 2000!
- 
- Ciencias sociales

“In the absence of
actual network data,
all this is speculation”



M. S. Granovetter. The strength of weak ties.
American Journal of Sociology 78(6), 1973

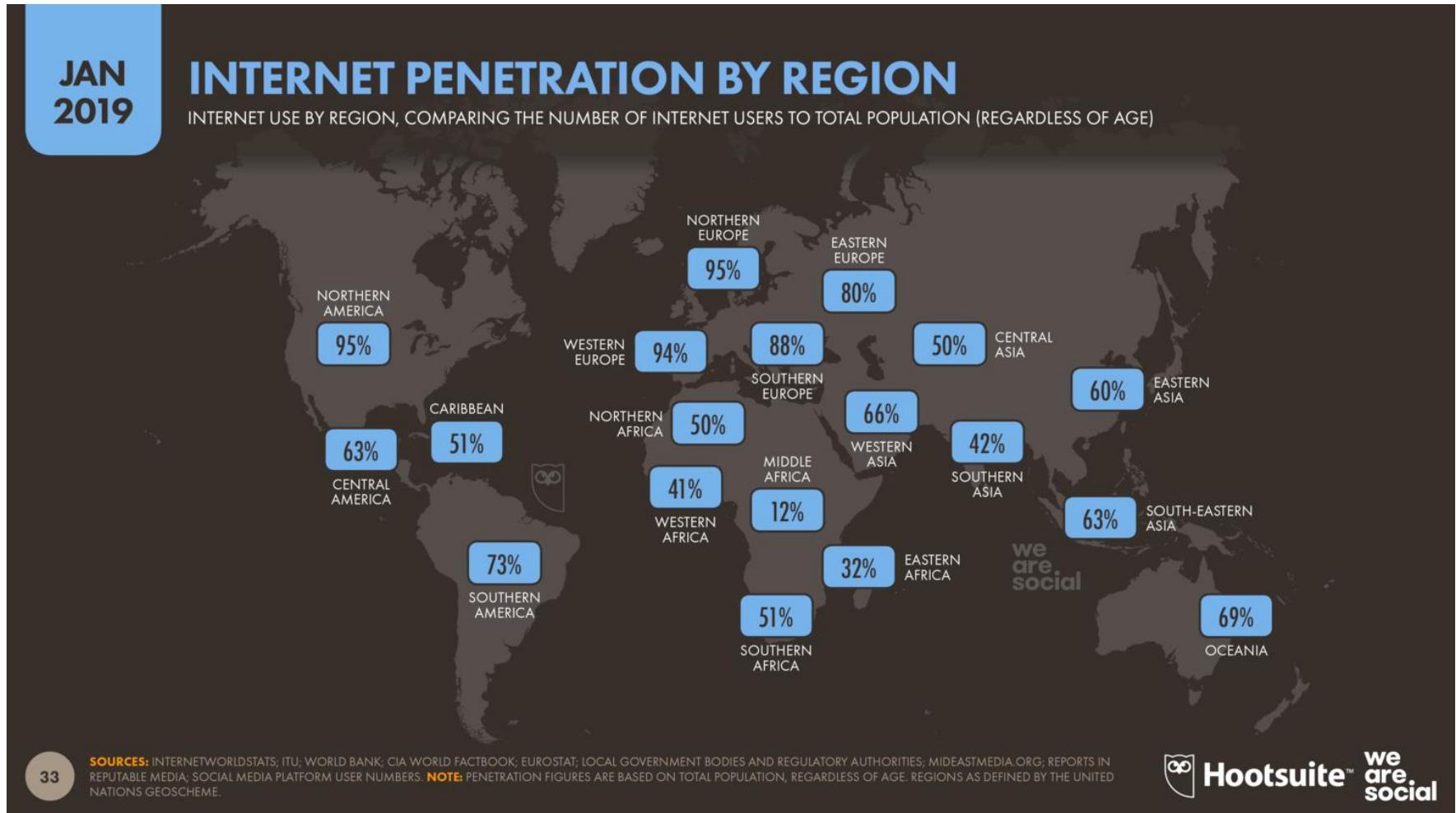
Redes sociales online

- ◆ Las redes online en los 2000 dan lugar a un boom de actividad en este campo
 - SixDegrees en 1997, Friendster en 2002, MySpace y LinkedIn en 2003, Orkut y Facebook en 2004, Twitter en 2006, Google+ en 2011...
- ◆ Tendencia a la integración de la red en plataformas de servicio

