

Web Mining ed Analisi delle Reti Sociali


Introduzione all'Analisi delle Reti Sociali

Dino Pedreschi
Dipartimento di Informatica
Università di Pisa
www.di.unipi.it/~pedre

Materialie didattico

- Slides by Jiawei Han, Univ. of Illinois at Urbana-Champaign
- M. E. J. Newman, *The structure and function of complex networks*, SIAM Review, Vol. 45, p. 167-256, 2003.
- Chapter 9.2 of the book: Jiawei Han and Micheline Kamber. *Data Mining: Concepts and Techniques*, 2nd ed. Morgan Kaufmann Publishers, 2006.
- Consultazione: Duncan J. Watts. *Six Degrees: The Science of a Connected Age*. (Norton, New York, 2003).

Social Network Analysis

- Social Network Introduction 
- Statistics and Probability Theory
- Models of Social Network Generation
- Networks in Biological System
- Mining on Social Network
- Summary

Society

Nodes: individuals

Links: social relationship
(family/work/friendship/etc.)



S. Milgram (1967)

John Guare

Six Degrees of Separation

Social networks: Many individuals with diverse social interactions between them.

Social networks: Actor Connectivity



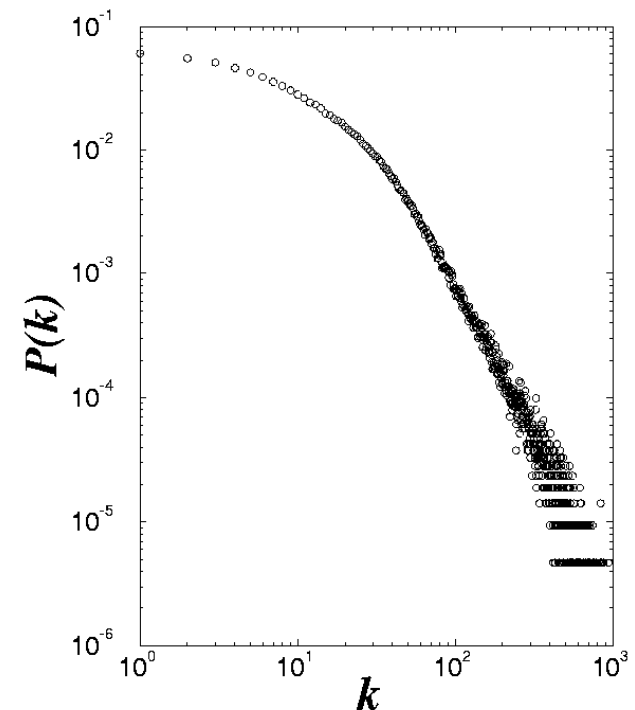
← →

Days of Thunder (1990)
Far and Away (1992)
Eyes Wide Shut (1999)

