

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

Materiale 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

EVERY SAGA HAS A BEGINNING

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

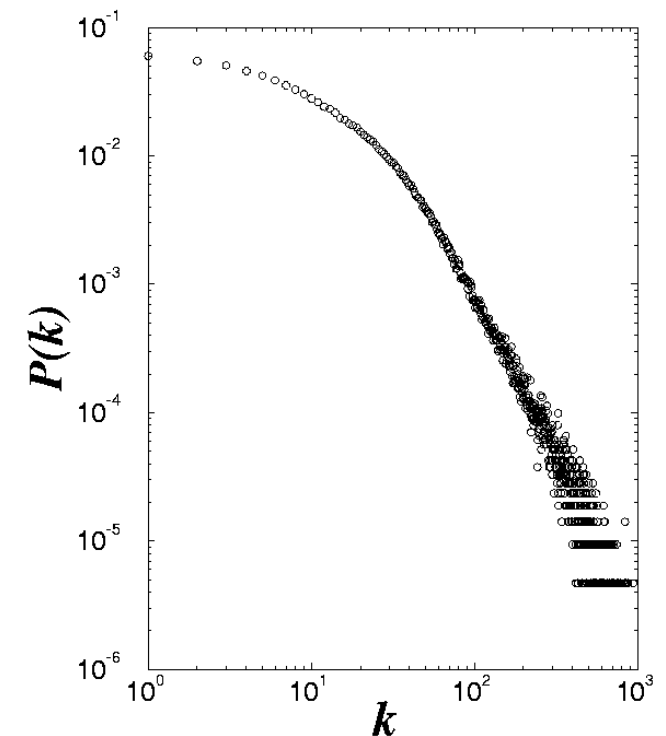
$N = 212,250$ actors
 $\langle k \rangle = 28.78$

