

Citation Networks

Filippo Radicchi, Santo Fortunato,
Alessandro Vespignani

Outline

★ General features:

- a) Degree distribution
- b) Community structure
- c) Other features

★ Models:

- a) Dynamical models: Price and his brothers
- b) Static models

★ Dynamics on citation networks

Networks of scientific papers ...

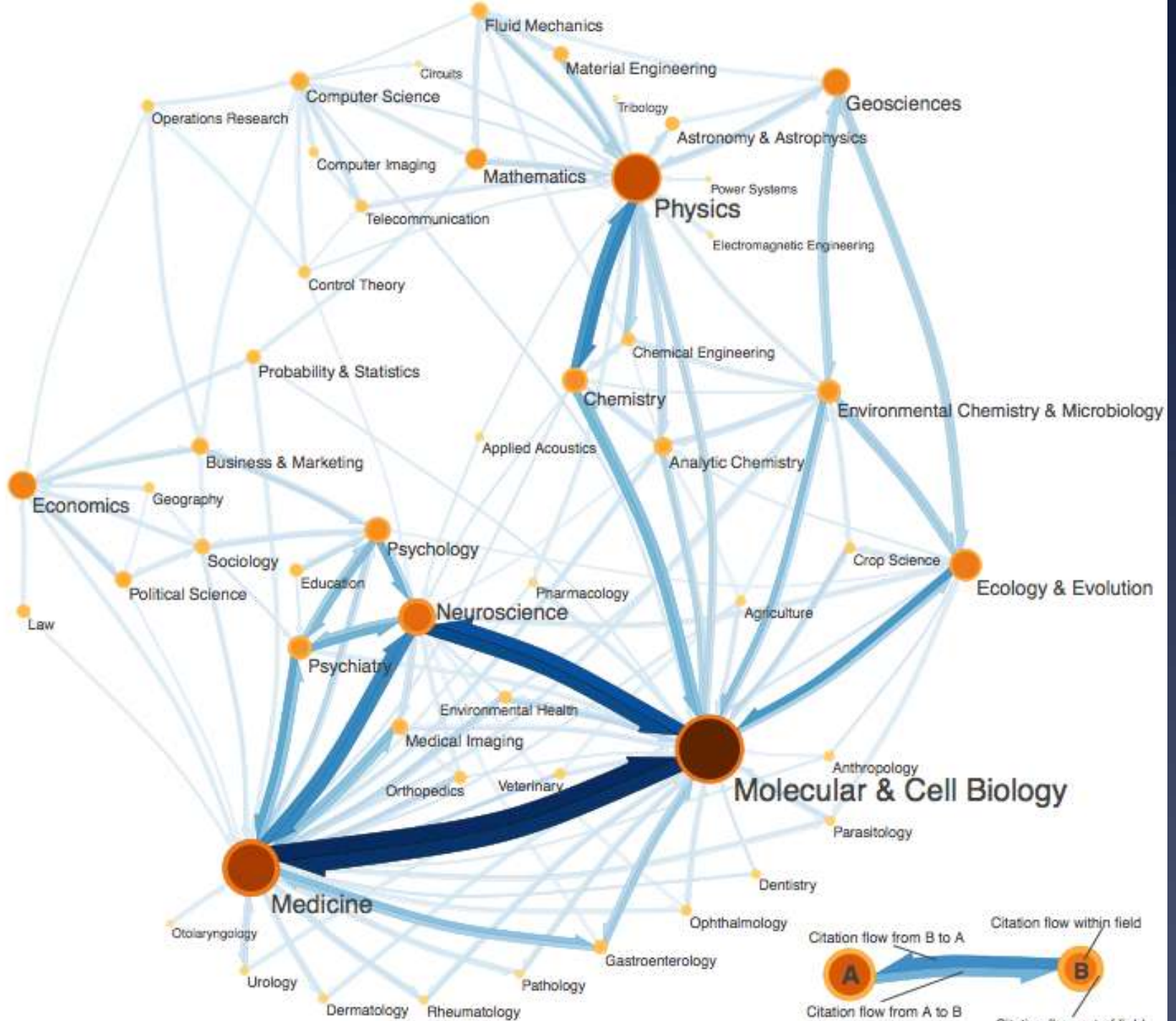
D. de Solla Price, *Networks of scientific papers*, *Science* **149**, 510 (1965)

