

Instituto de Ingeniería del Conocimiento



Introducción al Análisis de Redes Sociales

Grupo de Procesos Empresariales con Efectos de Red Social (PEERS)

¿Qué es una red social?

- Es una manera diferente de mirar a una colectividad

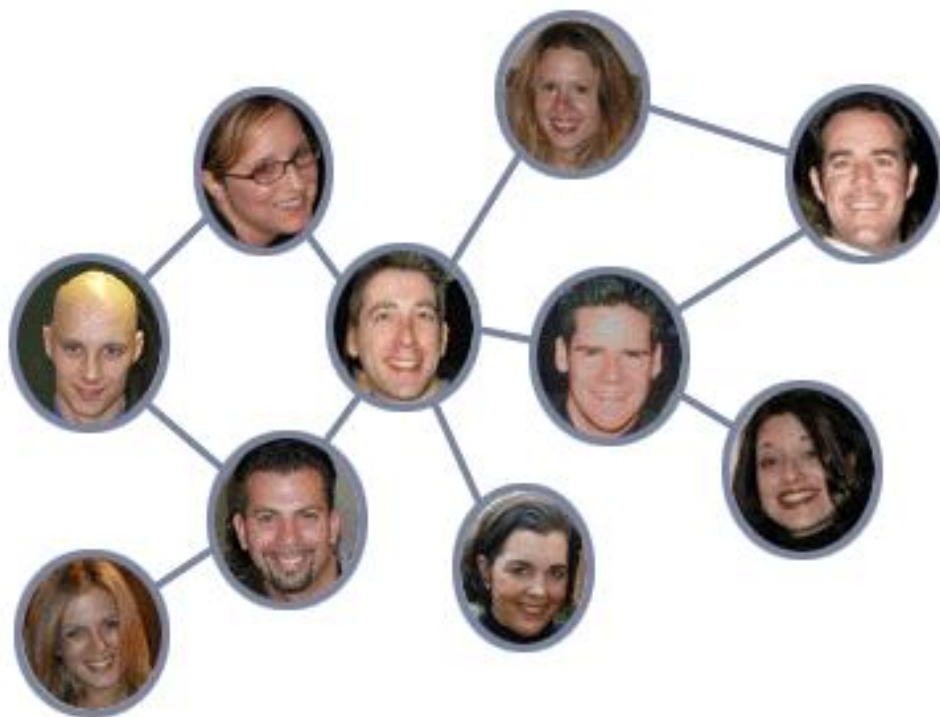
Colectividad = nodos + enlaces

- **Nodos**

- Personas
- Organizaciones
- Compañías

- **Enlaces:**

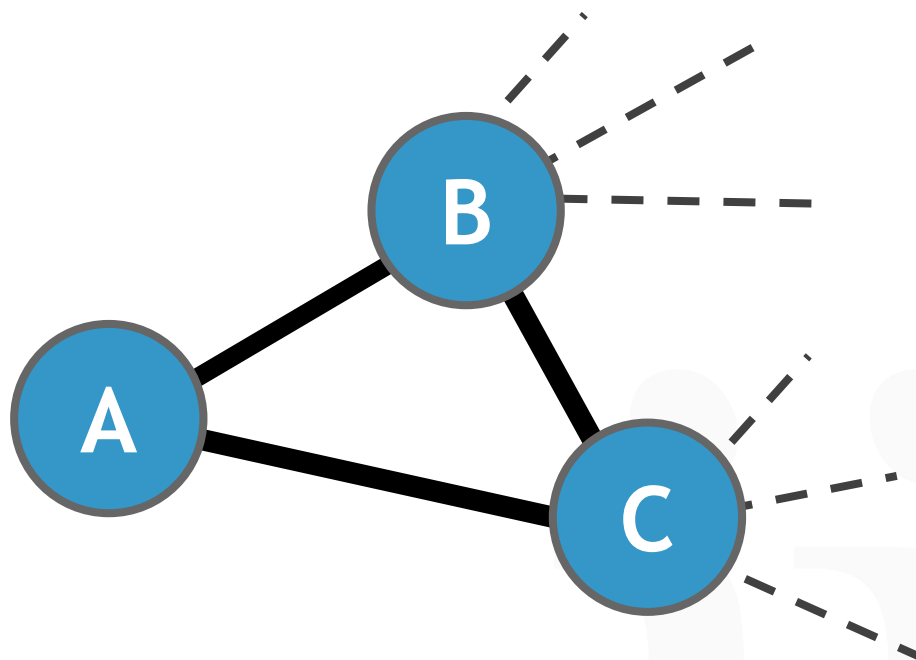
- Operaciones:
 - Llamadas/emails intercambiados
 - Transacciones
- Relaciones, amistad
- Afiliaciones
 - Membresía común