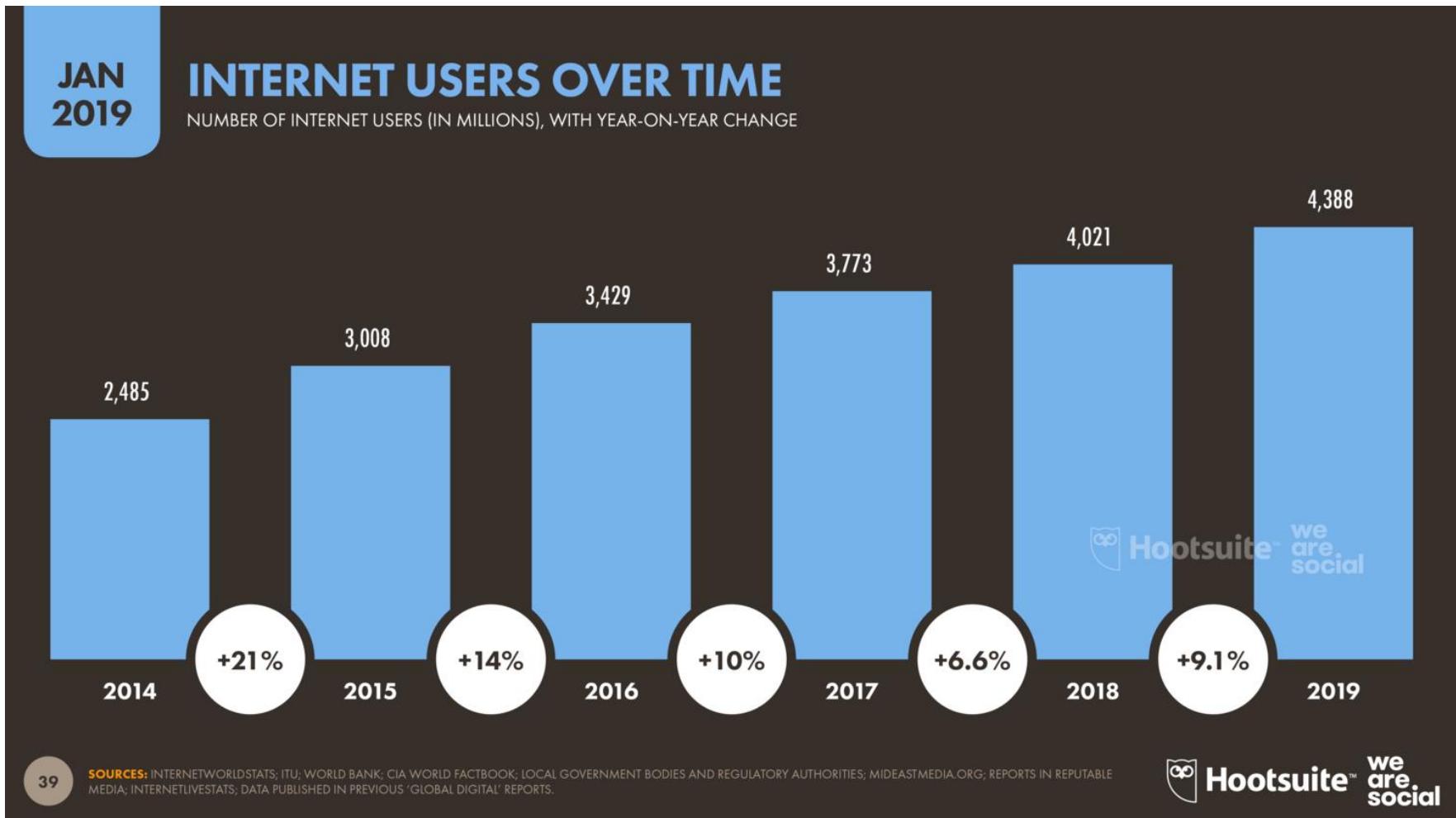
Cifras globales



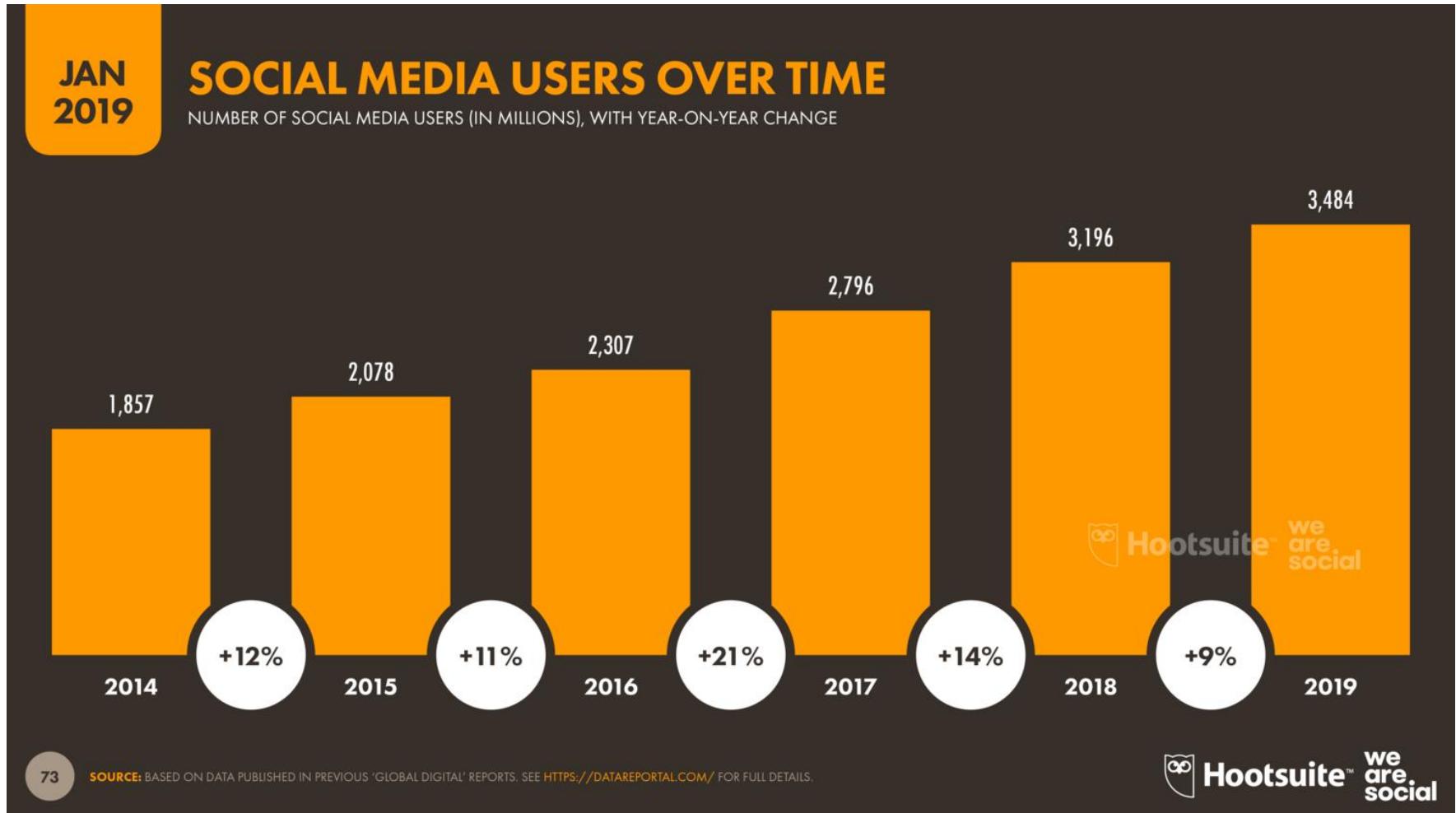
Cifras globales



Cifras globales



Cifras globales



Cifras globales

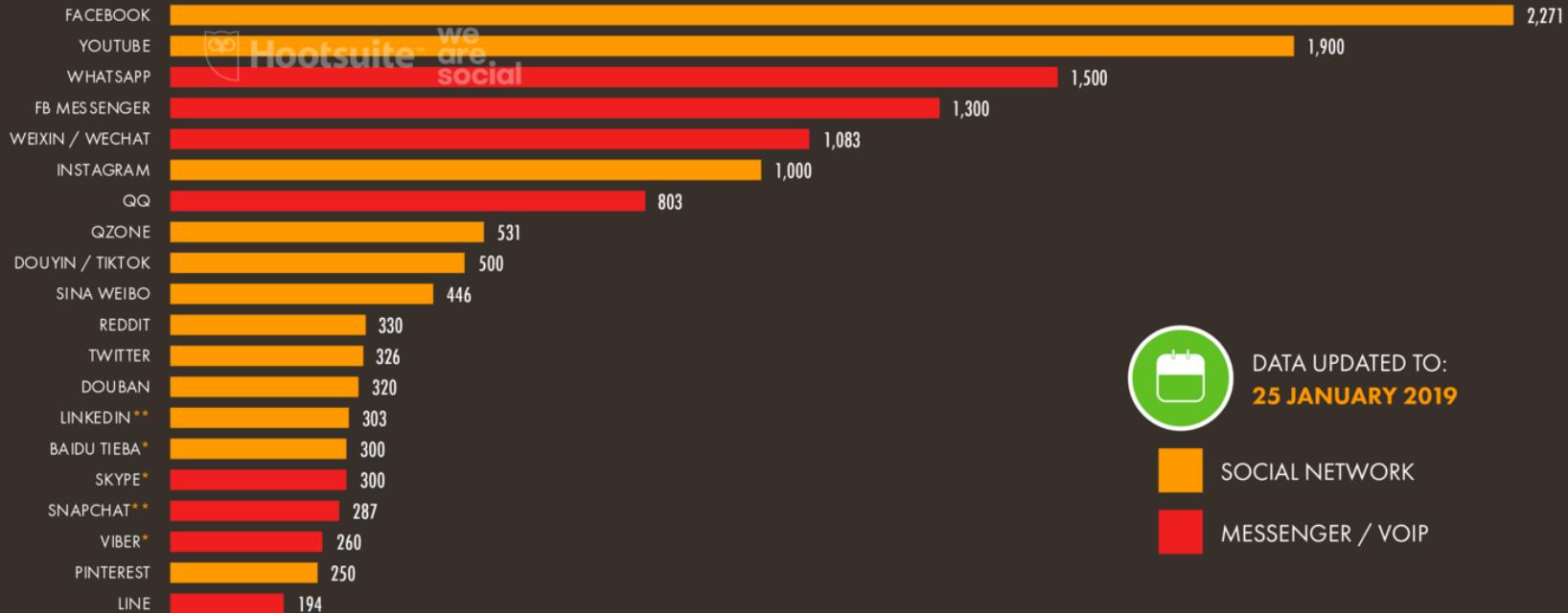


Cifras globales

JAN
2019

SOCIAL PLATFORMS: ACTIVE USER ACCOUNTS

BASED ON MONTHLY ACTIVE USERS, USER ACCOUNTS, OR UNIQUE VISITORS TO EACH PLATFORM, IN MILLIONS



DATA UPDATED TO:
25 JANUARY 2019

SOCIAL NETWORK

MESSENGER / VOIP

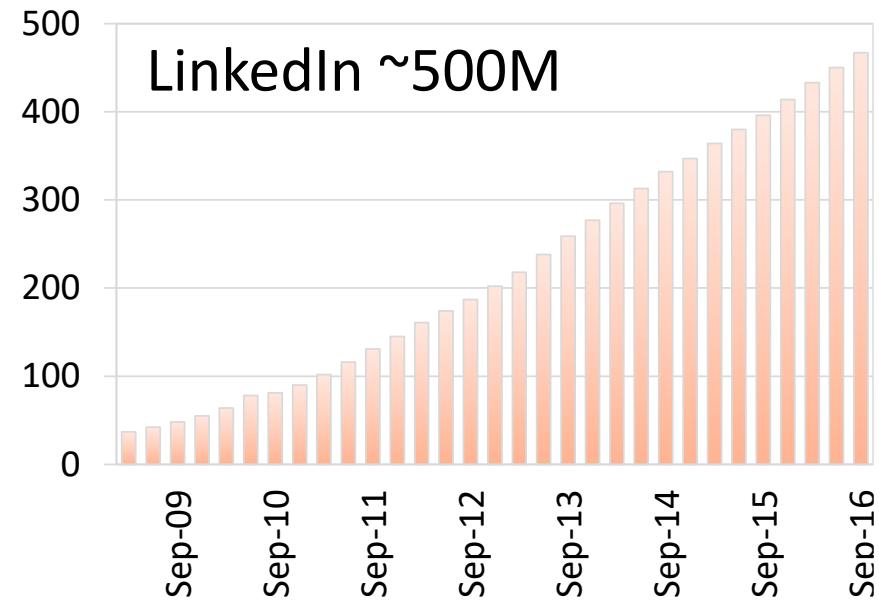
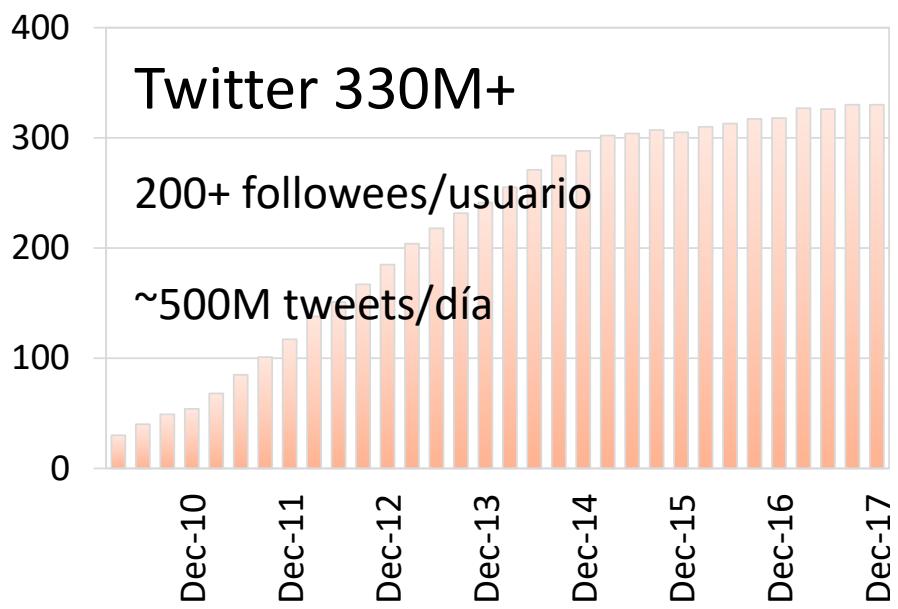
81

SOURCES: KEPIOS ANALYSIS; LATEST COMPANY EARNINGS RELEASES, PRESS RELEASES OR MEDIA STATEMENTS; REPORTS IN REPUTABLE MEDIA (ALL UP TO JAN 2019). *ADVISORY: PLATFORMS IDENTIFIED BY (*) HAVE NOT PUBLISHED UPDATED USER FIGURES IN THE PAST 12 MONTHS, SO FIGURES MAY BE LESS RELIABLE. **NOTES: THESE PLATFORMS DO NOT PUBLISH MAU DATA. LINKEDIN FIGURE IS BASED ON MONTHLY UNIQUE WEBSITE VISITORS IN DEC 2018, VIA SIMILARWEB. SNAPCHAT FIGURE EXTRAPOLATED FROM DATA REPORTED IN TECHCRUNCH (JUN 2017).

Hootsuite™ we
are social

Crecimiento de las redes sociales online

- ◆ Cientos de sitios (Wikipedia enumera 14 con 100M+ usuarios)
- ◆ Miles de millones de usuarios activos

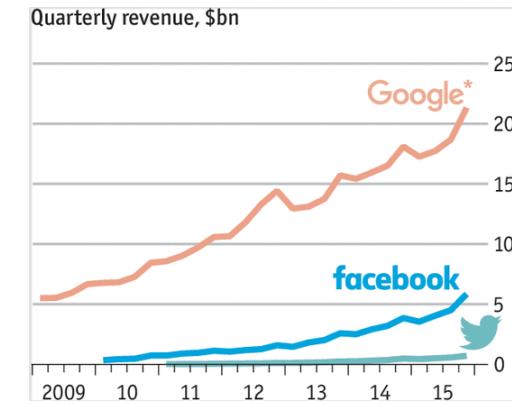


Volumen de negocio en las redes sociales

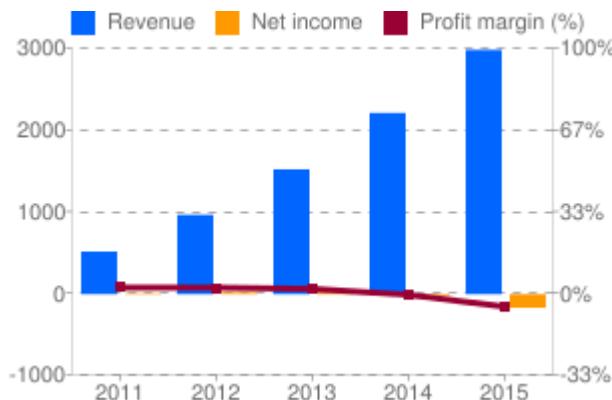
Facebook (17K+ empleados)



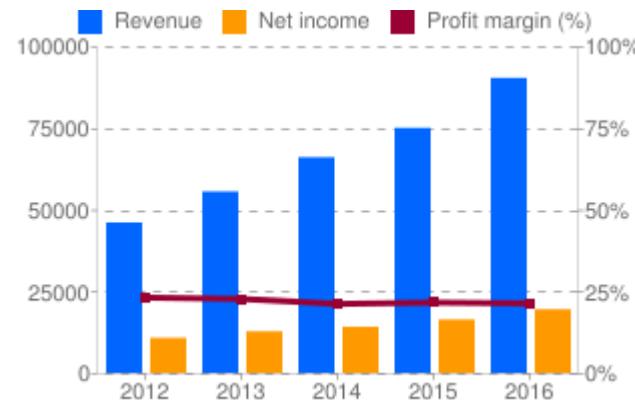
Twitter (~3.9 empleados)



LinkedIn (9.7K+ empleados)

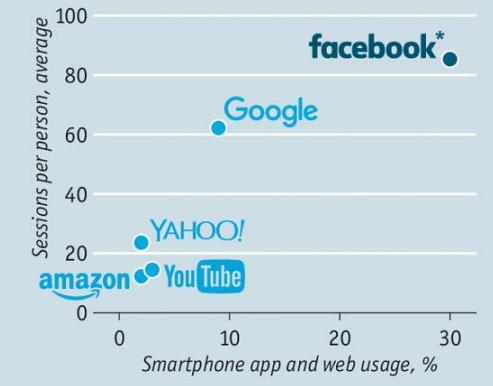


Google (70K+ empleados)