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

STAR WARS
EPISODE I
THE PHANTOM MENACE

Nodes: actors

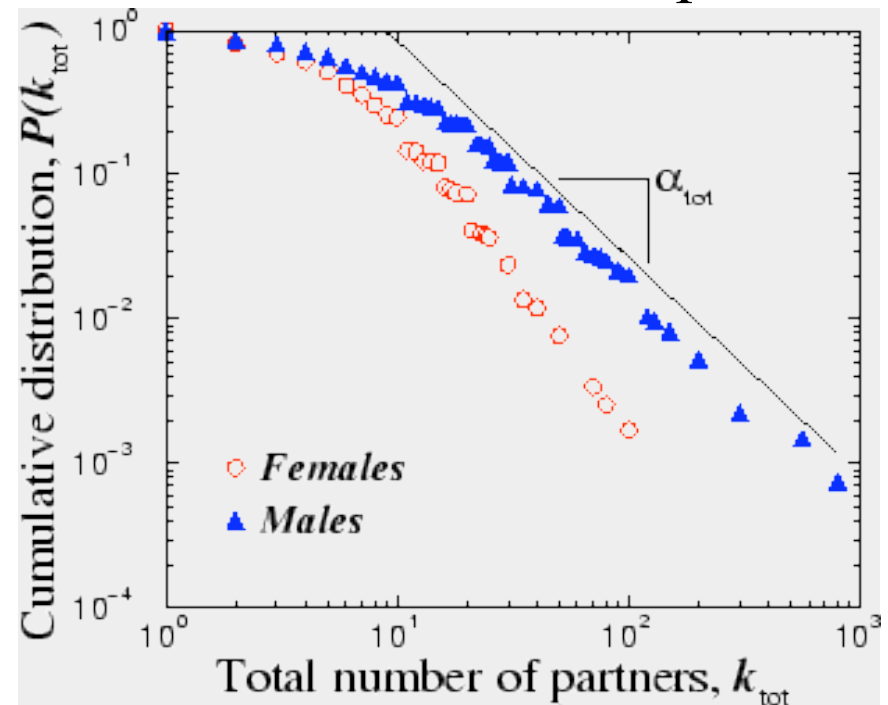
Links: cast jointly



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

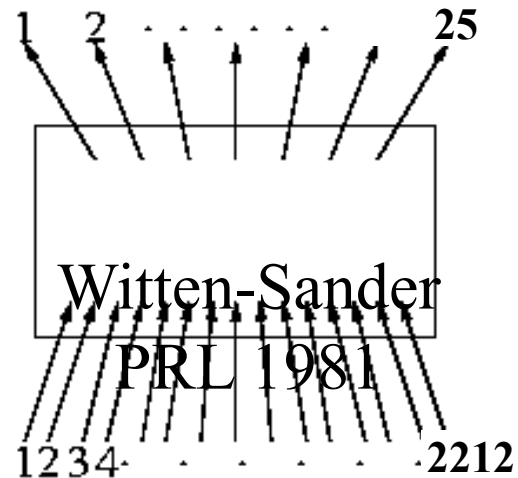
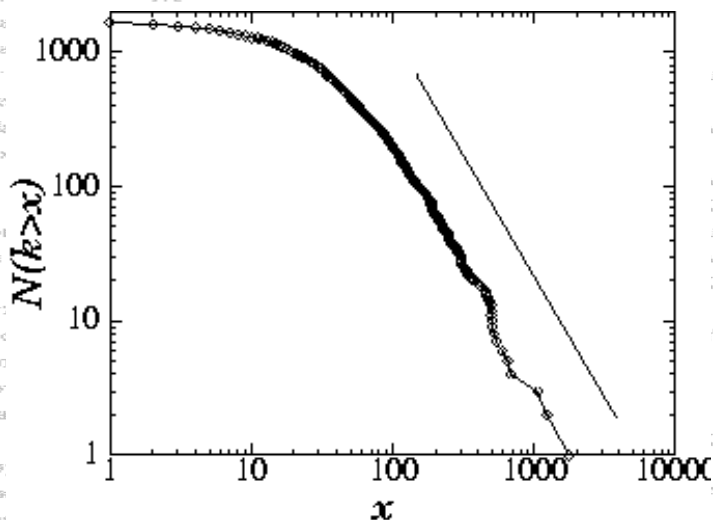
1,000 Most Cited Physicists
Out of over 500,000 E
(see <http://www.sst.nyu.edu>)

| Author name | Institution | Country | Field |
|---------------|--------------------|-------------|-------------------------|
| Witten | Princeton (U) | USA, NJ | High Energy Physics (E) |
| Gossard | UCSB (U) | USA, CA | Semiconductors (E) |
| Cava | Rutgers (U) | USA, NJ | Superconductivity (E) |
| Barlogg | Berkeley (U) | USA, NJ | Superconductivity (E) |
| Ploog | Max-Planck (NL) | Germany | Semiconductors (E) |
| Ellis | Euro Nuclear Cent. | Switzerland | Astronomy (E) |
| Fisk | Florida State (U) | USA, FL | Solid State (T) |
| Cardona | Max Planck (NL) | Germany | Semiconductors (E) |
| Nanopoulos | Texas A&M (U) | USA, TX | High Energy Physics (E) |
| Heeger | UCSB (U) | USA, CA | Polymer Physics (E) |
| Lee* | | | |
| Suzuki* | | | |
| Anderson | | NJ | Solid State (T) |
| Suzuki* | | | |
| Freeman | | | Solid State (T) |
| Tanaka | | | Solid State (T) |
| Muller | | | Superconductivity (E) |
| Schnee | | | Superconductivity (E) |
| Chen | | | Optics (E) |
| Morko | | | Semiconductors (E) |
| Miller | | | Semiconductors (E) |
| Chu | | | Superconductivity (E) |
| Bednorz | | | Superconductivity (E) |
| Cohen | | | Solid State (T) |
| Meng | | | Superconductivity (E) |
| Waszczykowski | | | Superconductivity (E) |
| Shirane | | | Superconductivity (E) |
| Wiegmann | | | Semiconductors (E) |
| Vanderbilt | | | Magnetism (E) |
| Uchida | | | Superconductivity (E) |
| Hor | | | Superconductivity (E) |
| Murphy | | | Astronomy (E) |
| Birgen | | | Superconductivity (E) |
| Jorgensen | | | Superconductivity (E) |
| Hinks | DG Argonne (NL) | USA, IL | Superconductivity (E) |