El valor de una red social

- **¿Por qué mirar a una colectividad como una red social?**

Por que el valor de una red social es mayor que la suma de sus partes.
Es por tanto un sistema complejo



- **¿Cuanta de la actividad de A con B y C depende del hecho de que B y C están conectados?**

Las Redes Sociales tienen una rica estructura

- **“Pequeño mundo”**: el diámetro de una red social es pequeño en comparación con el número de nodos.
- **Heterogéneas**: no todos los nodos en la red tienen las mismas propiedades estructurales locales. Existe una fracción significativa de nodos altamente conectados (hubs).
- **Clusterizadas**: a corto y medio rango, las redes sociales presentan una gran densificación debido a la existencia de comunidades, transitividad, etc.

Propiedades de una red social

Pequeño mundo

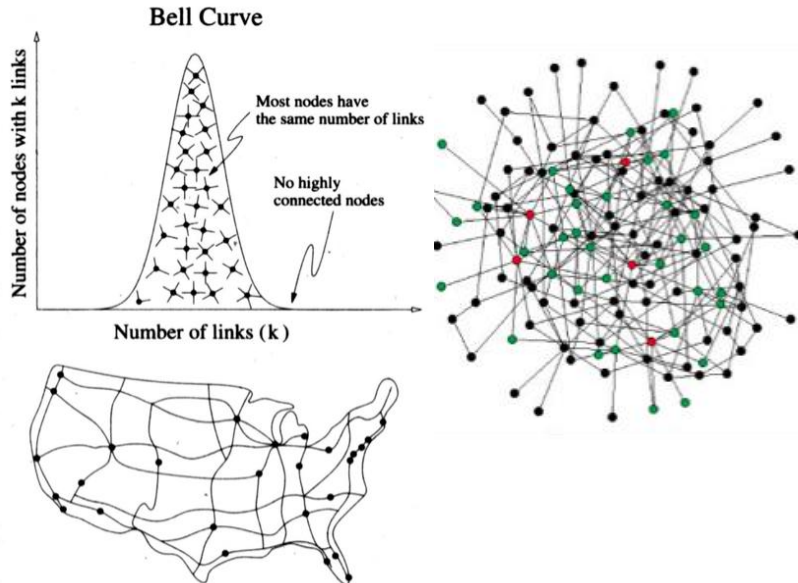
- Incluso en grandes redes sociales, la distancia social promedio es muy pequeña
 - Milgram 1967* (usando correo normal)
 - Distancia promedio: seis grados de separación
 - Dodds et al 2003* (usando *e-mail*)
 - Distancia promedio: seis grados de separación (!)