Tip-top tap

Smartphone usage in America, aged over 18
December 2015

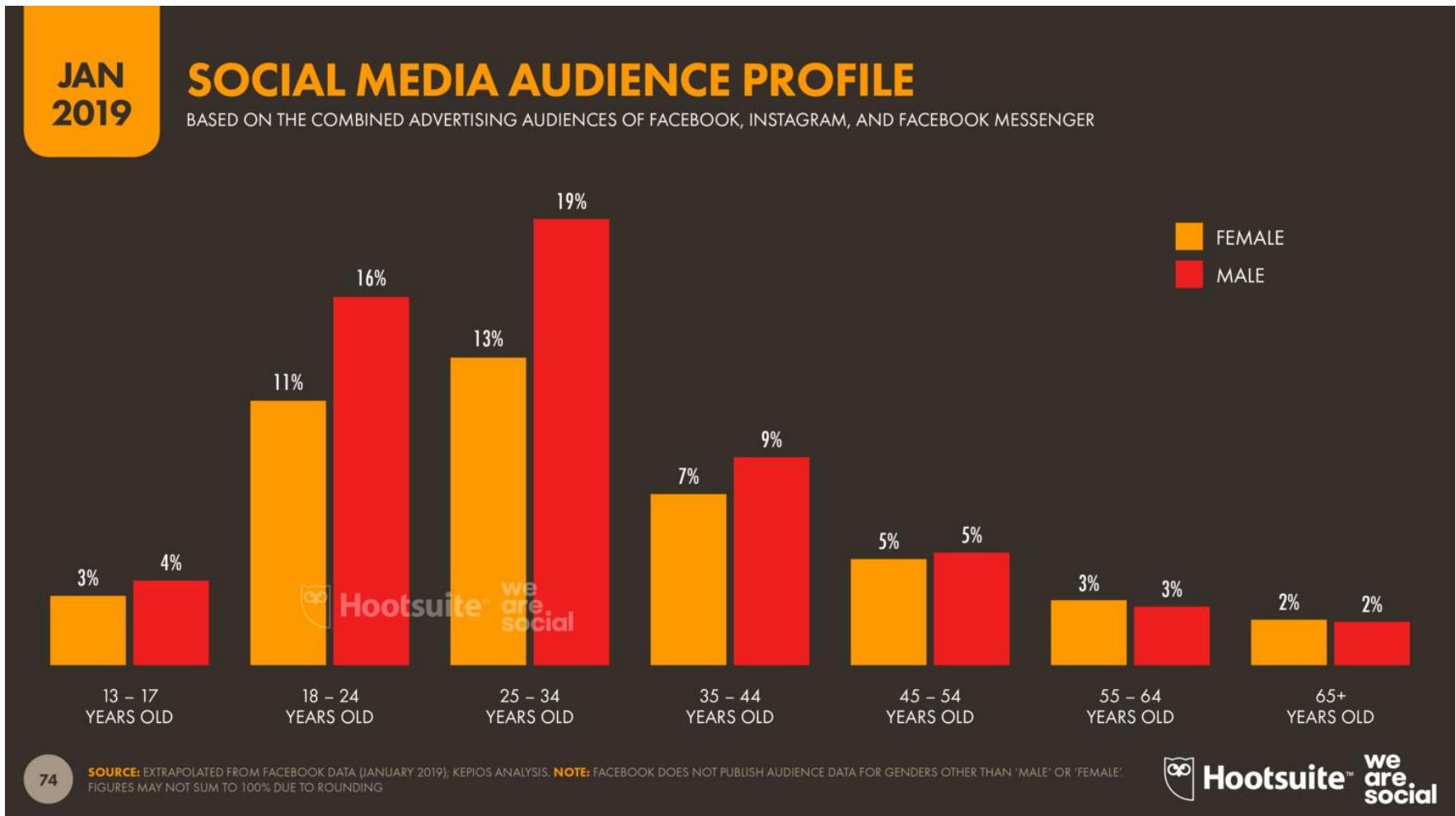


Source: Nielsen *Including Instagram and WhatsApp

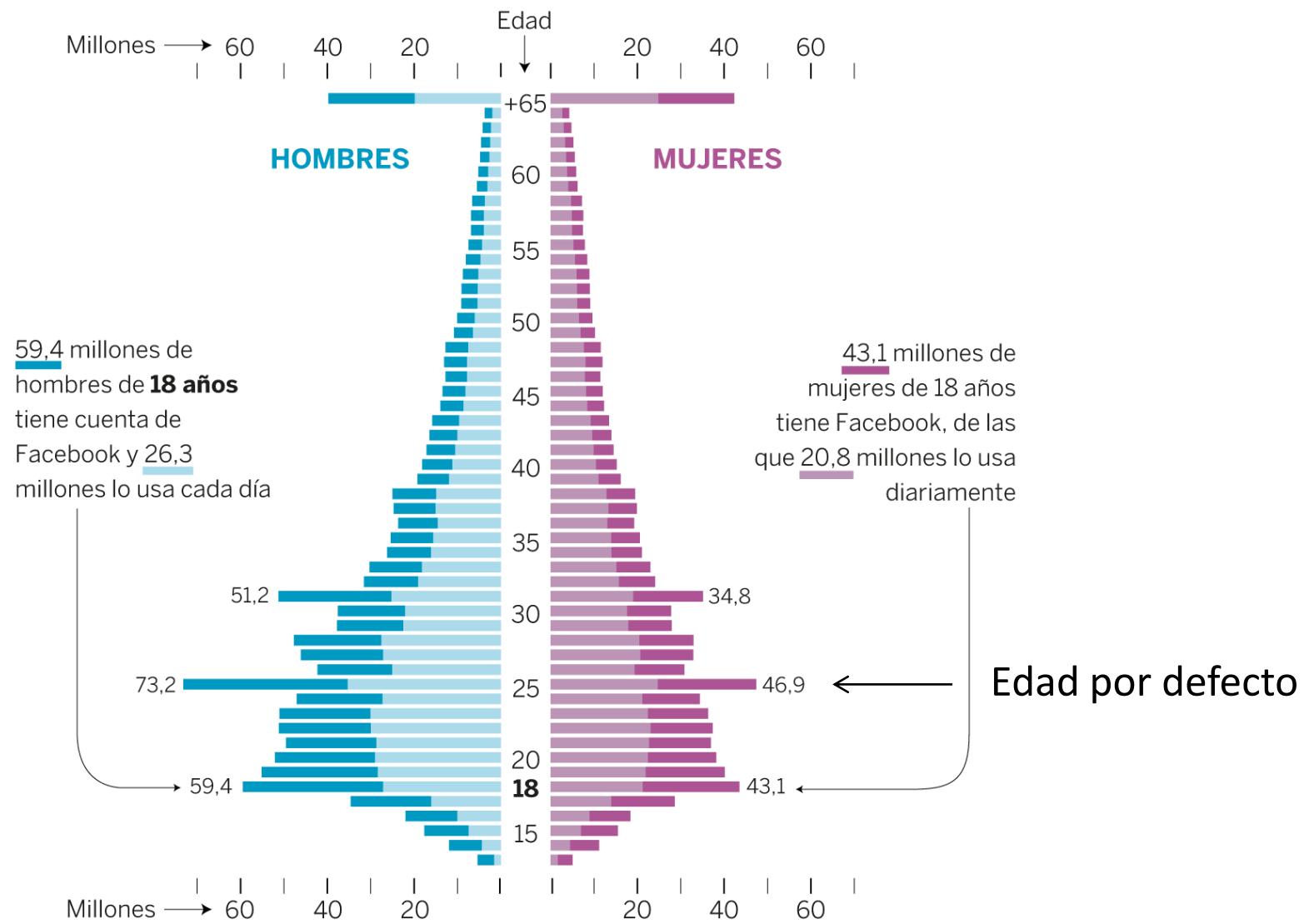
Economist.com

2

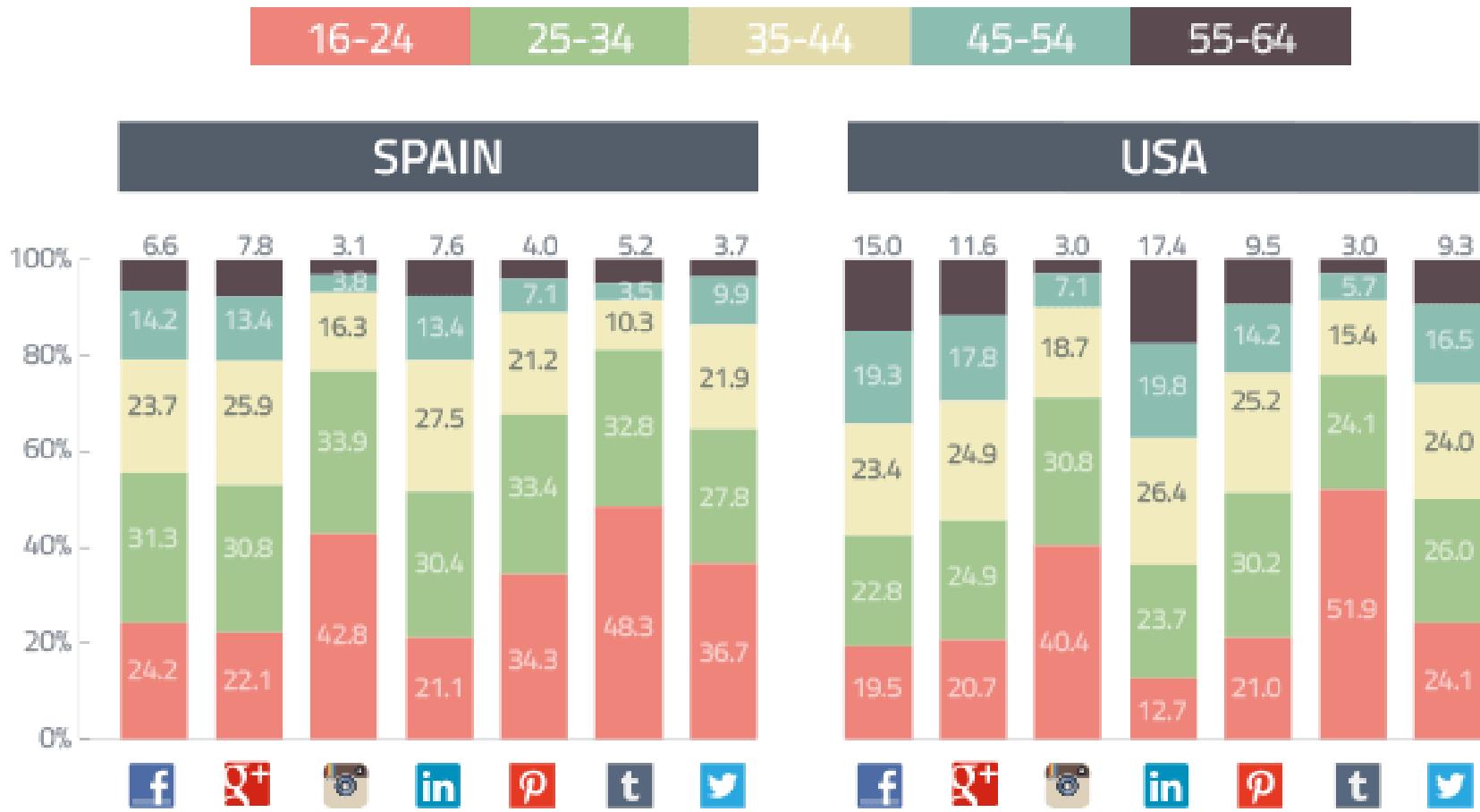
Cifras globales



Pirámide poblacional de Facebook



Distribución poblacional en medios sociales



Redes sociales (cont)

- ◆ Un campo muy abierto en potencial y soluciones
- ◆ Ejemplos de problemas de interés
 - A qué personas dirigir una campaña de publicidad; cómo ganar influencia en la red; descubrir, predecir, recomendar relaciones; cómo frenar una epidemia o maximizar la difusión de un mensaje; entender cómo se relacionan las personas (cómo eligen relaciones, cómo deciden asociarse, cómo interactúan, etc.); manipular unas elecciones 😬

Y con vistas a abordar tales problemas...

- ◆ Qué características nos interesa observar y analizar en una red social
 - P.e. aquéllas que determinan el comportamiento o respuesta global de la red ante estímulos: alcance y velocidad de la propagación de estados (información, enfermedades, opinión, decisiones), impacto de la eliminación de nodos y/o arcos, etc.
 - P.e. aquéllas que determinan el papel de ciertas personas o grupos: en procesos de difusión, influencia, toma de decisiones (comprar, votar, entrar en un grupo, asistir a un evento), etc.
- ◆ Explicar cómo y por qué se llegan a formar las características de una red, predecir su futuro desarrollo y evolución
- ◆ Explicar y predecir cómo se va a desarrollar un proceso (p.e. difusión) sobre la red

Redes sociales y minería/búsqueda de información

- ◆ Las redes sociales están cambiando la forma en que accedemos a la información
 - Flujo, compartición, propagación, búsqueda, acceso
 - Micro escala (nuestro entorno cercano) y macro escala (propagación viral, etc.)
- ◆ Se está empezando a descubrir la utilidad de las redes en algoritmos de recomendación
 - Nos pueden ser útiles actividades y hallazgos de nuestros amigos
- ◆ “Crowd power”
 - Se difumina la barrera entre productor y consumidor
 - Contenido pero también datos, estructura, respuestas, etc.
- ◆ Las personas son también objeto de tareas de búsqueda y análisis
 - Buscar/recomendar personas, expertos, candidatos a un trabajo, amigos, pareja, etc.
 - Analizar sus propiedades y relaciones: quién es influyente, qué comunidades se observan, qué factores determinan las interacciones, etc.

Análisis de redes sociales

- ◆ Medir y describir
 - Métricas
- ◆ Explicar
 - Modelos
- ◆ Predecir
 - Anticipar fenómenos
 - Influir en ellos
- ◆ Aspectos estáticos
 - Topología (local y global)
 - Propiedades de personas y enlaces individuales
- ◆ Aspectos dinámicos
 - Aparición y desaparición de enlaces
 - Formación y variación de propiedades
 - Interacción: flujo de información y estados

¿Qué se estudia sobre las redes sociales?

- ◆ Topología micro, meso, macro
 - Componentes conexas, densidad, cohesión, distancias...
 - Redes de mundo pequeño, ley de potencias
 - ◆ Propiedades estructurales de los nodos (y enlaces)
 - Cómo están posicionadas las personas en su entorno y qué nodos son “importantes”
 - Diferentes tipos de importancia: autoridad, centralidad, influencia, mediación...
 - Propiedades relacionadas con la cohesión del entorno
 - Enlaces fuertes / débiles
 - ◆ Modelos de estructura y formación de las redes
 - Tendencias elementales (micro) y topologías a las que dan lugar
 - Modelos de grafos “aleatorios”
 - ◆ Comunidades
 - ◆ Fenómenos de propagación
 - ◆ Optimización de algoritmos para grafos de muy alta escala
 - ◆ Visualización (un reto para escalas masivas)
 - ◆ ...
- 
- “Forma” de una red:
para escalas masivas
se mide en términos de
estadísticas y métricas
cuantitativas

2. Topologías de red: métricas y estadísticas

Nociones generales de grafos

- ◆ Grafo $G = (V, A)$, nodos $u \in V$, arcos $(u, v) \in A$, $|V| = n$, $|A| = m$
- ◆ $g(u) \equiv$ grado de un nodo $u \in V$
 - En grafos dirigidos indegree, outdegree
- ◆ Tipos de grafos
 - Dirigidos / no dirigidos → por defecto vamos a suponer grafos no dirigidos, salvo cuando digamos expresamente lo contrario
 - Ponderados / no ponderados
 - Se podrían considerar multigrafos
- ◆ Componentes (fuertemente) conexas
- ◆ Caminos de distancia mínima (a.k.a. geodésicas)
 - Y de coste mínimo en grafos con arcos ponderados
- ◆ Red ego de un nodo: subgrafo a distancia ≤ 1 del nodo

A qué aplican las métricas y el análisis

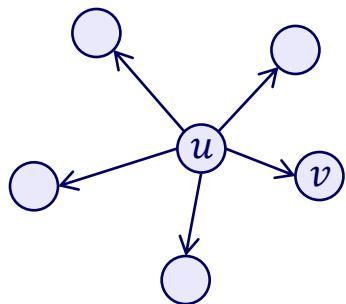
- ◆ Nodos individuales
- ◆ Arcos individuales
- ◆ La red en su totalidad