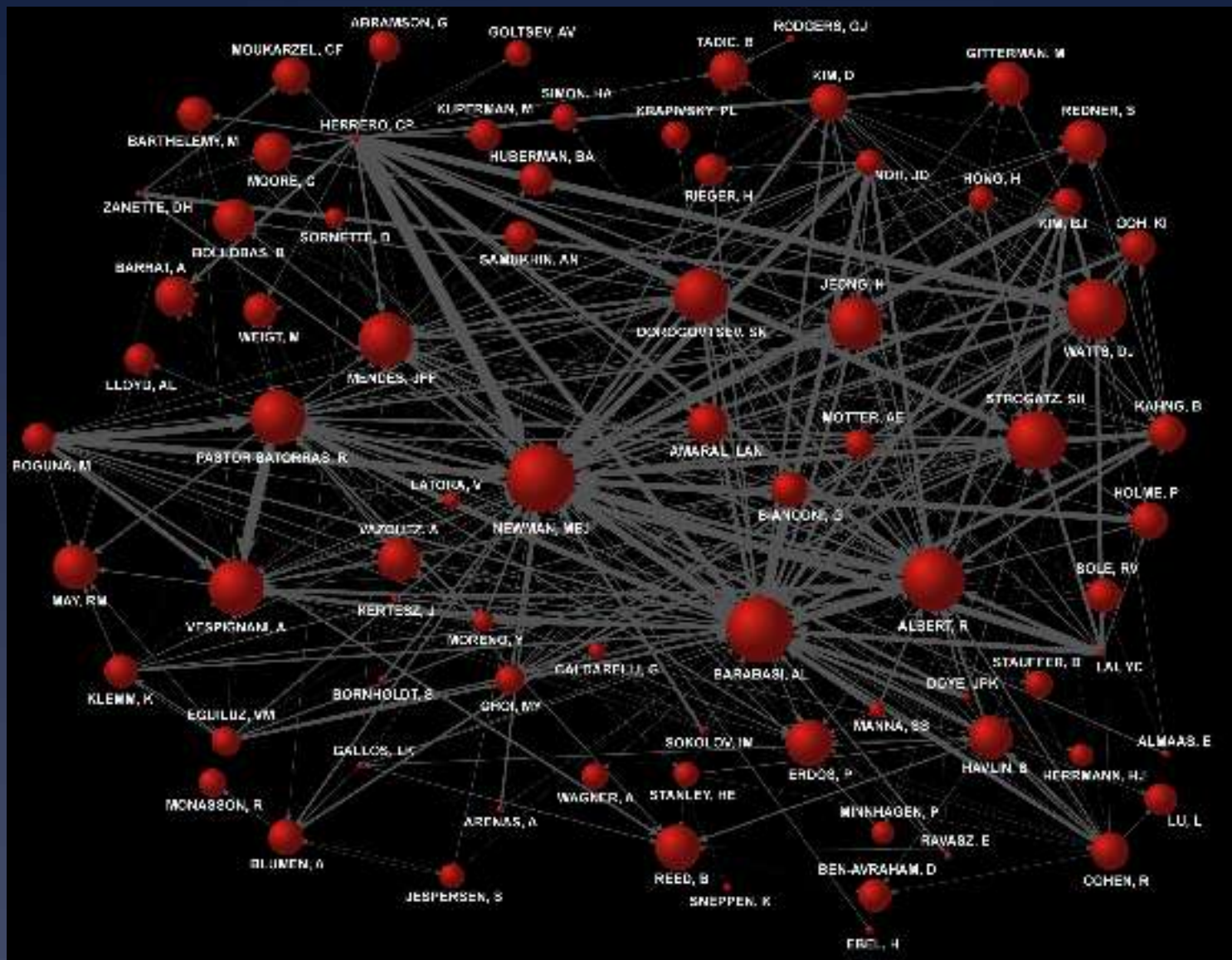
Bibliographic databases: Web of Science, Google Scholar, Scopus, Citebase, Citeseers, Spires, CrossRef, etc.