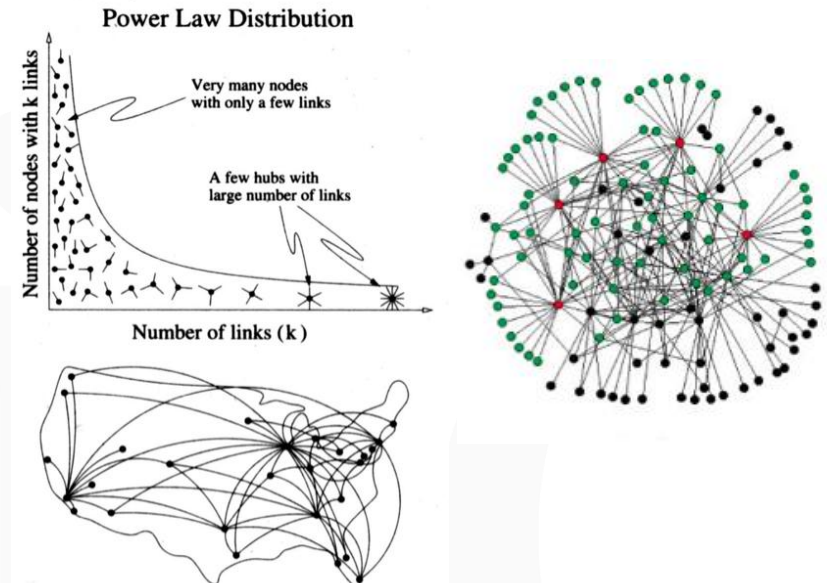
Propiedades de las redes sociales

- Las redes sociales son muy heterogéneas
 - Ley de Pareto (o ley de potencias):
 - 20% de los nodos tienen el 80% de la conectividad social
 - 80% de los nodos sólo tienen el 20% de la conectividad social

Homogeneous Network

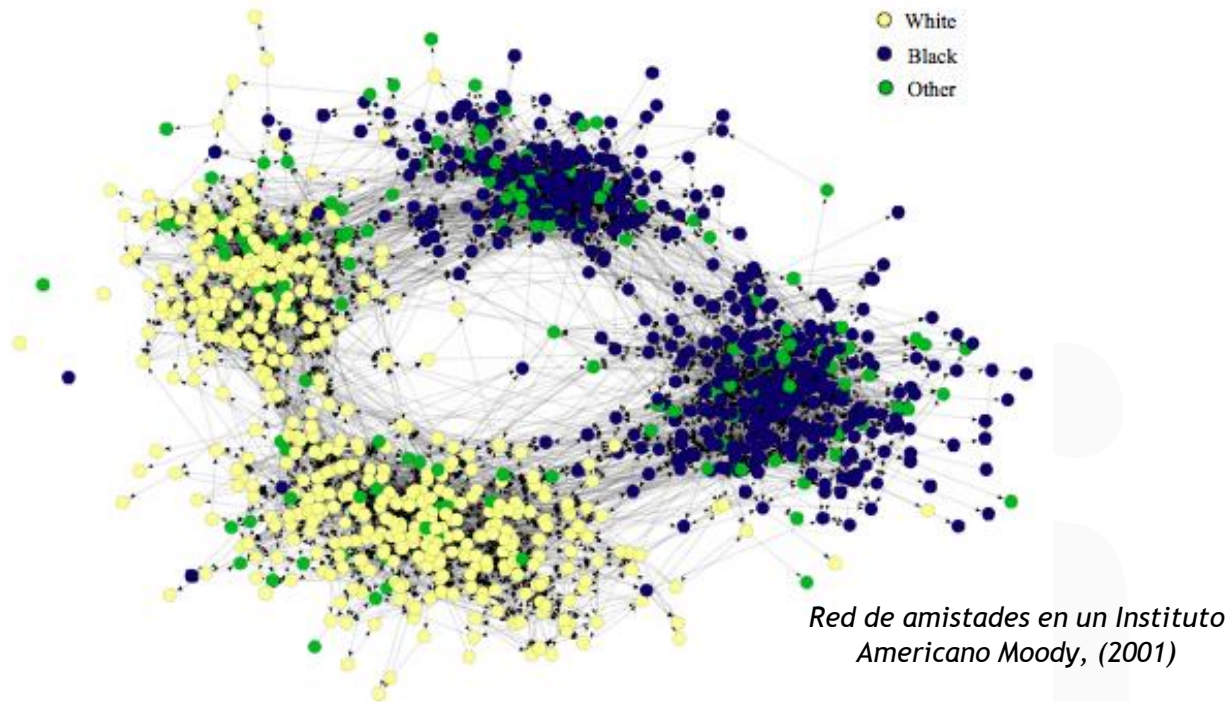


Heterogeneous Network



Propiedades de las redes sociales

- **Existencia de Comunidades**
- Grupos de nodos que tienen una diferente densidad de conexiones dentro del grupo que con el resto de la red