Propiedades de los nodos

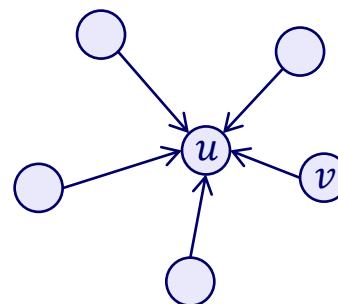
- ◆ Métricas que miden propiedades topológicas de nodos individuales, relacionadas con su importancia y el papel que pueden jugar en la red
- ◆ No hay un conjunto canónico de métricas
 - Se han definido cientos, veremos las más fundamentales y conocidas
 - Pero frecuentemente se “inventan” nuevas métricas para problemas particulares
- ◆ Grado
 - Distinción entre grado / indegree / outdegree en redes dirigidas
- ◆ Centralidad: betweenness, closeness, PageRank, autovector
- ◆ Cohesión local: coeficiente de clustering
- ◆ Existen ligeras variantes en la definición de algunas de estas métricas
 - Diferentes formas de normalizar, excluir o no el nodo en las sumas, etc.
 - Las diferencias son generalmente intrascendentes en tanto que preservan las comparaciones fijada una red (y en algunos casos para toda red)

Ejemplos

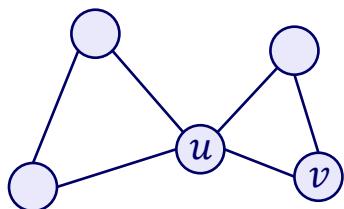
Outdegree



Indegree



Betweenness



Closeness



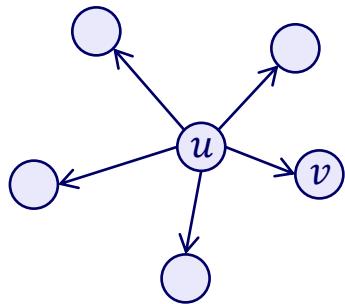
En todos estos ejemplos la propiedad es mayor en u que en v

Grado

- ◆ Nº de enlaces (en redes dirigidas, salientes / entrantes) en los que participa un nodo
- ◆ Es la métrica más simple, pero no menos significativa
 - Ya nos dice algo sobre el papel y/o importancia de los nodos
- ◆ Se considera a menudo una métrica de centralidad (es común que tenga relación p.e. con la influencia del nodo)
- ◆ Es relevante estudiar asimismo la distribución del grado como uno de los elementos característicos de una red

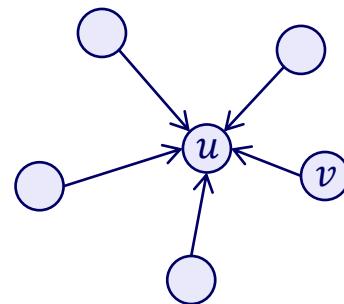
Grado

Outdegree



	g_{in}	g_{out}
u	0	5
v	1	0

Indegree

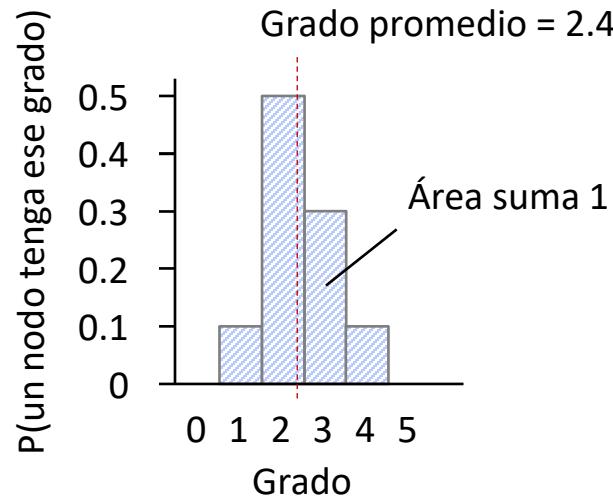
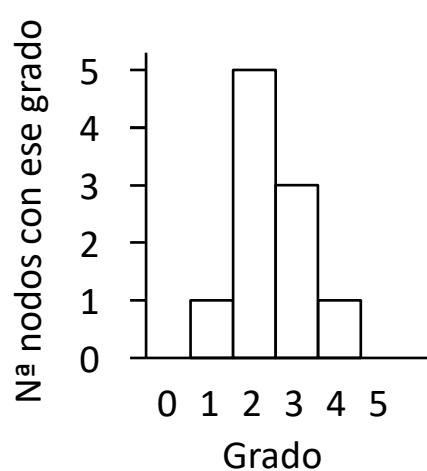
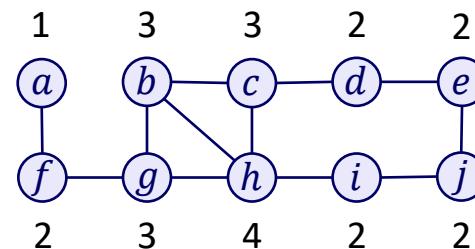
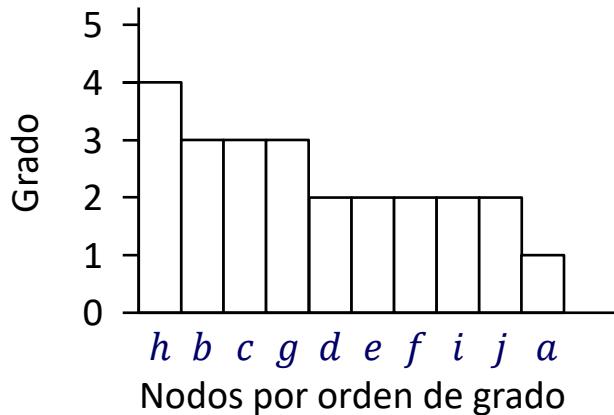


	g_{in}	g_{out}
u	5	0
v	0	1

Distribución del grado

- ◆ Forma parte esencial de la visión global de la estructura de una red
- ◆ La distribución puede visualizarse como una serie numérica: los grados ordenados de mayor a menor
- ◆ O bien, más comúnmente, se observa la frecuencia de los grados (cuántos nodos de grado 1, cuántos de grado 2, etc.)
 - Equivalente a la función de masa de la probabilidad de que un nodo al azar tenga un cierto grado (la diferencia está en dividir o no por el nº de nodos)
- ◆ Las distribuciones en redes naturales suelen estar típicamente muy sesgadas, como veremos
- ◆ También se estudia la distribución del grado en redes modelo (aleatorios), típicamente se consigue derivar una fórmula exacta

Ejemplo



Ejemplo

Grafo Facebook (J. Leskovec)

Red ego 10 usuarios

$$|V| = 4,039$$

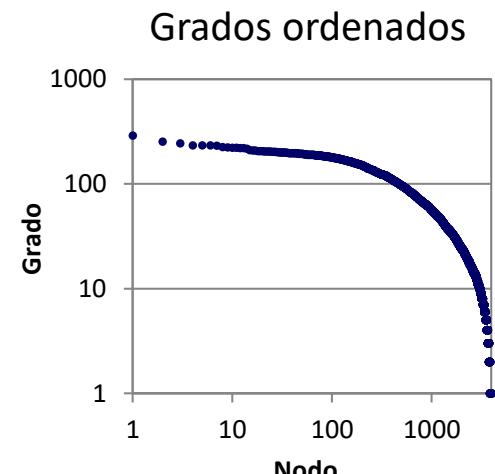
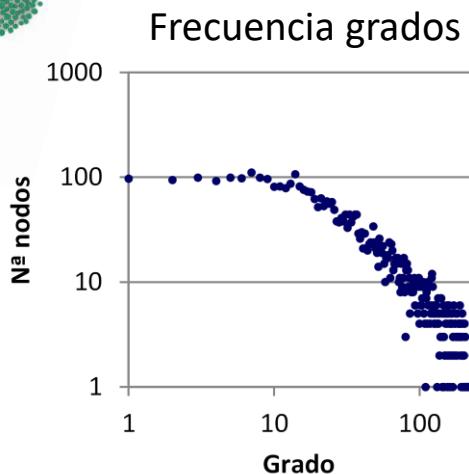
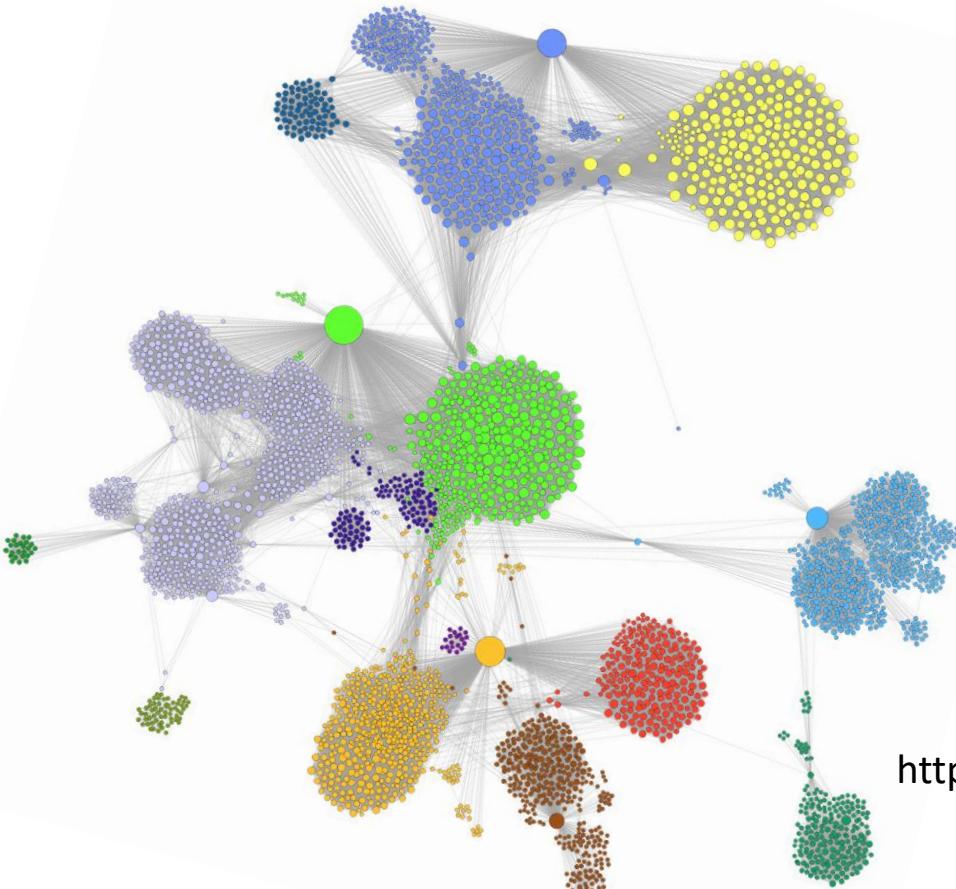
$$\text{avg}_u g(u) = 43.7$$

$$C_{\text{avg}} = 0.617$$

$$ASP = 3.7$$

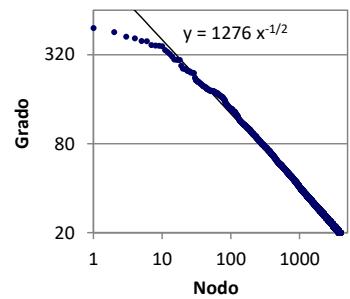
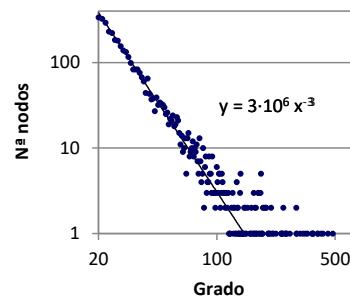
$$\text{Diámetro} = 8$$

<http://snap.stanford.edu/data/egonets-Facebook.html>



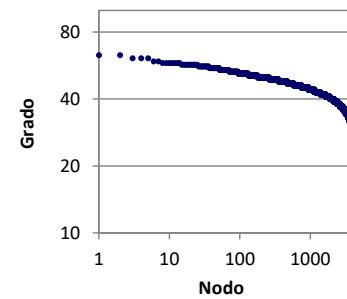
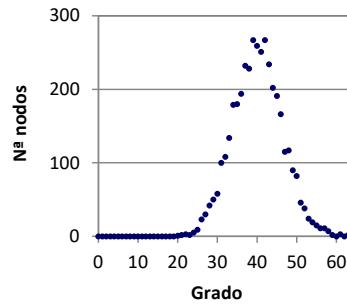
Más ejemplos

Grafo Barabási-Albert
 $|V| = 4,000$
 $\text{avg}_ug(u) = 40$
 $ASP = 2.56 (3.9)$
 $\text{Diámetro} = 4$
 $C_{\text{avg}} = 0.036 (0.002)$

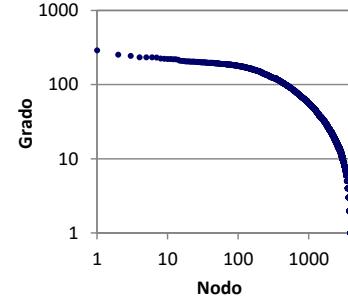
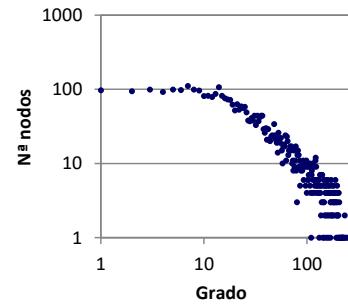