N = 212,250 actors
 $\langle k \rangle = 28.78$

$P(k) \sim k^{-\gamma}$
 $\gamma = 2.3$

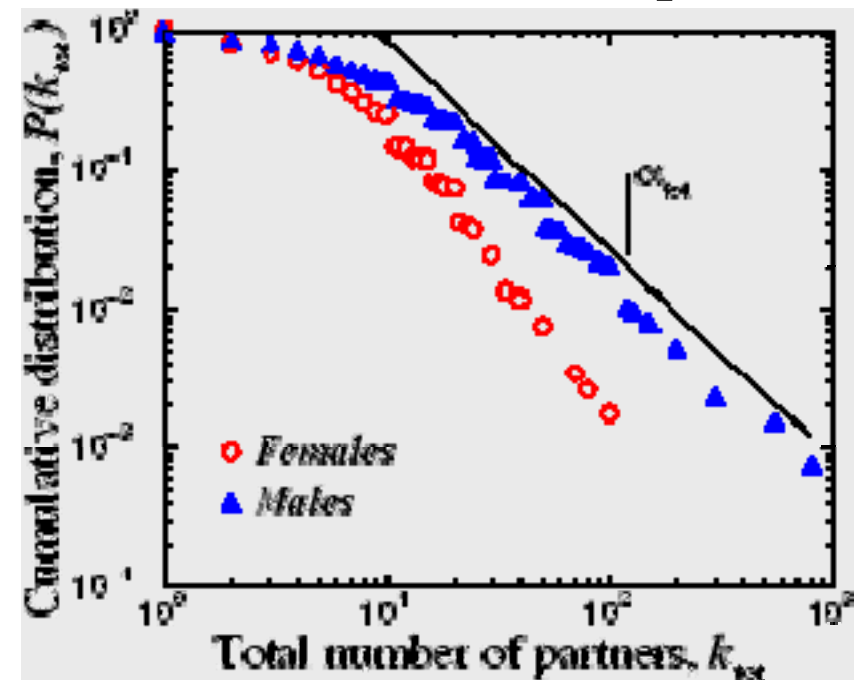
Nodes: actors
Links: cast jointly



Social networks: Sex-Web



Nodes: people (Females; Males)
Links: sexual relationships



4781 Swedes; 18-74;
59% response rate.
Liljeros et al. Nature 2001

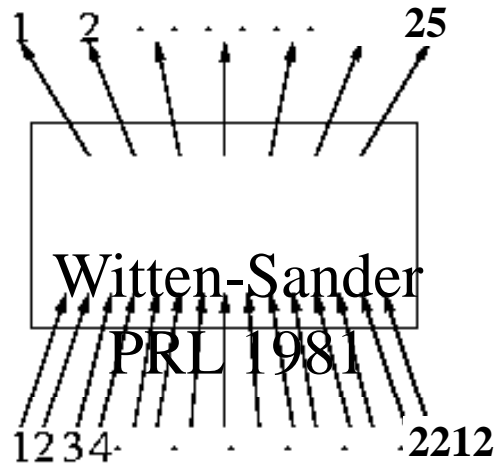
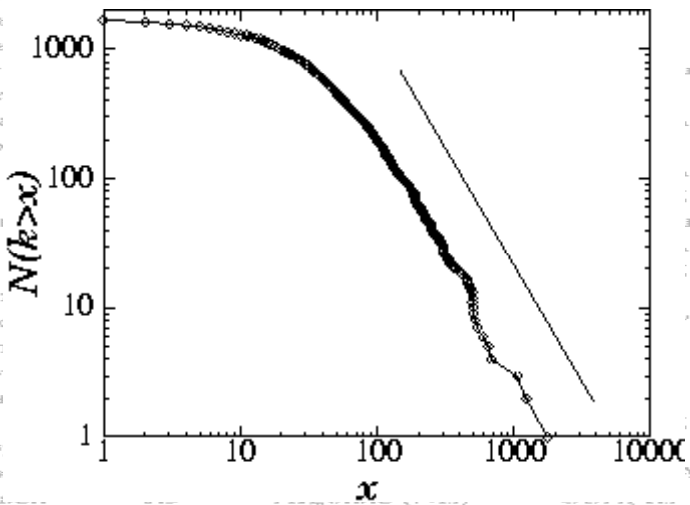
Information networks: Science Citation Index

1,000 Most Cited Physicists
Out of over 500,000 E
(see <http://www.sst.nyu.edu>)

Author name	Institution	Country	Field
Witten	MIT (U)	USA, NJ	High
Gossard	UCSB (U)	USA, CA	Sem
Cava	UCSB (U)	USA, NJ	Sup
Ballogg	UCSB (U)	USA, NJ	Sup
Ploog	Max-Planck (NL)	Germany	Sem
Ellis	Euro Nuclear Cent.	Switzerland	Astr
Fisk	Florida State (U)	USA, FL	Solid
Cardona	Max Planck (NL)	Germany	Sem
Nanopoulos	Texas A&M (U)	USA, TX	High
Heeger	UCSB (U)	USA, CA	Poly
Lee*			
Suzuki*			
Anderson		NJ	Solid
Suzuki*	M		
Freeman			
Tanaka			
Muller			
Schnee			
Chernik			
Morko			
Miller			
Chu			
Bednorz			
Cohen			
Metzger			
Waszczykowski			
Shirane			
Wiegmann			
Vando			
Uchida			
Hor			
Murphy			
Birgen			
Jorgensen			
Hinks	DG	Argonne (NL)	USA, IL