- **Transitividad: las redes sociales son densas localmente**
- La probabilidad de que dos vecinos de un nodo sean a su vez vecinos es muy alta (20%)

Análisis de redes sociales

- Representación y medición de redes sociales
 - Representación: De los datos a las relaciones y nodos
 - ¿Cuando un nodo es un nodo?
 - ¿Cuándo un enlace es un enlace?
 - Medición
 - Local
 - Conectividad social en la vecindad de un nodo
 - Clustering local o transitividad
 - Global
 - Caminos mínimos, diámetro de la red social
 - Centralidad del nodo (centros, puentes, fronteras, etc.)
 - Identificación de comunidades
 - Visualización

Midiendo :: Conectividad

■ A primeros vecinos

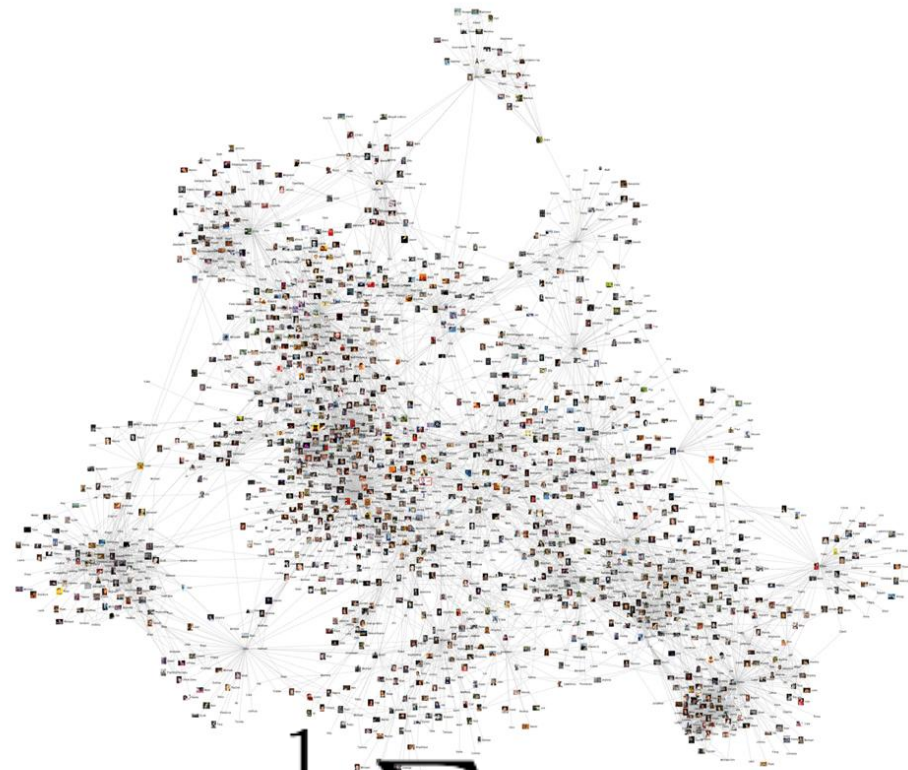
- Conectividad del nodo k_i
- Distribución de conectividad $P(k)$
- Momentos de la distribución:

$$\langle k^n \rangle = \sum_k k^n P(k)$$

■ A segundos vecinos

- Conectividad promedio de los vecinos
- Probabilidad condicionada $P(k'|k)$
- Conectividad promedio de los vecinos condicionada