Redes generadas por modelo

Grafo Erdös-Rényi
 $|V| = 4,000$
 $\text{avg}_ug(u) = 40$
 $ASP = 2.65 (2.76)$
 $\text{Diámetro} = 4$
 $C_{\text{avg}} = 0.01 (0.01)$



Grafo Facebook (J. Leskovec)
 Red ego 10 usuarios
 $|V| = 4,039$
 $\text{avg}_ug(u) = 43.7$
 $ASP = 3.7$
 $\text{Diámetro} = 8$
 $C_{\text{avg}} = 0.617$



Red datos reales

Paradojas de la amistad

- ◆ ¿Mis amigos tienen más amigos que yo?
- ◆ El promedio del nº de amigos de los amigos es mayor que el nº promedio de amigos por persona

$$\text{avg}_u g(u) \leq \text{avg}_{u,v:u \rightarrow v} g(v)$$

- Es un hecho estadístico fácil de comprobar
- Intuición: el grado de las personas con muchos amigos participa más veces en la suma que forma el promedio
- O bien: es estadísticamente más probable ser amigo de alguien con muchos amigos que de alguien con pocos

Paradojas de la amistad (cont)

- ◆ También se cumple siempre $\text{avg}_u g(u) \leq \text{avg}_u \text{avg}_{v:u \rightarrow v} g(v)$
 - Es una formulación ligeramente distinta, fácil de demostrar también
- ◆ No es teóricamente necesario sin embargo que la **mayoría** de personas tengan menos amigos que sus amigos
 - Es otra formulación distinta: que la **mediana** sea menor que la media $m < \mu$
 - Esto se cumple si la distribución del grado es monótona decreciente
 - Así suele ocurrir en las redes naturales
 - Sucedería lo contrario si los grados altos abundasen más que los bajos

Métricas de nodos basadas en distancias

- ◆ Closeness
- ◆ Excentricidad
- ◆ Betweenness

Closeness

- ◆ No necesariamente muchos contactos, ni punto de paso, pero en una posición cercana en promedio a todos los nodos
 - Intuitivamente, estar “en medio” de la red en términos de distancia

- ◆ Se manejan variantes con ligeras diferencias, por ejemplo:

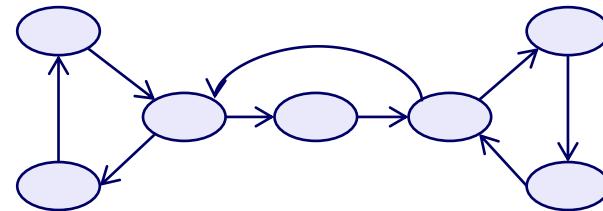
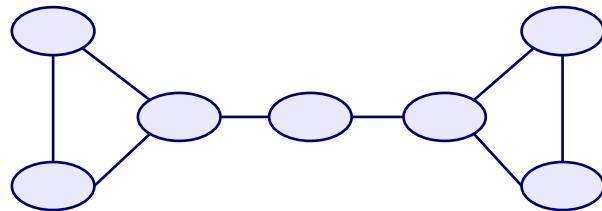
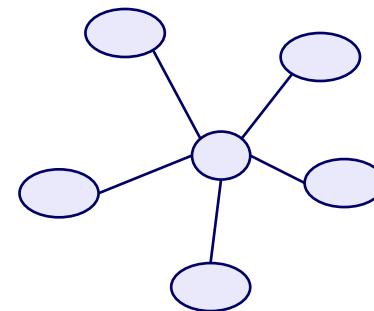
$$C(u) = \frac{n - 1}{\sum_{v \in V} \delta(u, v)} \quad // \text{ inversa de la distancia mínima media}$$

- ◆ Refleja una posición de influencia por la rapidez para llegar a los demás nodos p.e. en el paso de información
- ◆ En redes naturales las distancias $\delta(u, v)$ suelen ser muy cortas, por lo que $C(u)$ varía poco entre nodos, y es una métrica inestable a pequeños cambios en la red
 - P.e. con un solo enlace a un nodo muy central se dispara el valor
- ◆ Computación
 - Para todos los nodos: calcular todos los CDMs! $O(n(n + m))$ // u $O(nm)$ Brandes
 - Pero para un solo nodo, calcular un solo árbol CDM (con fuente en el nodo)

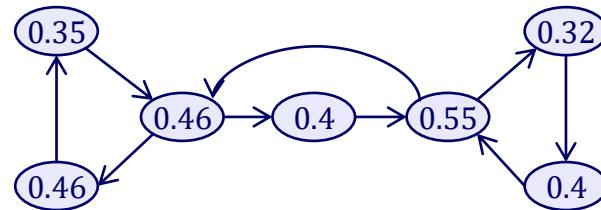
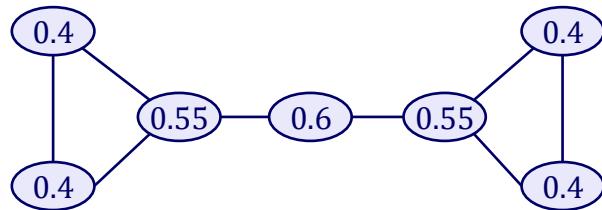
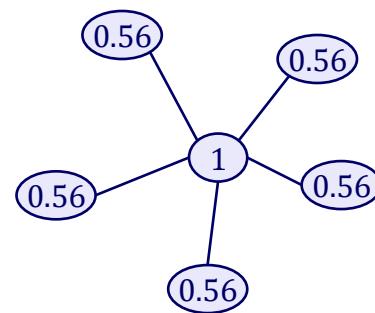
Closeness (cont)

- ◆ Cuando una red no es fuertemente conexa, todos los usuarios u tendrían $\delta(u, v) = \infty$ para algún v , y por tanto $C(u) = 0$
- ◆ Dos opciones para evitarlo
 - a) Calcular closeness en las componentes conexas como grafos separados
 - b) Closeness armónica: promedio de la inversa de las distancias en lugar de inversa de la distancia promedio: $C(u) = (n - 1) \sum_{v \neq u} 1/\delta(u, v)$

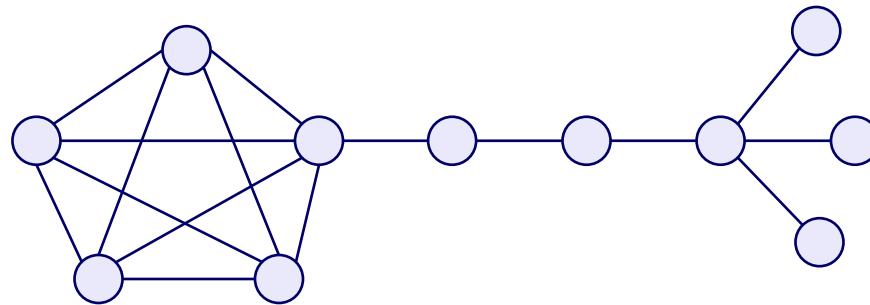
Ejemplos



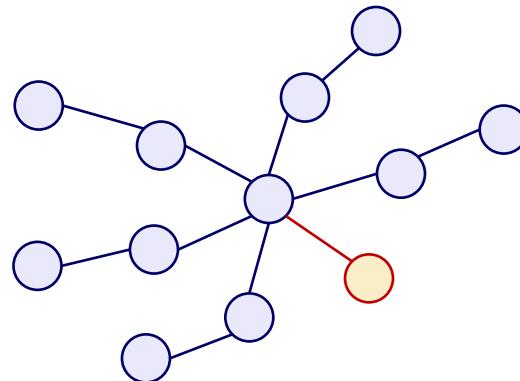
Ejemplos



Ejemplos (cont)



Grado no siempre implica closeness: este grafo contiene nodos con alto grado y bajo closeness, y viceversa. ¿Cuáles?



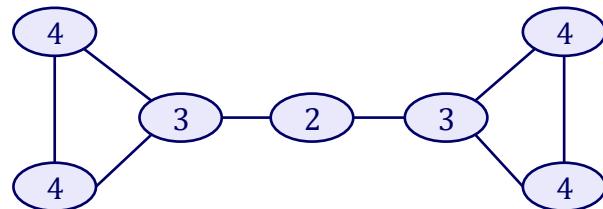
Closeness tampoco implica necesariamente betweenness

Excentricidad

- ◆ La distancia al nodo más lejano

$$e(u) = \max_{v \in V} \delta(u, v)$$

- ◆ Medida complementaria a closeness: en lugar de distancia mínima promedio, distancia mínima máxima



Betweenness

- ◆ No necesariamente muchos contactos, pero punto de paso entre muchos pares de nodos
- ◆ Ratio promedio de caminos de distancia mínima (CDM) de la red que pasan por el nodo

$$B(u) = \frac{2}{(n-1)(n-2)} \sum_{v,w \neq u} \frac{ns_{v,w}(u)}{ns_{v,w}}$$

En grafos dirigidos
omitimos esta condición

1 en grafos
dirigidos

$ns_{v,w}$ ≡ nº de CDM entre los nodos v y w

$ns_{v,w}(u)$ ≡ nº de CDM entre v y w que pasan por u

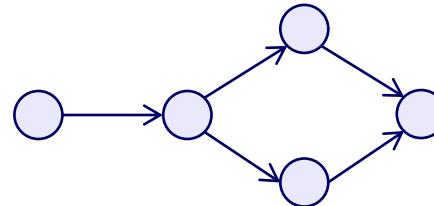
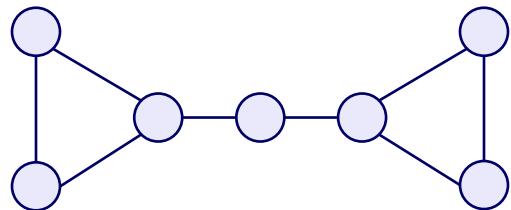
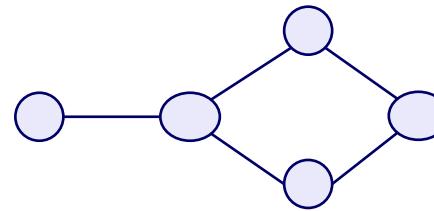
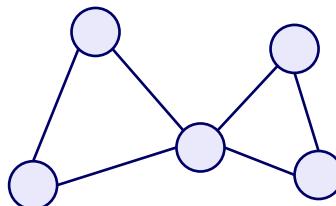
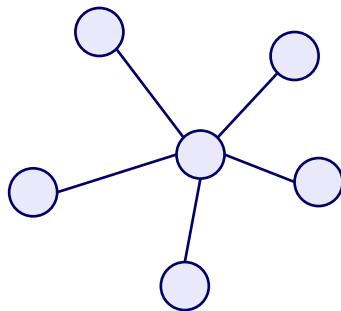
A menudo no se normaliza

Betweenness (cont)

- ◆ Si la red no es fuertemente conexa
 - Aplicamos la suma a los pares u, v tales que v es accesible desde u
 - Y normalizamos por el número de tales pares (o bien normalizamos dentro de cada componente fuertemente conexa)
- ◆ Cómputo: se necesita calcular todos los CDMs entre todos los pares de nodos (incluso para un solo nodo!)
 - BFS → árbol de CDMs desde cada nodo: $O(n(n + m))$
 - Una vez creado el bosque CDM, se calcula betweenness en $O(n(n + m))$
 - En grafos no dirigidos, $O(n m)$ con algoritmo de Brandes
 - En redes con pesos, Dijkstra / Johnson $O(n m + n^2 \log n)$, Floyd-Warshall $\Theta(n^3)$
- ◆ Los nodos con un valor alto en esta métrica tienen una posición de influencia por su papel en el paso de información
 - Su eliminación de la red tiende a crear disrupción del flujo de información

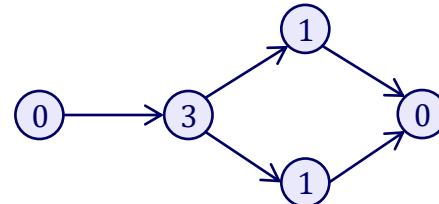
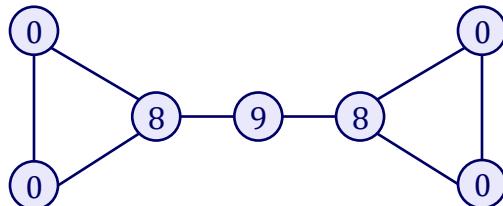
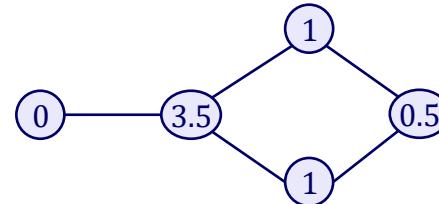
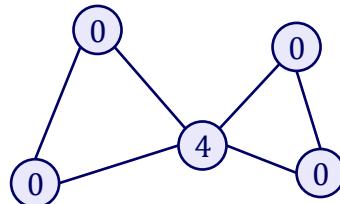
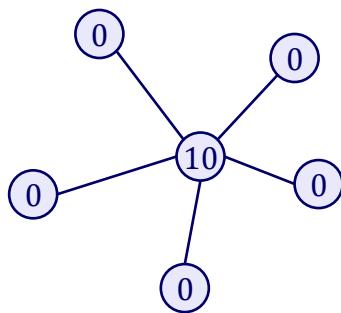
Ejemplos

(Valores sin normalizar)