Several possibilities:

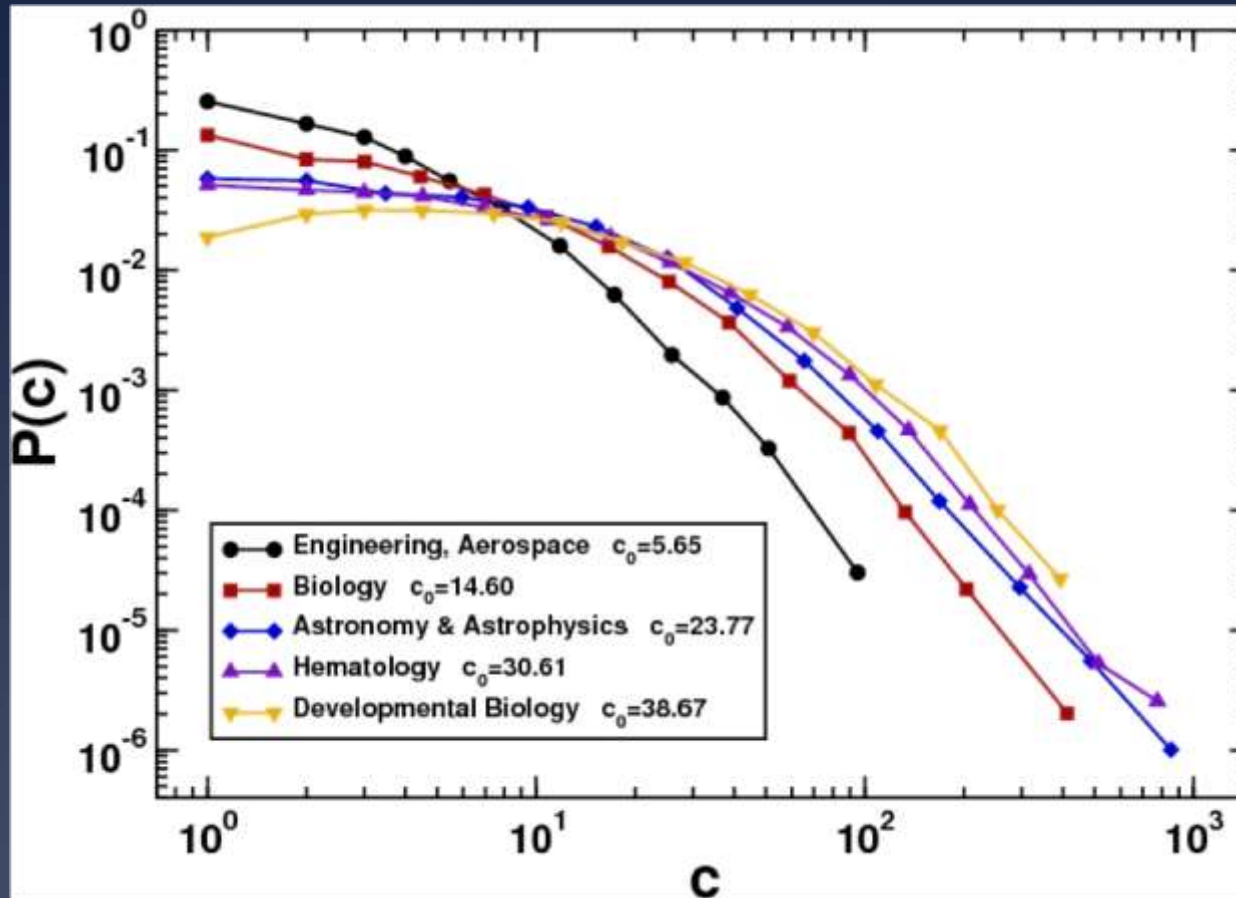
- 1) Citation networks of papers
- 2) Citation networks of journals
- 3) Citation networks of authors

Citation networks are **directed** and essentially **acyclic**!