Nodes: papers

Links: citations

1736 PRL papers (1988)



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

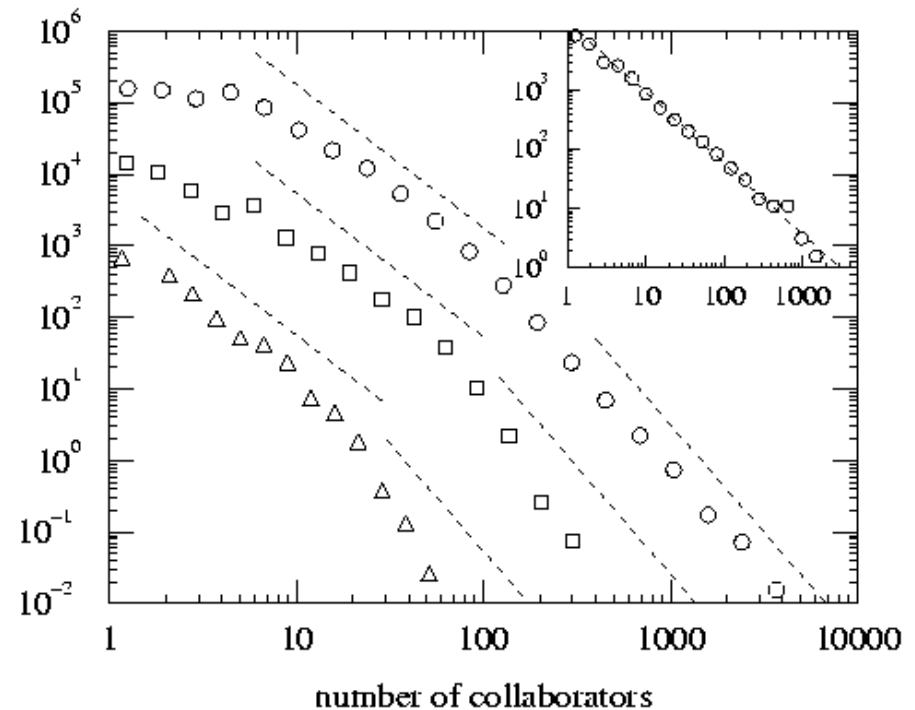
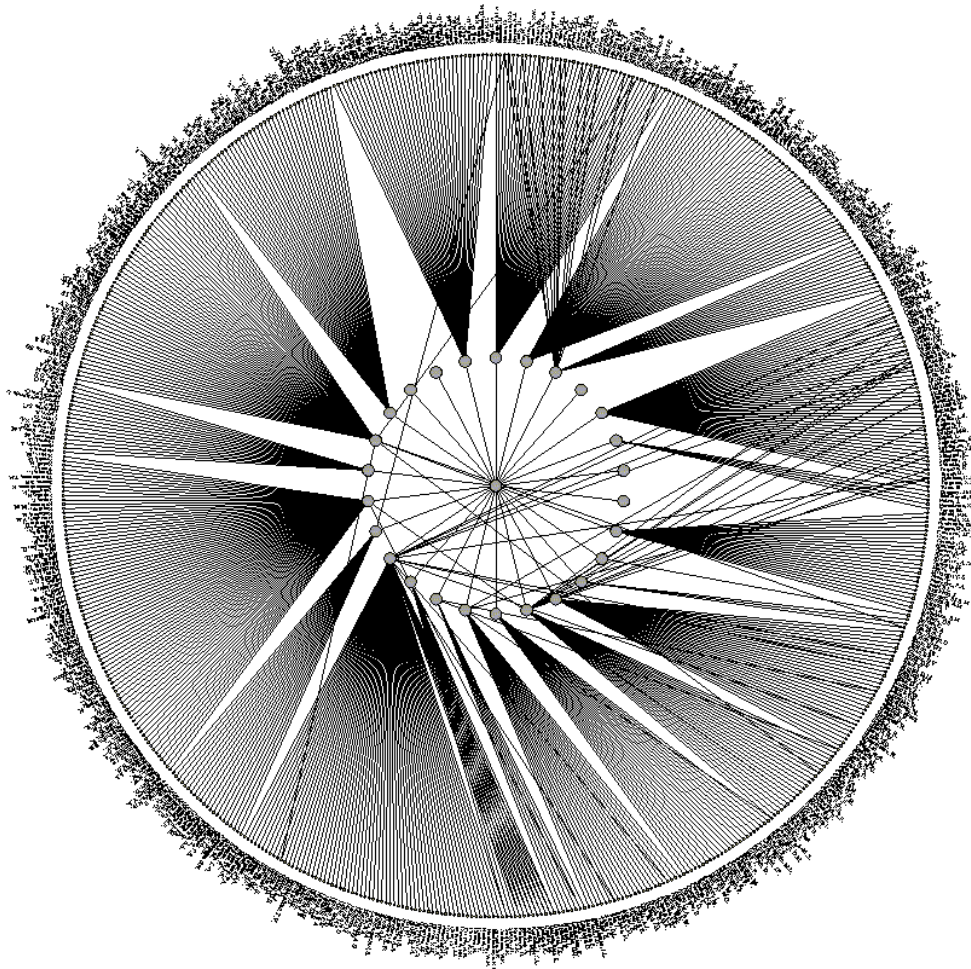
(S. Redner, 1998)