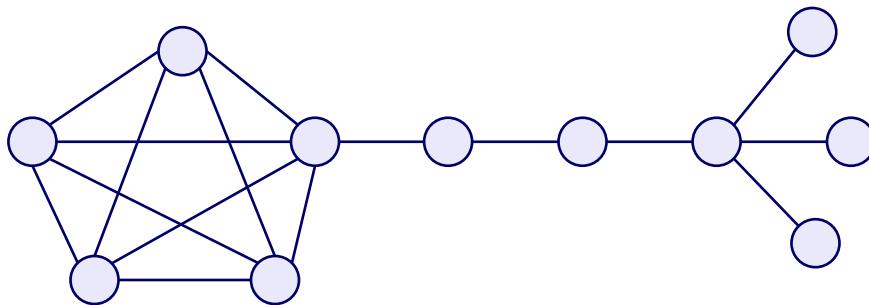


Ejemplos

(Valores sin normalizar)

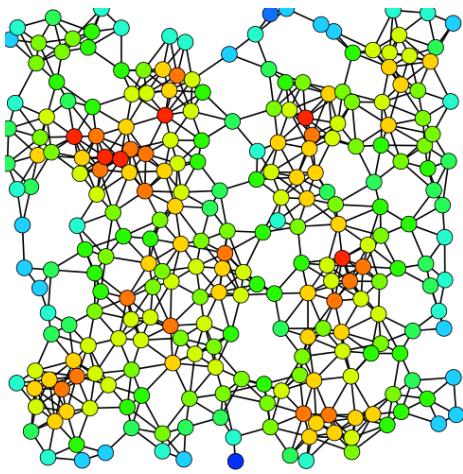


Ejemplos (cont)

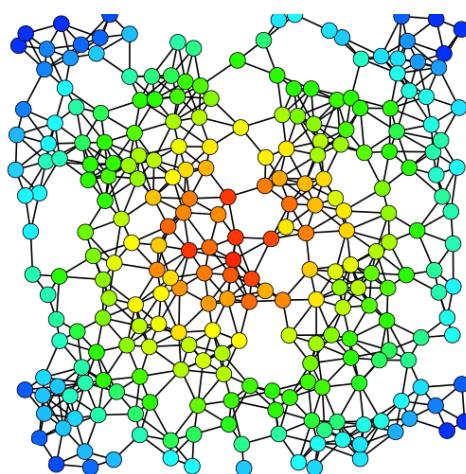


- ◆ Grado no siempre implica betweenness: este grafo contiene nodos con alto grado y bajo betweenness, y viceversa. ¿Cuáles?
- ◆ Betweenness tampoco implica closeness: ¿Qué nodos tienen alto betweenness y closeness moderado?

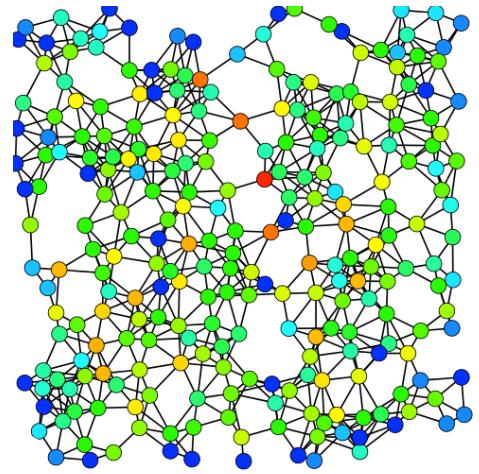
Comparación



Grado



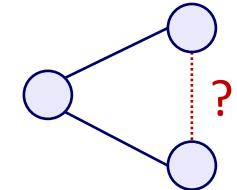
Closeness



Betweenness

Coeficiente de clustering local

- ♦ Refleja la cohesión del entorno de un nodo
- ♦ Se basa en la noción de cierre triádico
 - Transitividad: “los amigos de mis amigos son mis amigos”
 - En qué medida mis vecinos están conectados entre sí
 - Una forma de medir cómo de completo es el grafo entorno al nodo (“red ego”)
- ♦ $C(u) \equiv$ probabilidad de que dos vecinos de u tomados al azar sean vecinos



$$C(u) = p(v \rightarrow w | u \rightarrow v, u \rightarrow w) = \frac{\text{nº conexiones entre vecinos de } u}{\text{nº conexiones posibles entre vecinos de } u} \in [0,1]$$

$$g(u) < 2 \Rightarrow C(u) \triangleq 0$$

$$\text{nº conexiones posibles entre vecinos de } u = g(u)(g(u) - 1)/2$$

- ♦ El coef de clustering tiende a correlacionar inversamente con betweenness
 - Alto clustering \rightarrow redundancia en la comunicación
 - Bajo clustering \rightarrow posición ventajosa en la transmisión de información
 - Y es generalmente menos costoso de computar: $\Theta(\sum_{u \in V} g(u)^2)$
 - Aunque en redes power law muy sesgadas se puede disparar $\sum_{u \in V} g(u)^2$

Coeficiente de clustering global

- ◆ Refleja la cohesión global de entornos en la red
- ◆ Probabilidad de que dos nodos de la red con un amigo común tomados al azar estén conectados

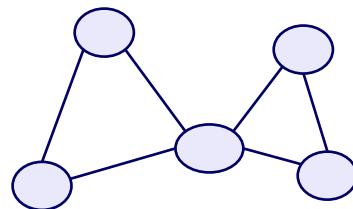
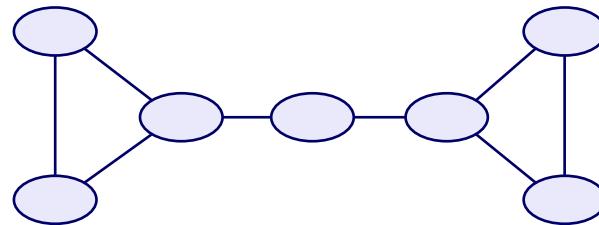
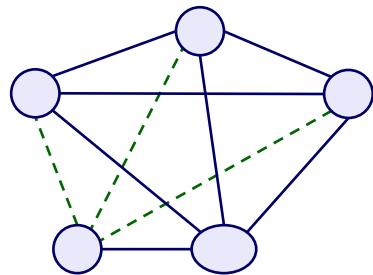
$$C(G) = \frac{\text{nº caminos cerrados de long 2}}{\text{nº de caminos de long 2}} \in [0,1]$$

- ◆ Equivalentemente a la definición anterior, fracción de tripletas transitivas

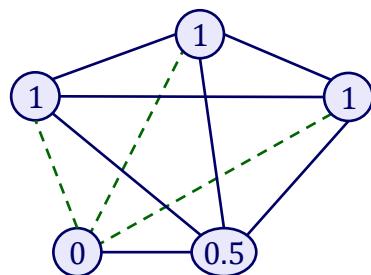
$$C(G) = \frac{3 \times \text{nº triángulos en la red}}{\text{nº de tripletas conectadas}}$$

- ◆ Definición alternativa: $C_{\text{avg}}(G) \equiv \text{avg}_u C(u)$
 - En general se prefieren las definiciones anteriores (en ésta dominan los nodos de bajo grado)
- ◆ El coef de clustering de una red depende de cómo se forman las amistades
 - Si se formasen al azar, sería bastante bajo
 - Mucho más alto si se forman por mediación de amigos, similitud, popularidad...

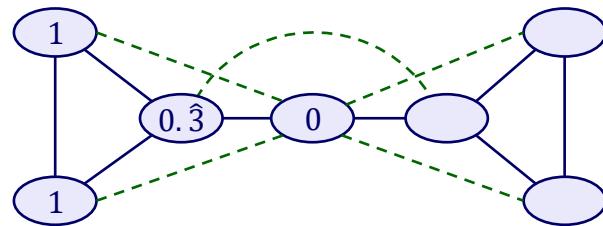
Ejemplos



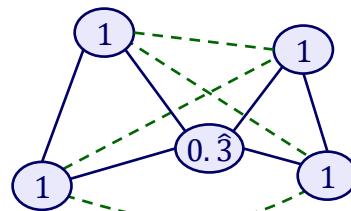
Ejemplos



$$C(G) = \frac{3 \cdot 4}{15} \quad C_{\text{avg}}(G) = \frac{3.5}{5}$$



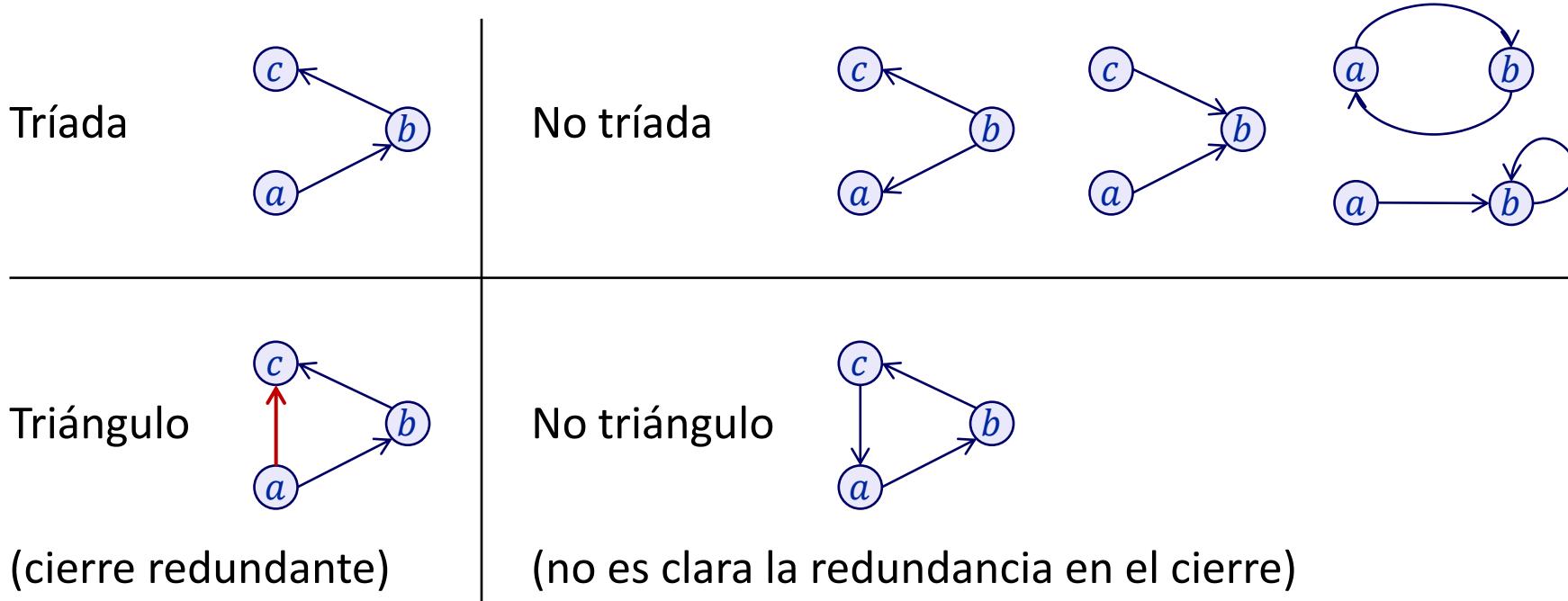
$$C(G) = \frac{3 \cdot 2}{11} \quad C_{\text{avg}}(G) = \frac{4.6}{7}$$



$$C(G) = \frac{3 \cdot 2}{10} \quad C_{\text{avg}}(G) = \frac{4.3}{5}$$

Coeficiente de clustering en redes dirigidas

- ◆ Caben diferentes generalizaciones, una opción común es:
 - Considerar como tríadas las formaciones de tipo $(a, b) + (b, c)$
 - Y como triángulos las formaciones de tipo $(a, b) + (b, c) + (a, c)$
 - Con $a \neq b, a \neq c, b \neq c$

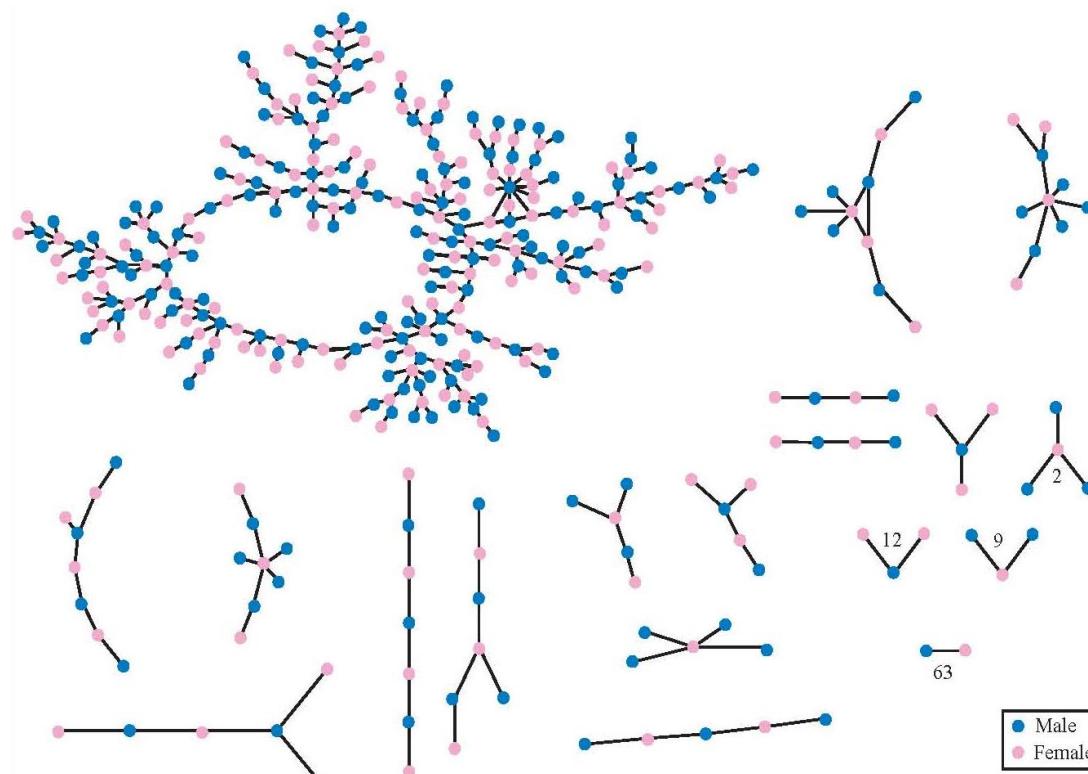


Coeficiente de clustering en redes sociales

- ◆ Habitualmente es “anómalamente” alto comparado con un desarrollo aleatorio de la red: la densidad de las redes sociales tiende a estar fragmentada en comunidades
- ◆ Un amigo común aumenta la oportunidad de enlace
- ◆ Cuando las redes son homófilas, la similitud tiende a menudo a ser transitiva
- ◆ Pueden derivarse ventajas en compartir contactos
- ◆ Existe un factor latente común a la formación de contactos, p.e. la participación en una actividad común (filiación)
- ◆ Y otros posibles factores y teorías...