$$k_{nn}(k) = \sum_{k'} k' P(k'|k)$$



$$k_{nn,i} = \frac{1}{k_i} \sum_{j \in \mathcal{N}_i} k_j$$

Midiendo :: Conectividad

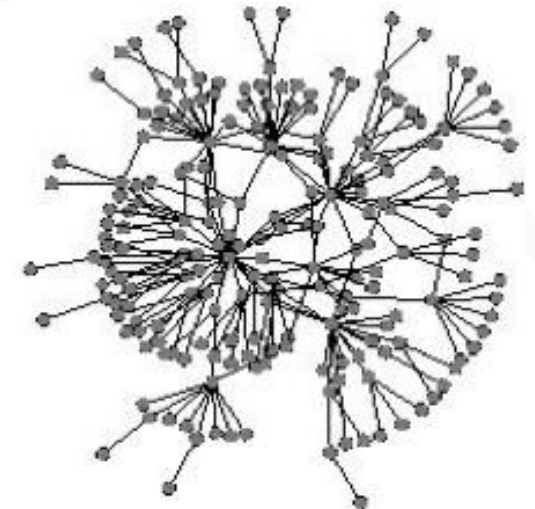
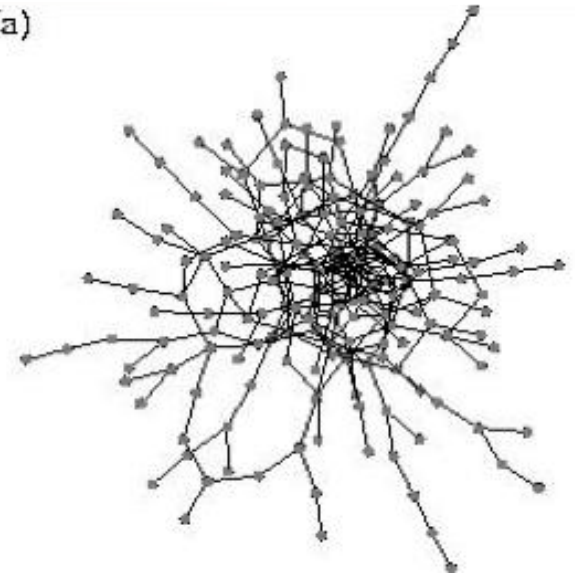
■ Asortatividad

- Los vecinos más próximos muestran una correlación en la conectividad social
- Red asortativa: nodos que están muy conectados tienden a estar rodeados de nodos con alta conectividad

$$P(k'|k) = k' P(k') / \langle k \rangle$$

- Redes disortativas: nodos que están muy conectados tienden a estar rodeados de nodos con baja conectividad

(a)

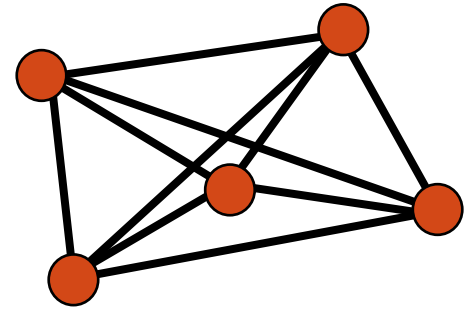


Midiendo :: Clustering

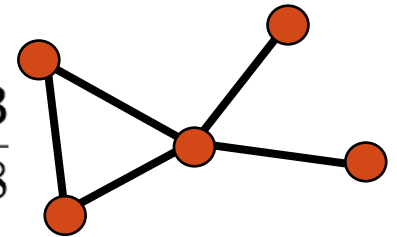
- Los vecinos de un nodo tienden a estar conectados entre ellos
 - Transitividad: probabilidad de que dos vecinos de un mismo nodo sean vecinos entre ellos

$$T = \frac{3 \times \# \text{ of triangles}}{\# \text{ of connected triples of vertices}}$$