* 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

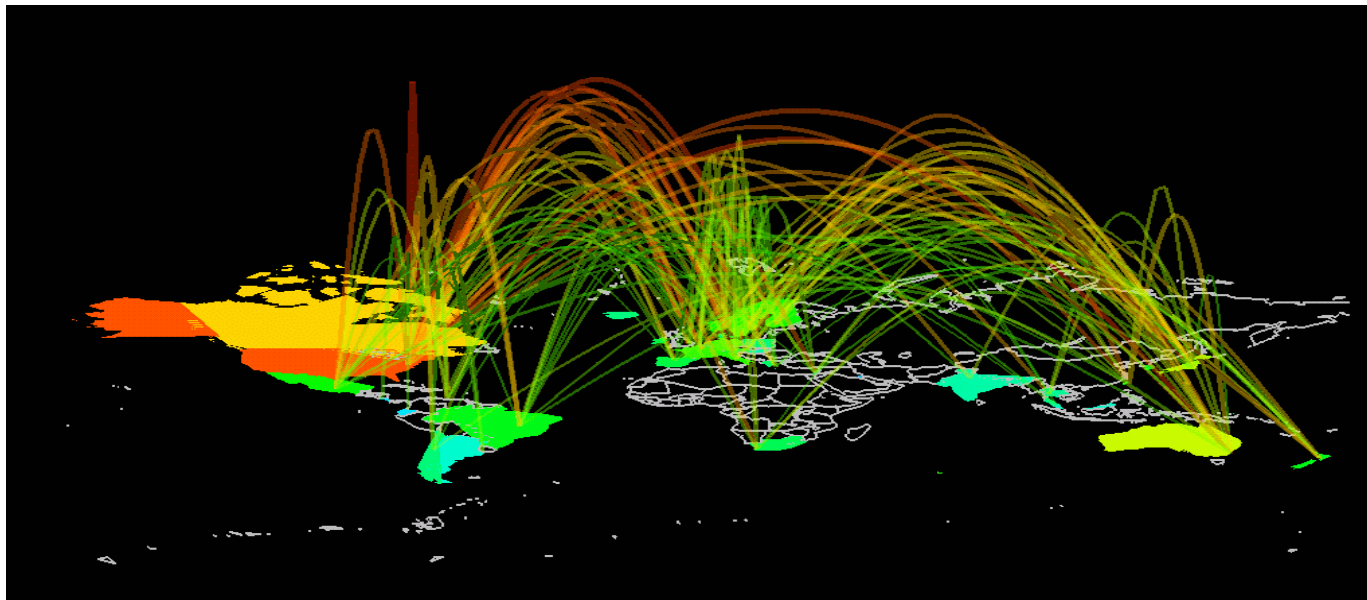
-routers

-satellites

↓ -phone lines

-TV cables

-EM waves