Coeficiente de clustering en redes sociales

- ◆ En ausencia de homofilia...
- ◆ Ejemplo: $CC = 0.005$ (frente a p.e. ~ 0.2 más típico)

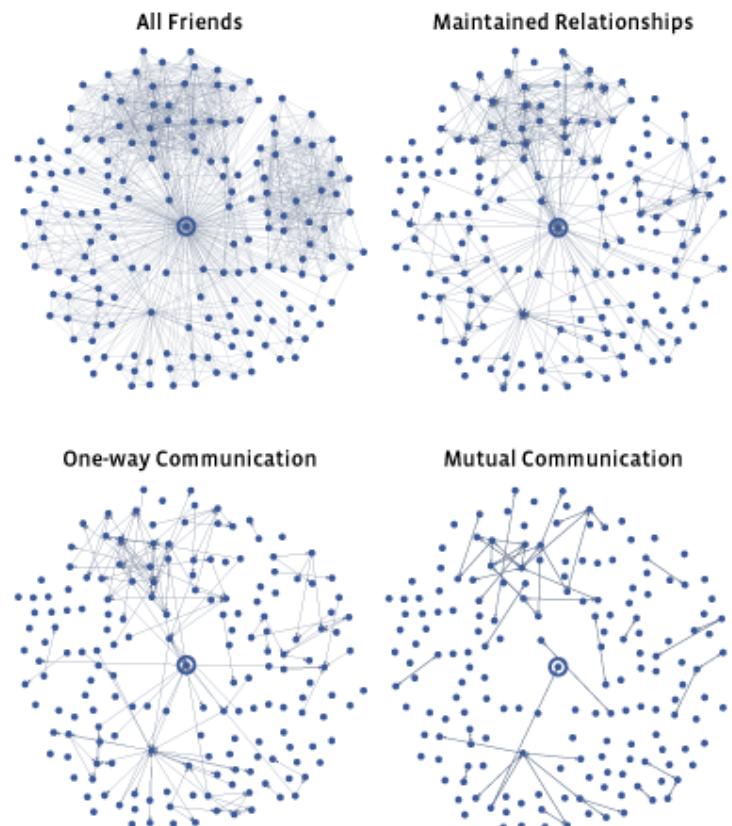


Métricas sobre enlaces

- ◆ Miden el papel o el valor de un enlace específico entre dos usuarios
 - Tanto si el enlace existe como si no
 - Efecto que aporta para los dos usuarios pero también para la red
- ◆ Muchas medidas giran en torno a nociones de **enlace débil / fuerte**
 - Los enlaces “débiles” suelen tener un valor especial
- ◆ ¿Qué es un enlace débil o fuerte?
 - Definiciones relativas a la interacción: cómo es la relación entre las personas (tipo de relación, frecuencia, duración, semántica, etc.)
 - Definiciones estructurales: cómo es la red en el entorno del enlace (arraigo, puentes, betweenness)

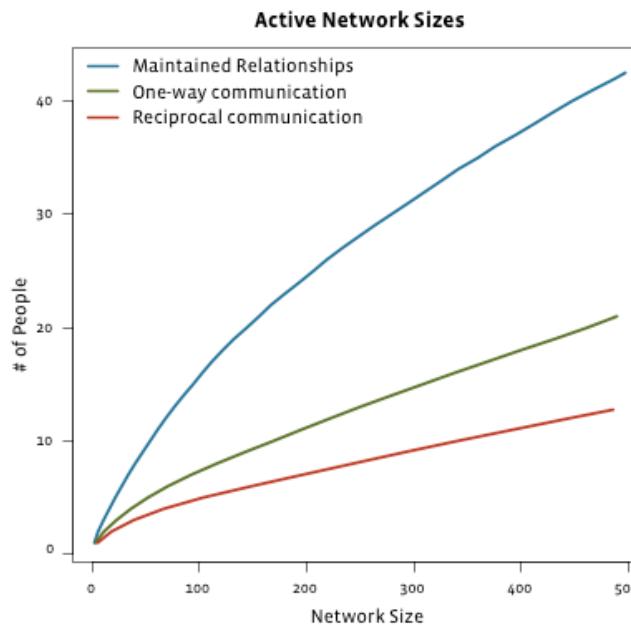
Fuerza de los enlaces: nociones de dominio

- ◆ Nociones de fuerza/debilidad relativas a la interacción y tipo de relación
- ◆ Propias del dominio



[http://overstated.net/2009/03/09/
maintained-relationships-on-facebook](http://overstated.net/2009/03/09/maintained-relationships-on-facebook)

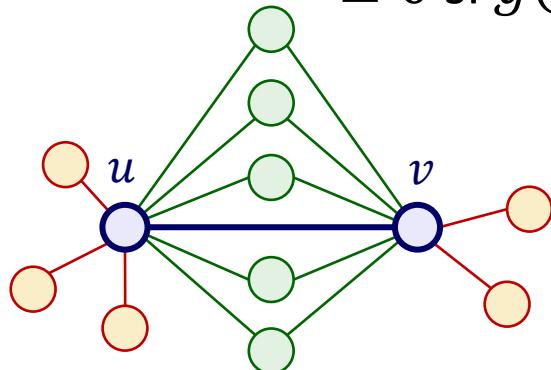
- P.e. simple conexión en una red online vs. frecuencia de interacción directa
- P.e. solidez y estabilidad de la conexión, grados de intensidad, confianza, relación activa vs. pasiva, etc.
- Enlace positivo vs. (implícitamente) negativo
- Correlación: interactuamos más en los entornos con más densidad de enlace (trabajo, etc.)



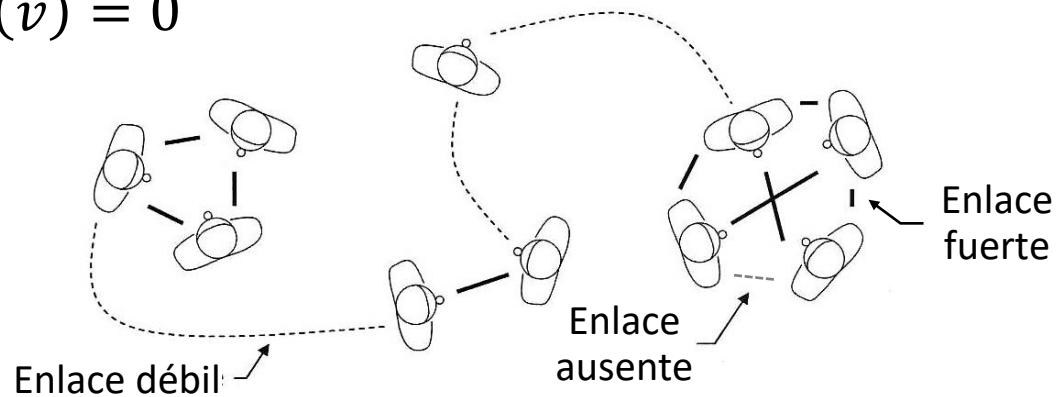
Fuerza estructural: arraigo

- ◆ Solapamiento de vecindarios (embeddedness, overlap...)
- ◆ Arcos arraigados conectan nodos con muchos contactos en común

$$\begin{aligned}\text{Arraigo}(u, v) &= \text{Jaccard}(\text{vecinos}(u) \setminus \{v\}, \text{vecinos}(v) \setminus \{u\}) \\ &\triangleq 1 \text{ si } (u, v) \in A \text{ y } g(u)g(v) = 1 \\ &\triangleq 0 \text{ si } g(u)g(v) = 0\end{aligned}$$



$$\text{P.e. Arraigo}(u, v) = 0.5$$

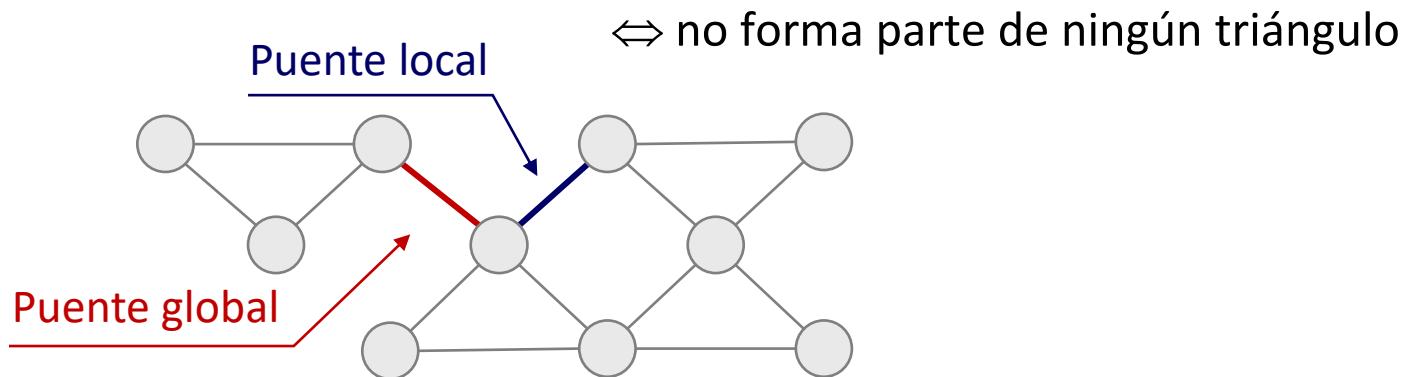


http://en.wikipedia.org/wiki/Interpersonal_ties

- ◆ Es común que el arraigo correlacione con otras nociones de fuerza/debilidad de enlaces propias del dominio

Fuerza estructural: puentes

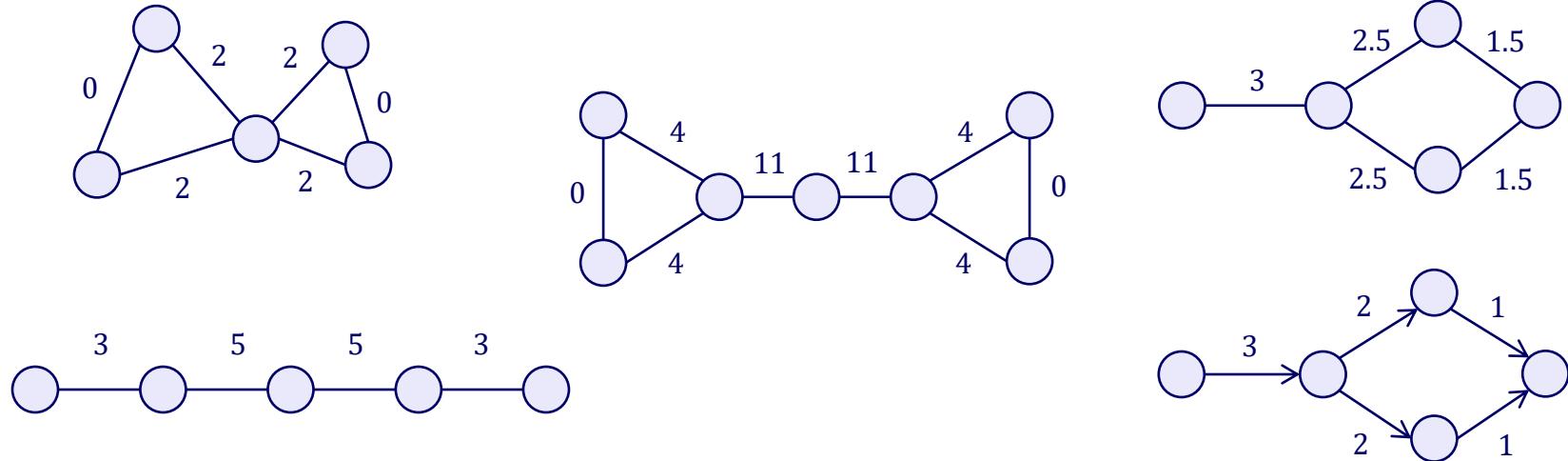
- ◆ Noción relacionada con el arraigo
- ◆ Global: si se elimina el enlace se crea una componente conexa más (cabe considerar puente, puente fuerte –por defecto–, puente débil)
- ◆ Local: si se elimina un enlace (a, b) , entonces $\delta(a, b) > 2$
 - En grafos no dirigidos puente local \Leftrightarrow arraigo 0