Nodes: papers
Links: citations

1736 PRL papers (1988)



	rank by total cit.				
	1				
	2				
	3				
	4				
	5				
	6				
	7				
	8				
	9				
	10				
	11				
	12				
	13				
	14		898	10417	
	15	Solid State (T)	27	389	10411
	16		11	963	10404
	17	nd Superconductivity (E)	82	122	10049
	18	Superconductivity (E)	63	156	9768
	19	Optics (E)	60	162	9668
	19	Semiconductors (E)	20	477	9668
	21	Semiconductors (E)	67	174	9652
	22	Superconductivity (E)	54	313	9453
	23	nd Superconductivity (E)	110	85	9311
	23	Solid State (T)	33	284	9311
	25	Superconductivity (E)	86	108	9300
	26	Superconductivity (E)	57	162	9170
	27	Superconductivity (E)	39	269	8841
	28	Semiconductors (E)	8	104	8822
	29	Magnetism (E)	67	129	8686
	30		28	301	8520
	31	Superconductivity (E)	72	119	8512
	32	Astronomy (E)	111	76	8439
	33	Superconductivity (E)	105	85	8375
	34	Superconductivity (E)	107	82	8288
	35	Superconductivity (E)	37	223	8263

$P(k) \sim k^{-\gamma}$
($\gamma = 3$)

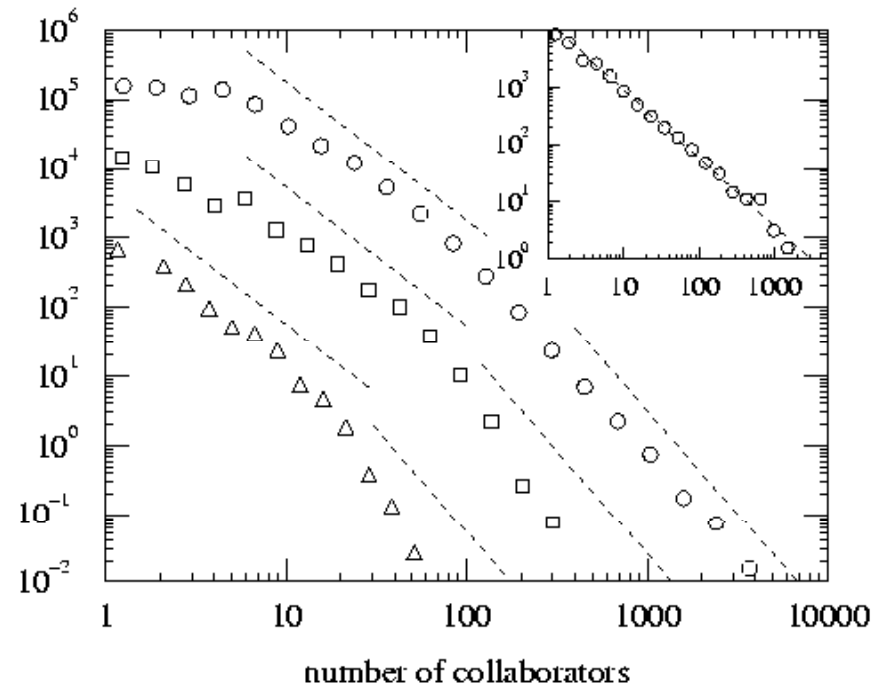
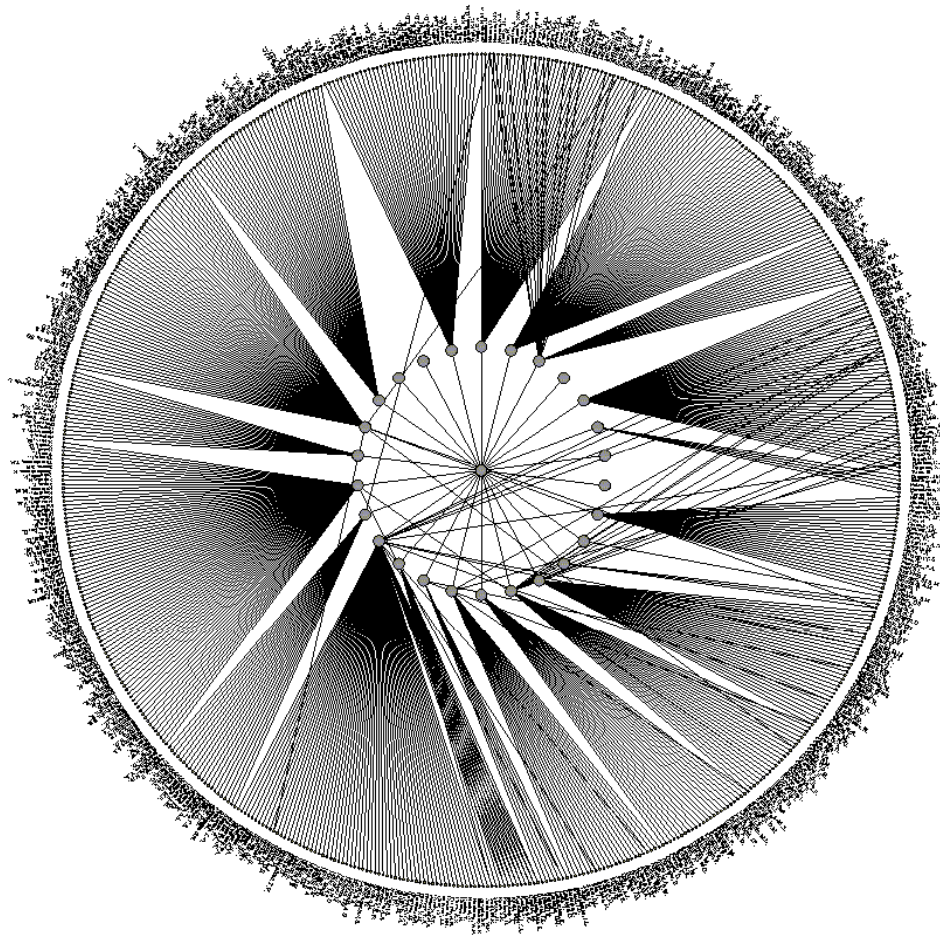
(S. Redner, 1998)