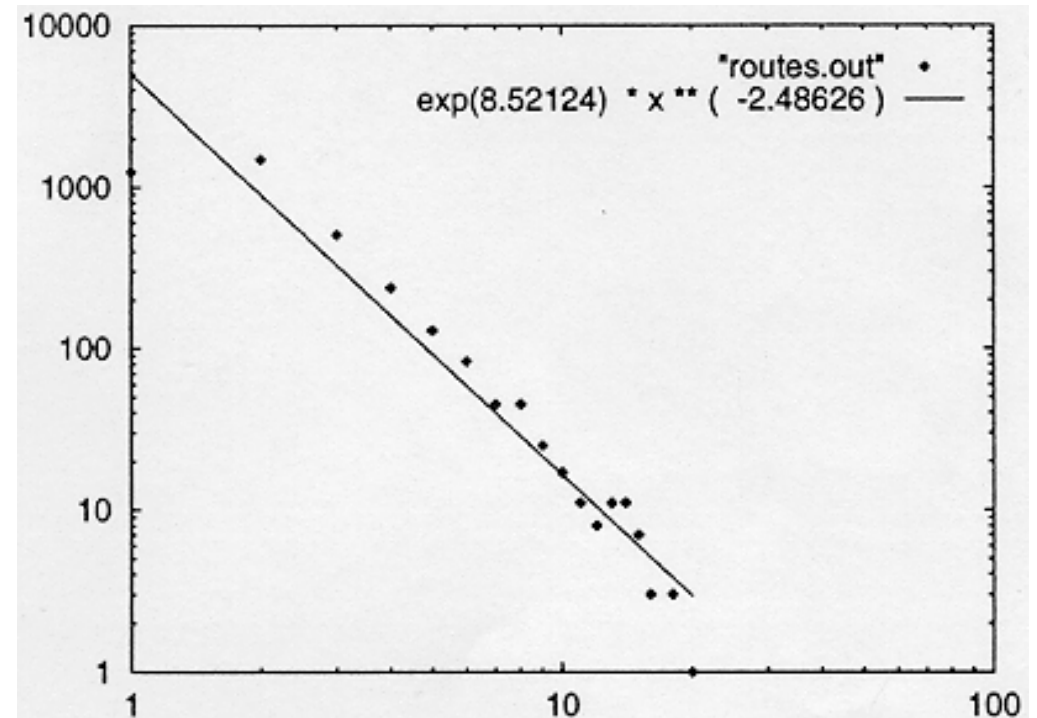
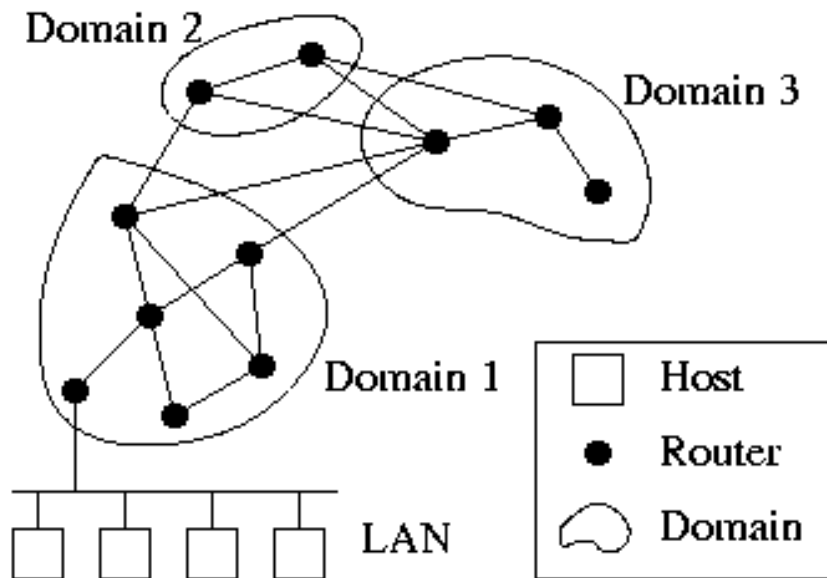


Communication networks: Many non-identical components with diverse connections between them.