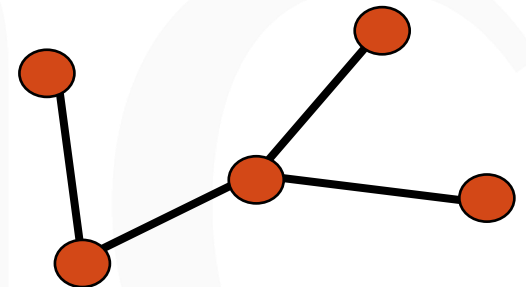
$$T = 1$$
$$C = 1$$



$$T = 3 \times \frac{1}{8} = \frac{3}{8}$$



$$T = 0$$
$$C = 0$$



Midiendo :: Caminos mínimos, diámetro de la red

- Longitud del camino mínimo entre dos nodos d_{ij}

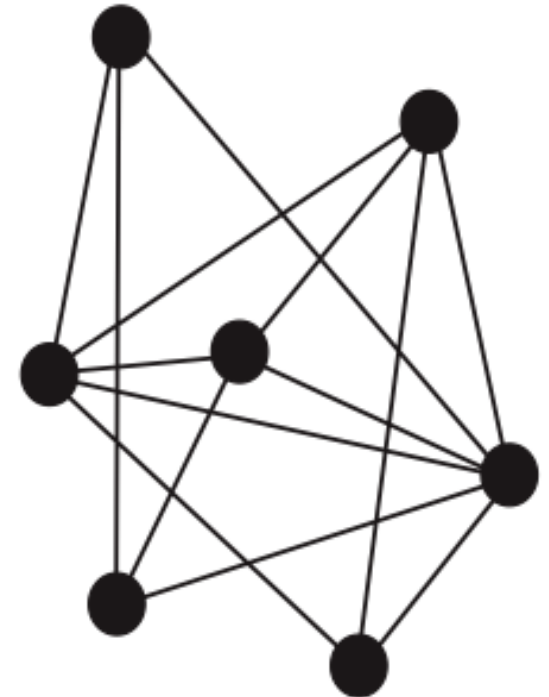
- Diámetro $D = \max_{i \neq j} d_{ij}$

- Promedio de la longitud de los caminos mínimos

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

- Eficiencia del grafo

$$E = \frac{1}{N(N-1)} \sum_{ij} \frac{1}{d_{ij}}$$



Midiendo :: Centralidad

- Identificar los “centros” de la red

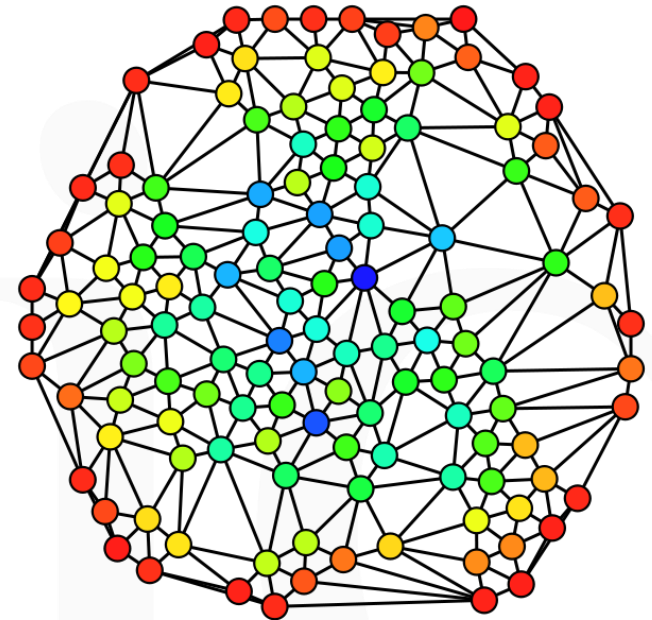
- Varias maneras de hacerlo:

- Intermediación

- Número de caminos mínimos entre j y k : n_{jk}
 - Número de caminos mínimos entre j y k que pasan por i : $n_{jk}(i)$
 - Grado de intermediación

$$b_i = \sum_{j \neq k} \frac{n_{jk}(i)}{n_{jk}}$$

- Autovectores (*PageRank*)



Ejemplos de métricas sobre redes sociales

		diámetro					clustering				
	Network	Type	n	m	z	ℓ	α	$C^{(1)}$	$C^{(2)}$	r	Ref(s).
Social	film actors	undirected	449 913	25 516 482	113.43	3.48	2.3	0.20	0.78	0.208	[20, 415]
	company directors	undirected	7 673	55 392	14.44	4.60	–	0.59	0.88	0.276	[105, 322]
	math coauthorship	undirected	253 339	496 489	3.92	7.57	–	0.15	0.34	0.120	[107, 181]
	physics coauthorship	undirected	52 909	245 300	9.27	6.19	–	0.45	0.56	0.363	[310, 312]
	biology coauthorship	undirected	1 520 251	11 803 064	15.53	4.92	–	0.088	0.60	0.127	[310, 312]
	telephone call graph	undirected	47 000 000	80 000 000	3.16		2.1				[8, 9]
	email messages	directed	59 912	86 300	1.44	4.95	1.5/2.0		0.16		[136]
	email address books	directed	16 881	57 029	3.38	5.22	–	0.17	0.13	0.092	[320]
	student relationships	undirected	573	477	1.66	16.01	–	0.005	0.001	–0.029	[45]
	sexual contacts	undirected	2 810				3.2				[264, 265]
Information	WWW nd.edu	directed	269 504	1 497 135	5.55	11.27	2.1/2.4	0.11	0.29	–0.067	[14, 34]
	WWW Altavista	directed	203 549 046	2 130 000 000	10.46	16.18	2.1/2.7				[74]
	citation network	directed	783 339	6 716 198	8.57		3.0/–				[350]
	Roget's Thesaurus	directed	1 022	5 103	4.99	4.87	–	0.13	0.15	0.157	[243]
	word co-occurrence	undirected	460 902	17 000 000	70.13		2.7		0.44		[119, 157]
Technological	Internet	undirected	10 697	31 992	5.98	3.31	2.5	0.035	0.39	–0.189	[86, 148]
	power grid	undirected	4 941	6 594	2.67	18.99	–	0.10	0.080	–0.003	[415]
	train routes	undirected	587	19 603	66.79	2.16	–		0.69	–0.033	[365]
	software packages	directed	1 439	1 723	1.20	2.42	1.6/1.4	0.070	0.082	–0.016	[317]
	software classes	directed	1 377	2 213	1.61	1.51	–	0.033	0.012	–0.119	[394]
	electronic circuits	undirected	24 097	53 248	4.34	11.05	3.0	0.010	0.030	–0.154	[155]
	peer-to-peer network	undirected	880	1 296	1.47	4.28	2.1	0.012	0.011	–0.366	[6, 353]
Biological	metabolic network	undirected	765	3 686	9.64	2.56	2.2	0.090	0.67	–0.240	[213]
	protein interactions	undirected	2 115	2 240	2.12	6.80	2.4	0.072	0.071	–0.156	[211]
	marine food web	directed	135	598	4.43	2.05	–	0.16	0.23	–0.263	[203]
	freshwater food web	directed	92	997	10.84	1.90	–	0.40	0.48	–0.326	[271]
	neural network	directed	307	2 359	7.68	3.97	–	0.18	0.28	–0.226	[415, 420]

asortatividad

Extraído de “Structure and Function of Complex Networks” M.E.J. Newman 2003

¿Qué cuestiones son relevantes en el análisis de redes sociales?

▪ Complejidad estructural

- Detección y análisis de comunidades: identificación de grupos densamente conectados.
- Capital social, centralidad: identificación de nodos que juegan papeles importantes en la estructura de la red (centros, puentes, etc.)