* citation total may be skewed because of multiple authors with the same name

Information network: Science Coauthorship

Nodes: scientist (authors)

Links: write paper together



(Newman, 2000, H. Jeong et al 2001)

Communication networks

The Earth is developing an electronic nervous system, a network with diverse nodes and links are

-computers

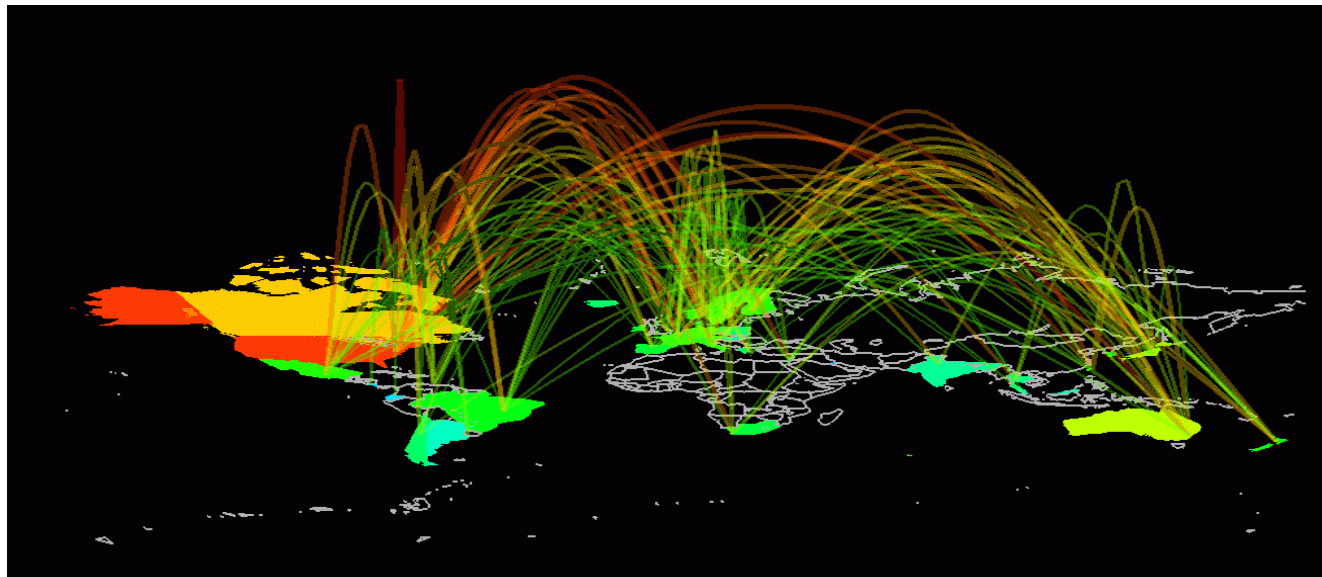
-routers

-satellites

-phone lines

-TV cables

-EM waves

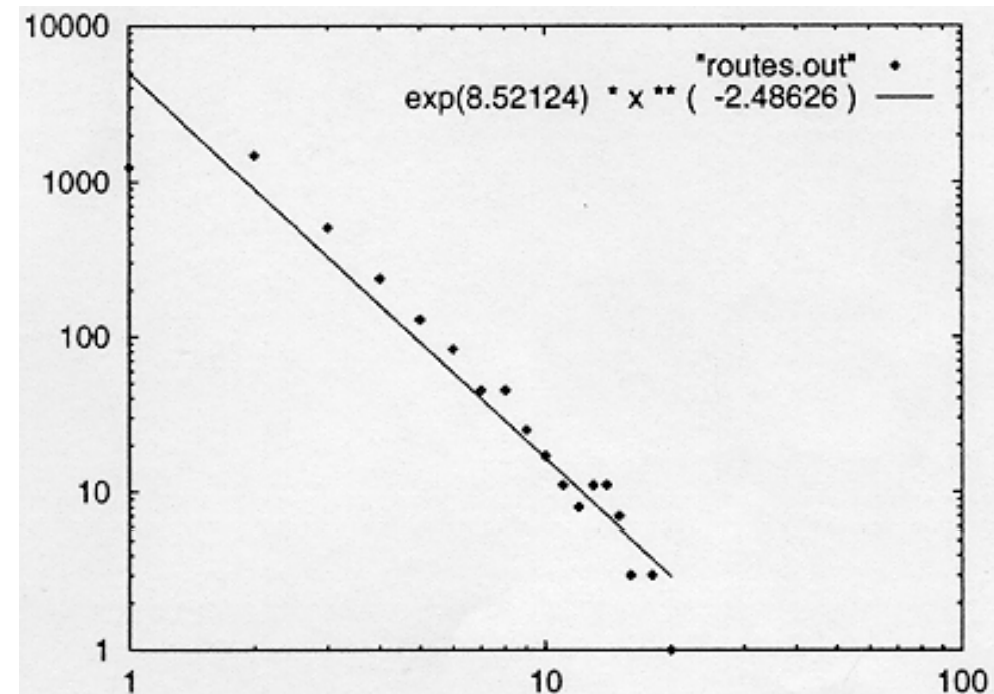
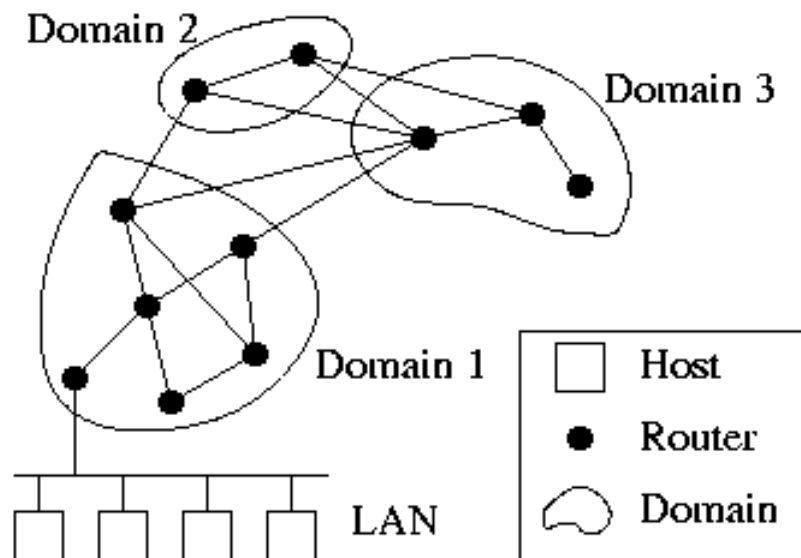


Communication networks: Many non-identical components with diverse connections between them.