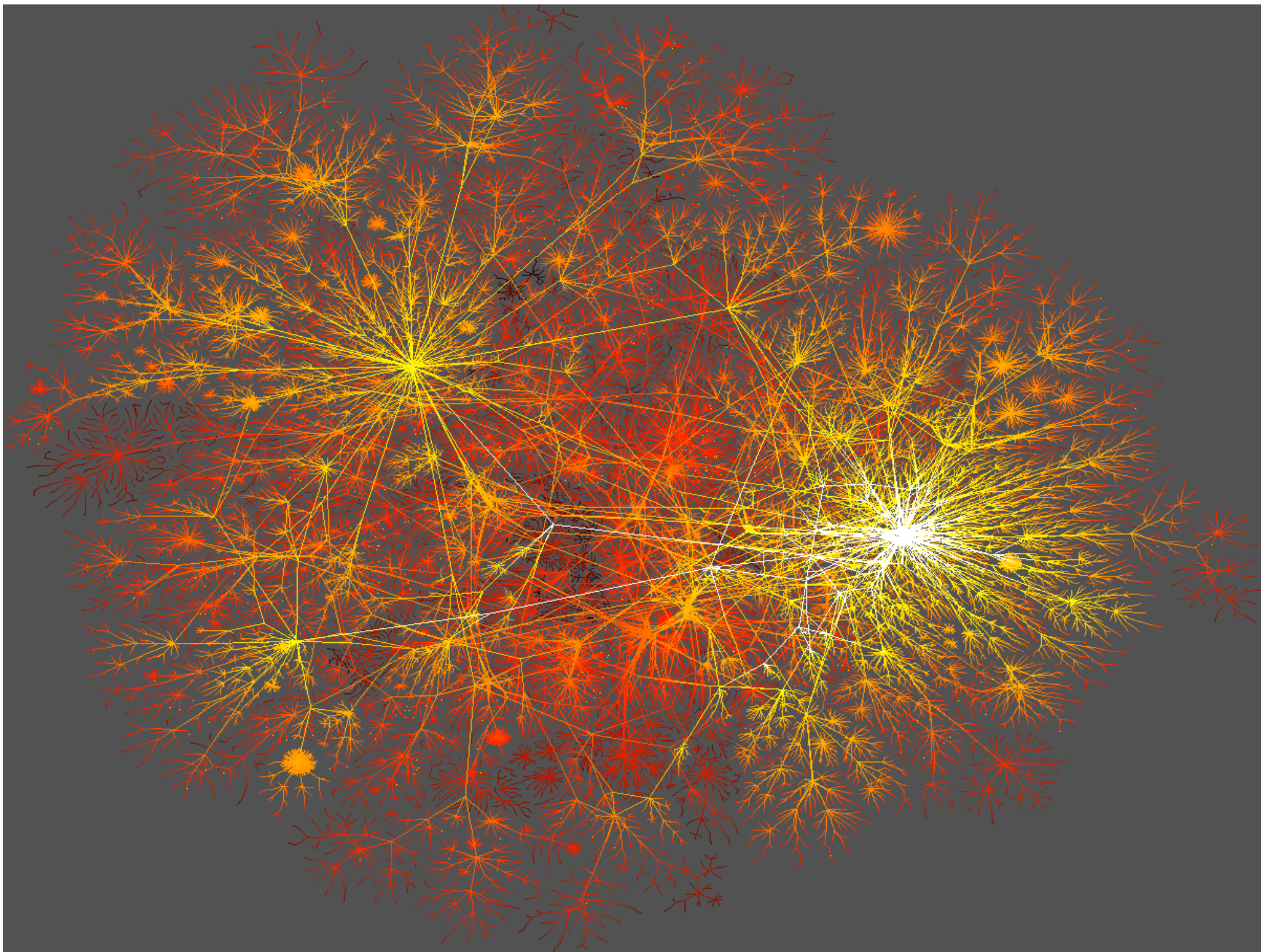
Tech. networks: Internet Backbone

Nodes: computers, routers

Links: physical lines

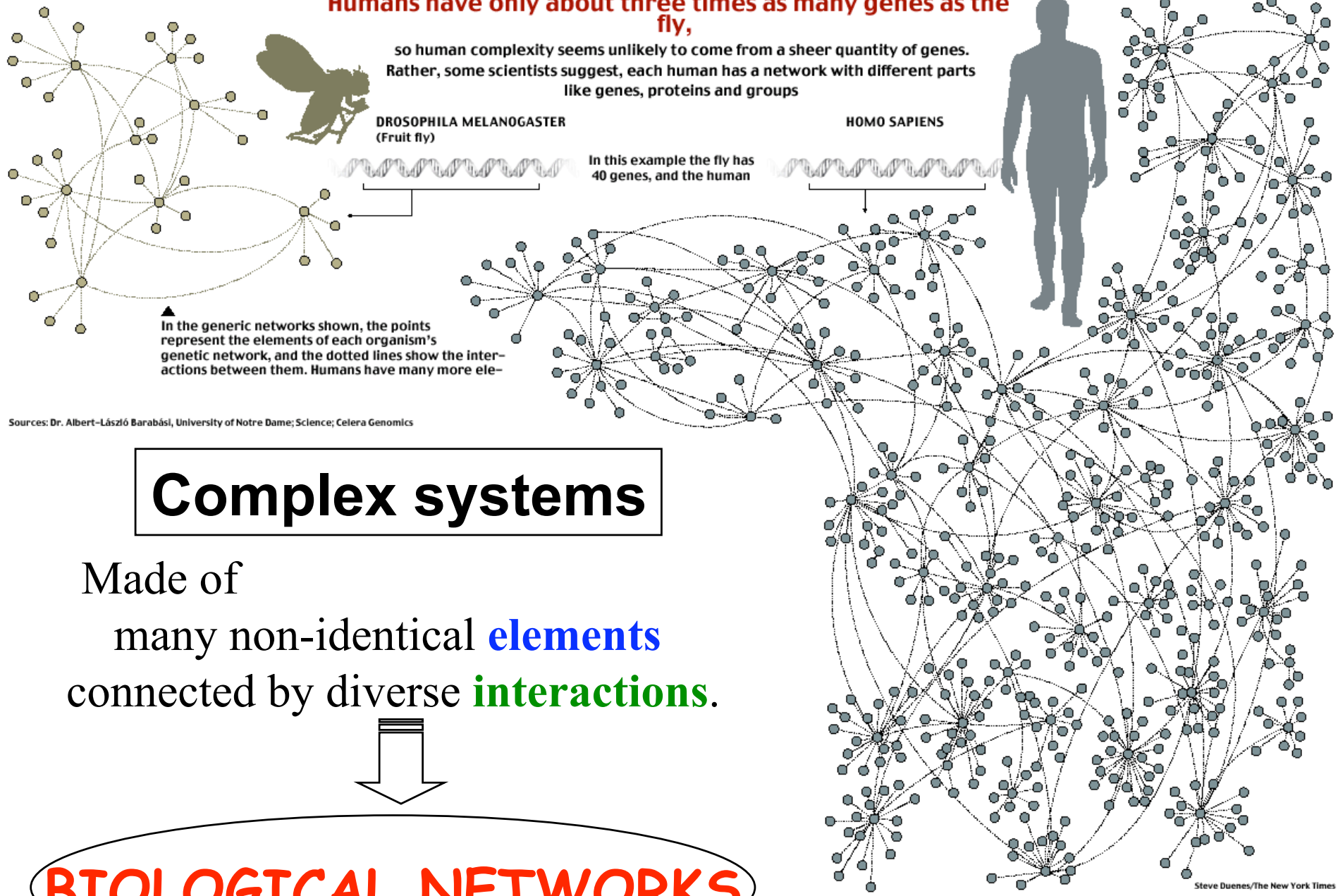


(Faloutsos, Faloutsos and Faloutsos, 1999)



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

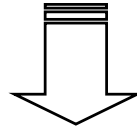
so human complexity seems unlikely to come from a sheer quantity of genes. Rather, some scientists suggest, each human has a network with different parts like genes, proteins and groups