- ◆ Punto de vista complementario al coeficiente de clustering local: usuarios con muchos enlaces débiles tienden a bajo clustering

Fuerza estructural: betweenness

- ◆ Se define betweenness de los arcos, igual que de los nodos
- ◆ Promedio (sobre todos los pares de nodos) del ratio de CDMs que pasan por el arco
- ◆ Es decir, para un arco (a, b) , $ns_{v,w}(a, b)$ en lugar de $ns_{v,w}(u)$



Fuerza estructural: interpretación

- ◆ Nodos arraigados vs. mediadores
 - Los enlaces poco arraigados favorecen el alcance de búsqueda y expansión (menos ciclos/redundancias), a pesar de ser débiles (p.e. menos transitados)
 - El comportamiento de los nodos mediadores es clave también: qué harán éstos con sus enlaces débiles, con qué frecuencia los usan, qué grado total tiene el nodo, etc.
- ◆ “Signo” de los enlaces
 - Amigos vs. “enemigos”
 - P.e. ausencia de enlace con fuerte arraigo (i.e. entre personas con muchos contactos comunes) puede indicar aversión

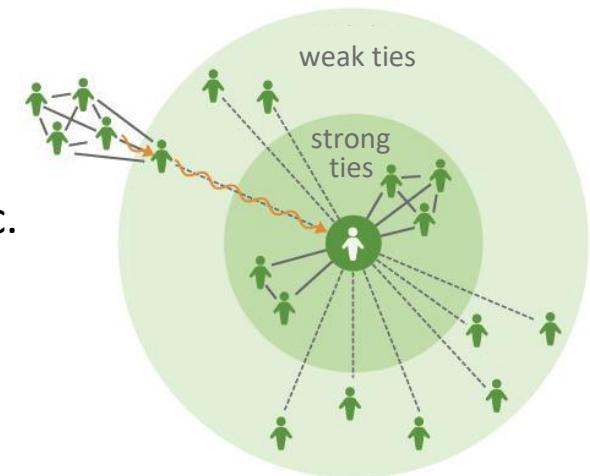
Valor de los enlaces débiles



- ◆ Los enlaces débiles y fuertes tienen distinta utilidad

- Fuertes: ventajas a nivel individual 1-1, p.e. más disponibilidad, fiabilidad, etc. – vivir sin ellos es difícil!
 - Débiles: ventajas en la interacción global, enriquecen y aceleran el flujo de información global, “exclusividad” del contacto a nivel individual, etc.
 - M. S. Granovetter. The strength of weak ties. American Journal of Sociology 78(6), 1973

Mark S. Granovetter
(1943-)



- ◆ Agujeros estructurales

- “People who stand near the holes are at higher risk of having good ideas”
 - R. S. Burt. Structural Holes: The Social Structure of Competition. Harvard University Press, 1995
 - R. S. Burt. Structural Holes and Good Ideas. American Journal of Sociology 110(2), 2004



Ronald S. Burt
(1949-)

Valor de los enlaces débiles (cont)

- ◆ P.e. en Facebook, la mayoría de enlaces son débiles
 - P. De Meo, E. Ferrara, G. Fiumara, A. Provetti. On Facebook, most ties are weak. Communications of the ACM 57(11), 2014
- ◆ Además la mayor parte del flujo de información en Facebook transcurre a través de enlaces débiles
 - E. Bakshy, I. Rosenn, C. Marlow, L. Adamic. The role of social networks in information diffusion. WWW 2012

Relación entre métricas

- ◆ En redes naturales, las siguientes métricas tienden a tener distribución power law y correlacionan con el grado
 - Betweenness
 - Coeficiente de clustering (inversamente)
 - PageRank
- ◆ Tiene distribución irregular más o menos centrada en la media, y correlaciona menos con el grado
 - Closeness
- ◆ Estas correlaciones pueden variar sensiblemente según los grafos
 - Ver p.e. presentación de P. Boldi en WOA 2012
<http://boldi.di.unimi.it/woa.pdf>

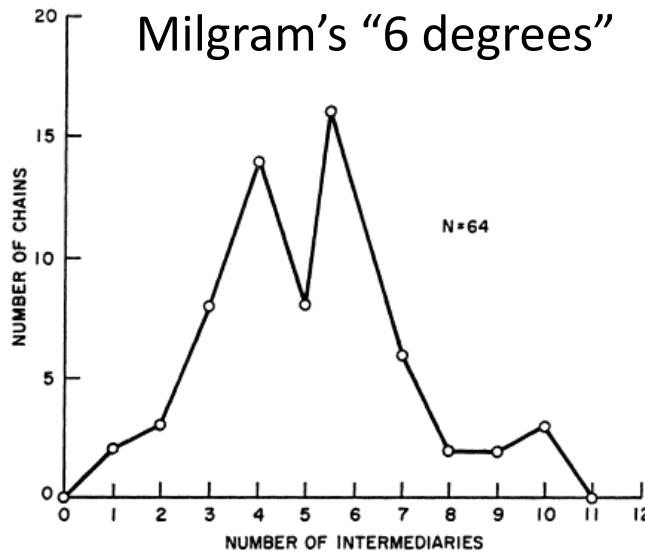
Propiedades globales de las redes

- ◆ Ya vistas:
 - Grado promedio (equivalente a densidad)
 - Distribución de los arcos (i.e. del grado de los nodos)
 - Coeficiente de clustering (i.e. cohesión)
- ◆ Distancias mínimas
- ◆ Estructura de comunidades: componentes conexas, cliques & cores, comunidades
- ◆ Asortatividad: en qué medida los usuarios se relacionan con usuarios similares (homofilia) o diferentes (heterofilia)
 - Similitud de tipo (categórica)
 - Similitud escalar
 - Similitud de grado

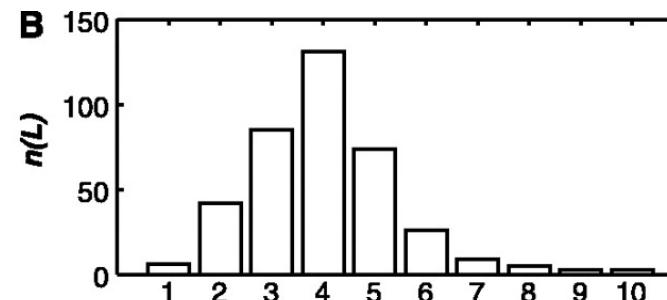
Las redes de gran escala (p.e. $> \sim 100K$ nodos) son difíciles de analizar cualitativamente (p.e. por inspección visual)

Por ello típicamente se observan mediante estas métricas, propiedades y estadísticas globales

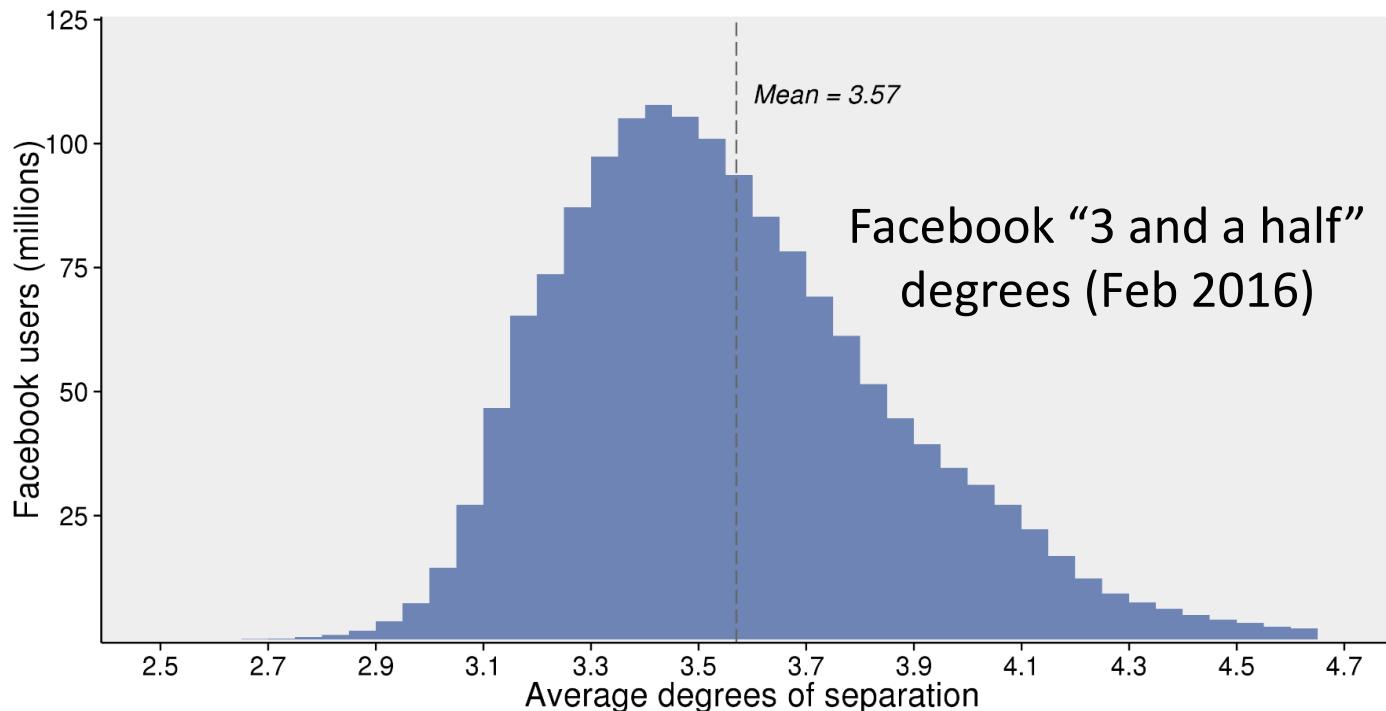
Milgram's "6 degrees"



Distancias mínimas



Dodds et al 2003



Distancias mínimas: promedio

- ◆ “Average shortest path” (ASP)
- ◆ Promedio de las distancias entre todos los pares de nodos

en grafos no dirigidos, 2
si en el sumatorio no se
repiten los pares de nodos

$$\text{ASP} = \frac{1}{n(n-1)} \sum_{u,v \in V} \delta(u, v)$$

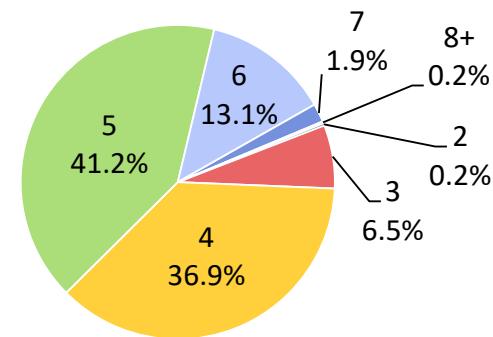
- ◆ En grafos no conexos, se puede medir ASP por separado en cada componente conexa
- ◆ O se puede formar un promedio conjunto, p.e.:

$$\text{ASP} = \left(\sum_m \sum_{u,v \in \mathcal{C}_m} \delta(u, v) \right) / \sum_m |\mathcal{C}_m| |\mathcal{C}_m - 1|$$

- ◆ O contar la media restringida a los pares accesibles

Distancias mínimas

- ◆ Diámetro: distancia mínima máxima $\max_{u,v} \delta(u, v)$
 - Diámetro exacto, p.e. > 40 en Facebook (estudio de P. Boldi)
 - Diámetro efectivo en percentil 90%: descartando el 10% de diámetros más largos
- ◆ % de usuarios a determinada distancia
 - Promedio por usuario: % de la red accesible a través de k pasos
 - Promedio por pares, p.e. 92% de pares de usuarios a distancia ≤ 5 en Facebook (P. Boldi), 87% en Twitter
- ◆ Etc.
- ◆ Radio: distancia mínima máxima
mínima $\min_u \max_v \delta(u, v)$
- ◆ $\text{radio} \leq \text{diametro} \leq 2 \text{ radio}$



Distribución de distancias en Twitter
<http://www.sysomos.com/insidetwitter>, 2010

Algunos ejemplos de métricas en redes reales

	$ V $	$\text{avg}_u g(u)$	$C(G)$	$C_{\text{avg}}(G)$	ASP	% GC	r	α
Facebook	> 1.000M	140	–	–	4.7	99.9%	–	–
Actores	~450.000	113.4	0.20	0.78	3.48	98%	0.208	2.3
Coautoría math	~250.000	3.92	0.15	0.34	7.57	82.2%	0.120	–
Mensajes email	~60.000	1.44	–	0.16	4.95	95.2%	–	1.5/2
Jefferson High	573	1.66	0.005	0.001	16.01	50%	- 0.029	–
WWW 2000	~200M	7.2	–	–	16.18	91.4%	–	2.1/2.7
Internet 1999	~10.000	5.98	0.035	0.39	3.31	100%	- 0.189	2.5