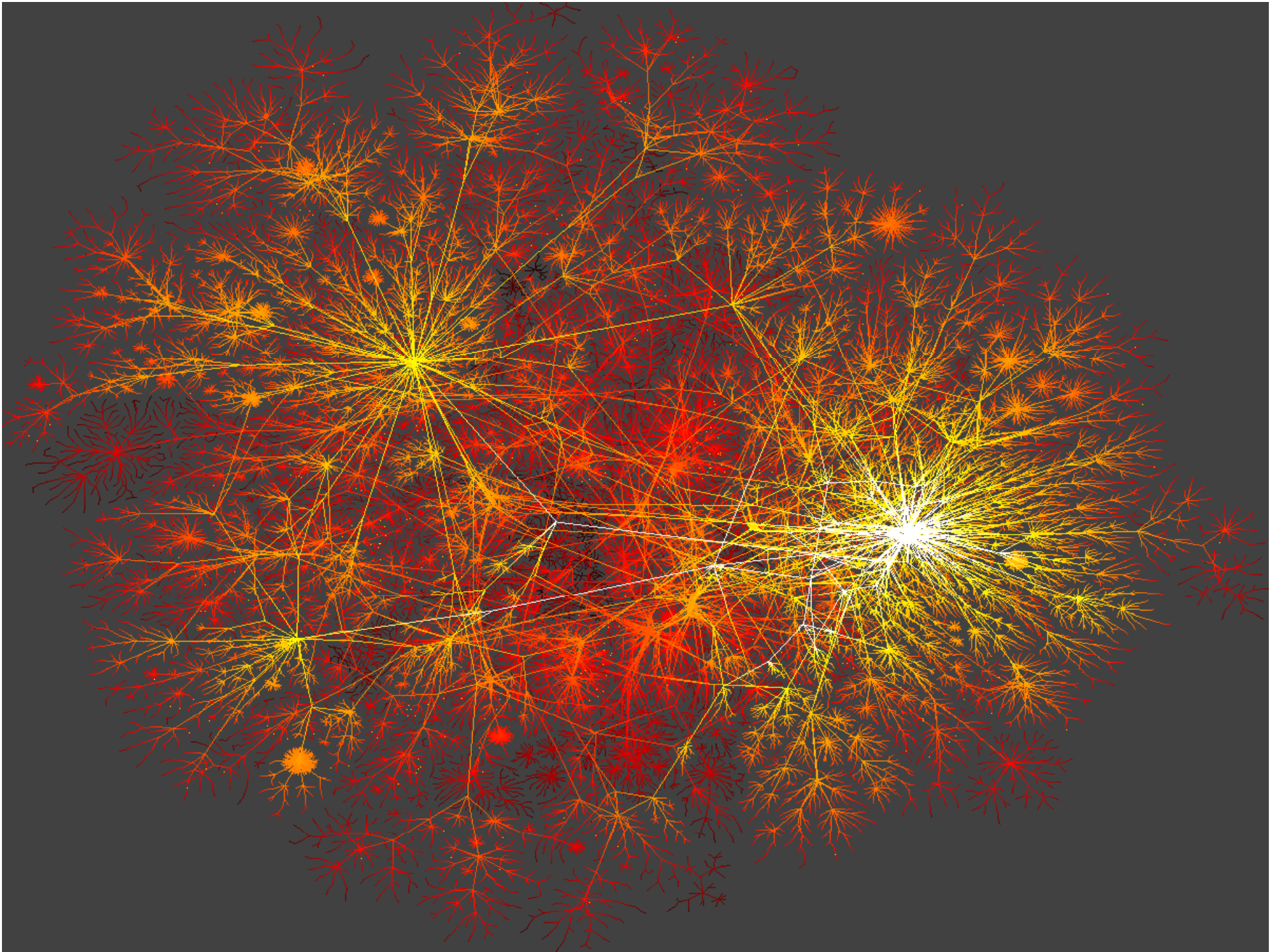
Tech. networks: Internet Backbone

Nodes: computers, routers

Links: physical lines

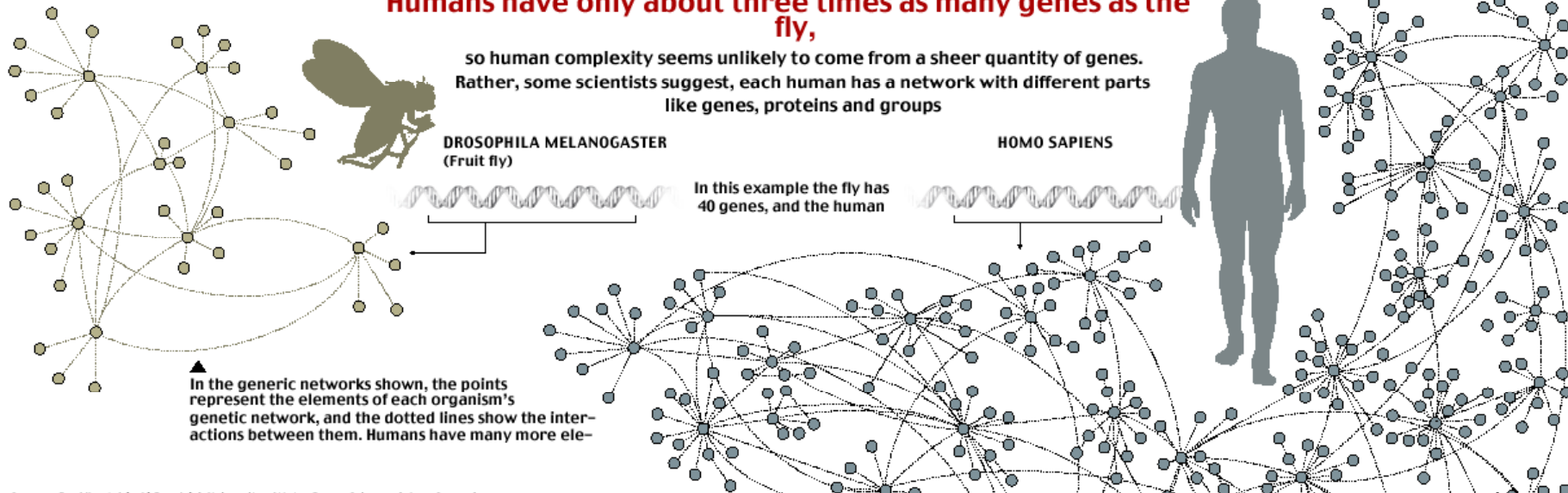


(Faloutsos, Faloutsos and Faloutsos, 1999)



Humans have only about three times as many genes as the fly,

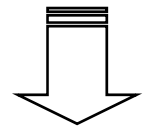
so human complexity seems unlikely to come from a sheer quantity of genes. Rather, some scientists suggest, each human has a network with different parts like genes, proteins and groups



Sources: Dr. Albert-László Barabási, University of Notre Dame; Science; Celera Genomics

Complex systems

Made of many non-identical **elements** connected by diverse **interactions**.