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

Complex systems

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

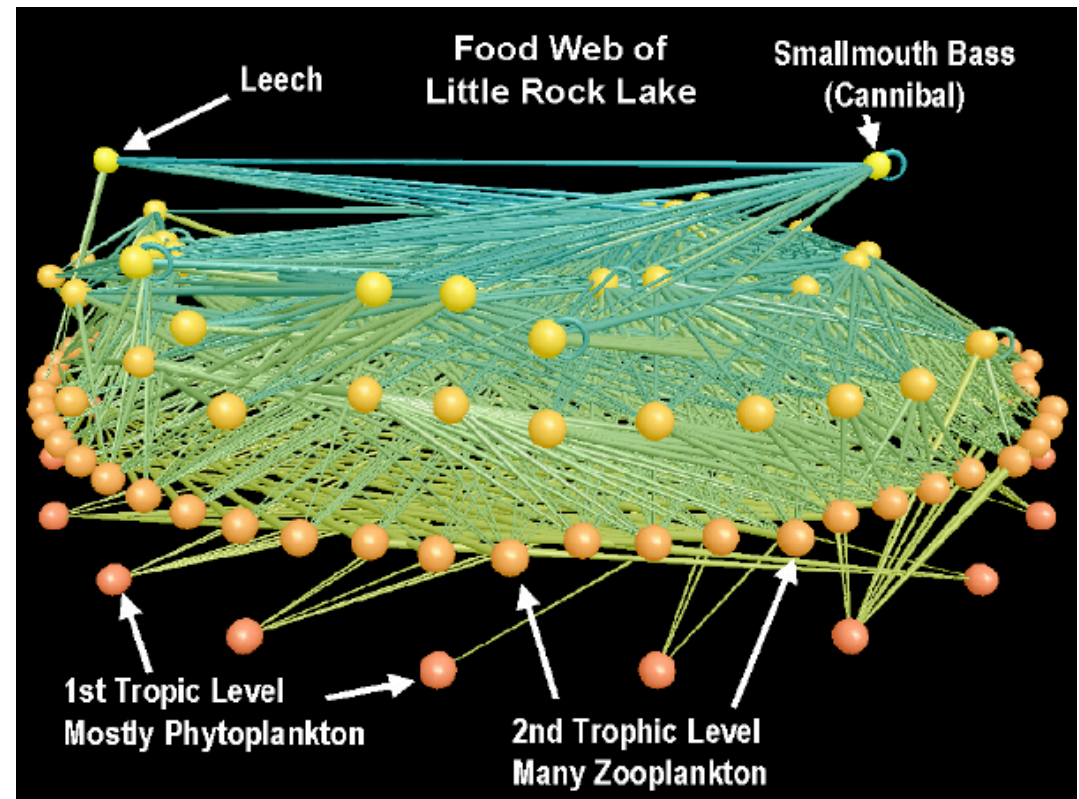
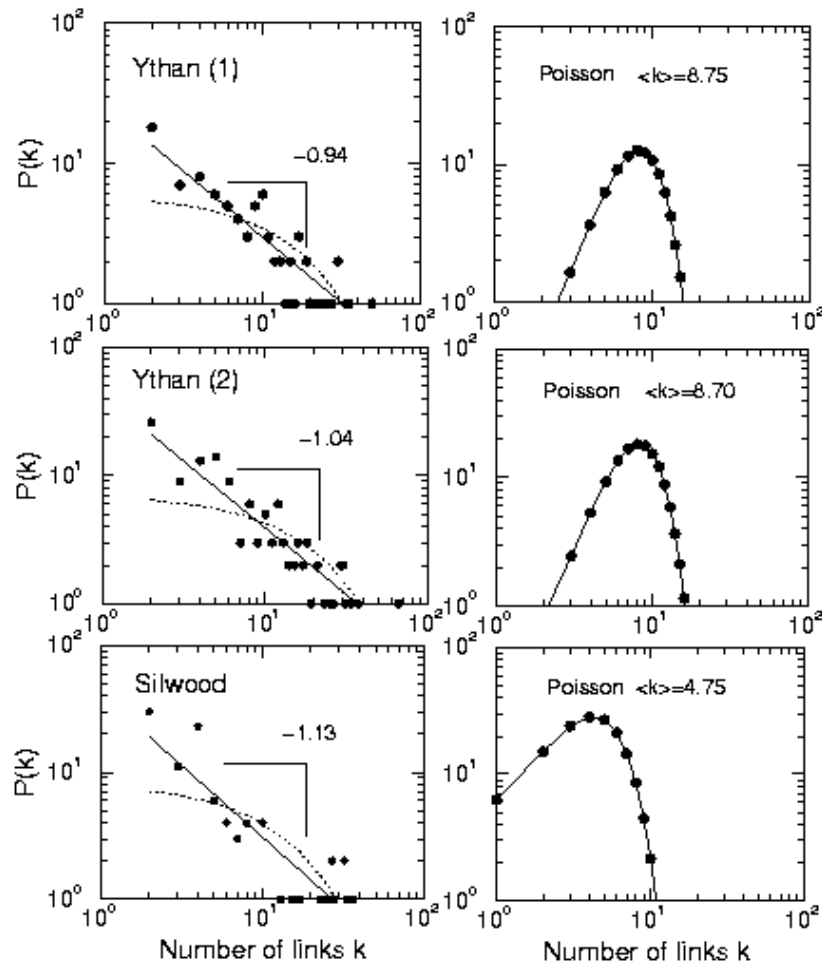


BIOLOGICAL NETWORKS

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*