▪ Complejidad dinámica

- ¿Cómo fluye la información en las redes sociales?
- ¿Cómo navega la gente en la red social buscando talento, experiencia, contactos, información, etc.?
- ¿Cómo de eficiente es la red social en la diseminación de innovaciones, ideas, opiniones, etc.?

Software de Análisis de Redes Sociales

- Listas de recursos

- General

- Cytoscape: http://www.cytoscape.org/cgi-bin/moin.cgi/Network_analysis_links
 - Center for complex network research <http://www.nd.edu/~networks/resources.htm>
 - Infovis: <http://iv.slis.indiana.edu/sw/#packages>

- Visualización

- VisualComplexity: <http://www.visualcomplexity.com/vc/links.cfm>
 - Cosin Project: <http://www.cosinproject.org/>

- Software genérico

- Pajek: <http://vlado.fmf.uni-lj.si/pub/networks/pajek/>
 - UCINET: <http://www.analytictech.com/downloaduc6.htm>
 - DyNet: <http://www.atalab.com/software/>

- Visualización

- Graphviz: <http://www.graphviz.org/>
 - Walrus (Caida): <http://www.caida.org/tools/visualization/walrus/>
 - JGraph: <http://www.jgraph.com/>
 - InFlow: <http://www.orgnet.com>

References

▪ Introductory (general audience)

- M. Buchanan, *Nexus: Small Worlds and the Groundbreaking Science of Networks*, Norton, New York, 2002.
- A.-L. Barabási, *Linked: The New Science of Networks*, Perseus, Cambridge, 2002.
- D.J. Watts, *Six Degrees: The Science of a Connected Age*, Norton, New York, 2003.

▪ Reviews on Social Networks

- [1] S.H. Strogatz, *Nature* 410 (2001) 268.
- [2] R. Albert, A.-L. Barabási, *Rev. Mod. Phys.* 74 (2002) 47.
- [3] S.N. Dorogovtsev, J.F.F. Mendes, *Adv. Phys.* 51 (2002) 1079.
- [4] M.E.J. Newman, *SIAM Rev.* 45 (2003) 167.
- [5] D.J. Watts, *Small Worlds: The Dynamics of Networks between Order and Randomness*, Princeton University Press, Princeton, NJ, 1999.
- [6] S. Bornholdt, H.G. Schuster (Eds.), *Handbook of Graphs and Networks: From the Genome to the Internet*, Wiley-VCH, Germany, 2003.
- [7] S.N. Dorogovtsev, J.F.F. Mendes, *Evolution of Networks*, Oxford University Press, Oxford, 2003.
- [8] R. Pastor-Satorras, A. Vespignani, *Evolution and Structure of the Internet: A Statistical Physics Approach*, Cambridge University Press, Cambridge, 2004.
- [9] R. Pastor-Satorras, M. Rubi, A. Diaz-Guilera (Eds.), *Statistical Mechanics of Complex Networks*, Springer, Berlin, 2003.
- [10] E. Ben-Naim, H. Frauenfelder, Z. Toroczkai (Eds.), *Complex Networks*, Springer, Berlin, 2004.

Referencias

▪ Sobre Análisis de Redes Sociales

- S. Wasserman, K. Faust, Social Networks Analysis, Cambridge University Press, Cambridge, 1994.
- J. Scott, Social Network Analysis: A Handbook, 2nd ed., Sage Publications, London, 2000.

▪ Algoritmos de grafos

- T.H. Cormen, C.E. Leiserson, R.L. Rivest, C. Stein, Introduction to Algorithms, MIT University Press, Cambridge, 2001.
- R. Sedgwick, Algorithms in C++. Part 5: Graph Algorithms, Addison-Wesley, Boston MA, 1988.
- R.K. Ahuja, T.L. Magnati, J.B. Orlin, Network Flows: Theory, Algorithms, and Applications, Prentice-Hall, Englewood Cliffs, NJ, 1993.