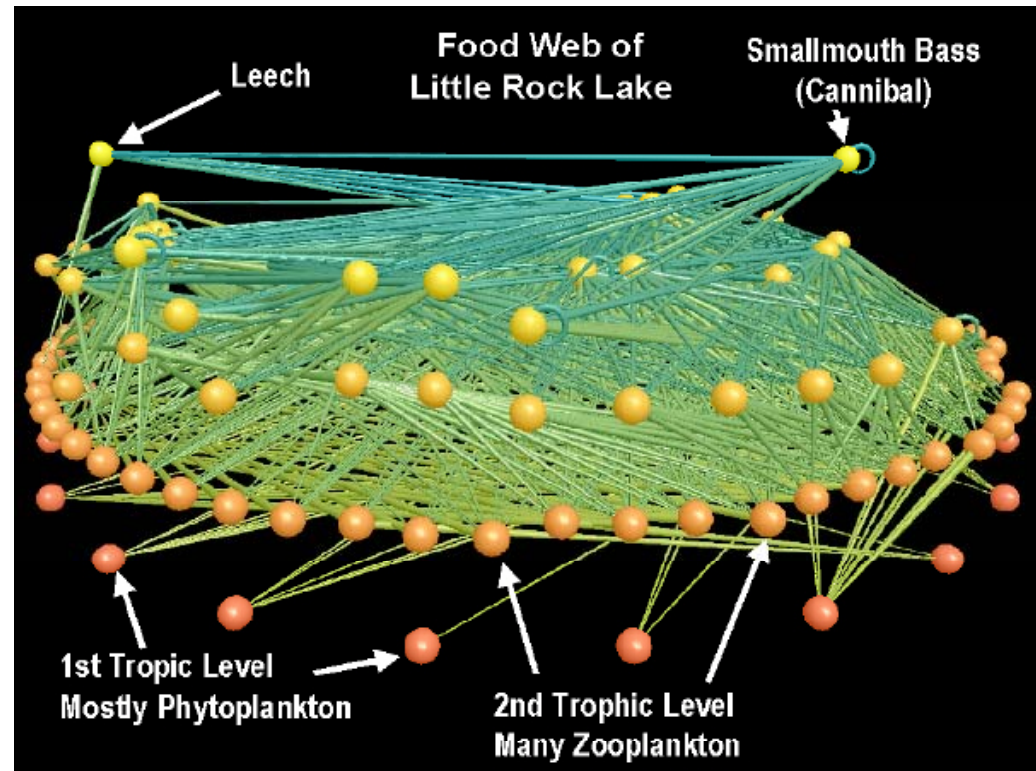
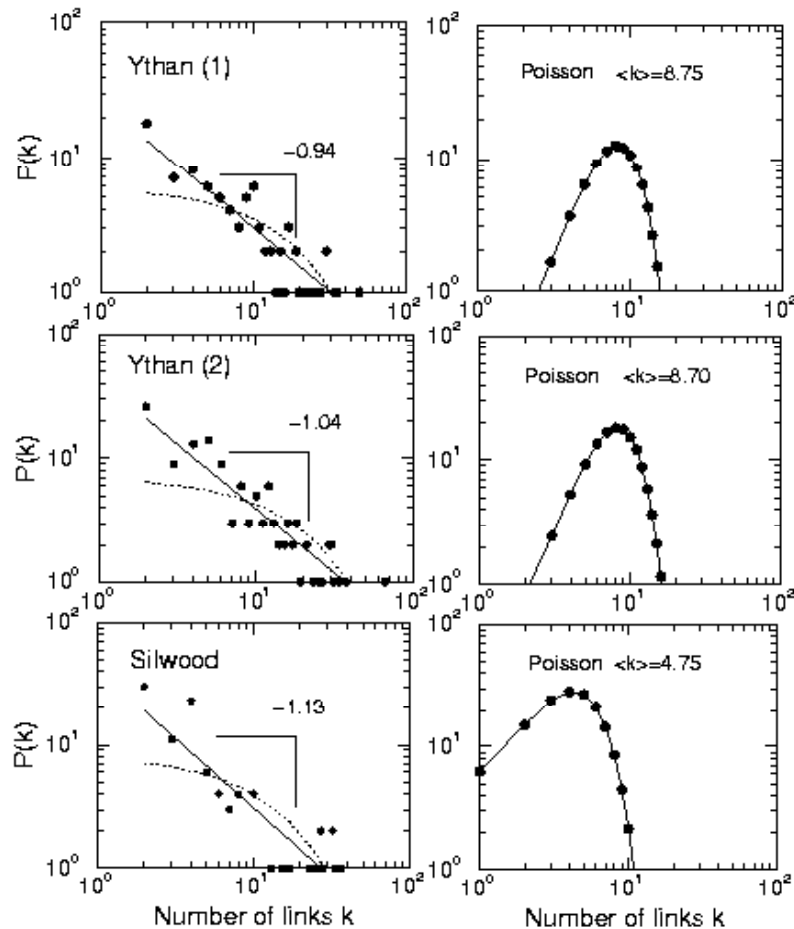


BIOLOGICAL NETWORKS

Biological networks: Food Web

Nodes: trophic species

Links: trophic interactions



R. Sole (cond-mat/0011195)

R.J. Williams, N.D. Martinez *Nature* (2000)

“Natural” Networks and Universality

- Consider many kinds of networks:
 - social, technological, business, economic, content,...
- These networks tend to share certain *informal* properties:
 - large scale; continual growth
 - distributed, organic growth: vertices “decide” who to link to
 - interaction restricted to links
 - mixture of local and long-distance connections
 - abstract notions of distance: geographical, content, social,...
- Do natural networks share more *quantitative* universals?
- What would these “universals” be?
- How can we make them precise and measure them?
- How can we explain their universality?
- This is the domain of *social network theory*
- Sometimes also referred to as *link analysis*