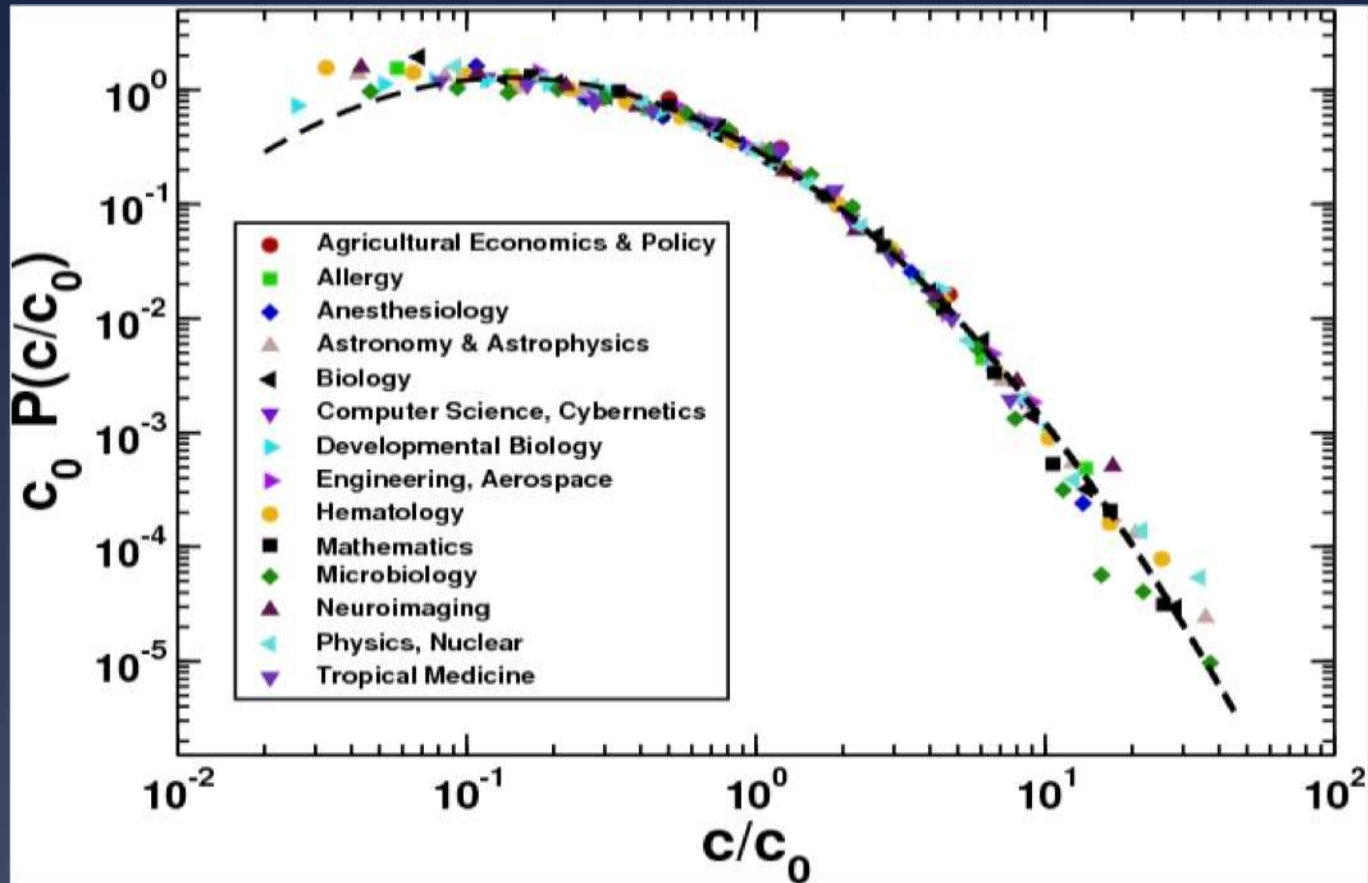




Degree distribution



Degree distribution



F. Radicchi, S. Fortunato and C. Castellano,
Proc. Natl. Acad. Sci. USA 105, 17268 (2008)

Which distribution? Controversial!

Power law

- de Solla Price, D.J.: *Networks of Scientific Papers*. Science **149**, 510–515 (1965)
- Redner, S.: *How popular is your paper? An empirical study of the citation distribution*. Eur. Phys. J. B **4**, 131–134 (1998)
- Seglen, P.O.: *The skewness of science*. J. Am. Soc. Inform. Sci. **43**, 628–638 (1999)
- Vazquez, A.: *Statistics of citation networks*. arXiv:cond-mat/0105031 (2001)
- Lehmann, S., Lautrup, B., Jackson, A.D.: *Citation networks in high energy physics*. Phys. Rev. E. **68**, 026113 (2003)

Lognormal

- Redner, S.: *Citation Statistics from 110 Years of Physical Review*. Phys. Today **58**, 49 (2005)
- Stringer, M.J., Sales-Pardo, M., Amaral, L.A.N.: *Effectiveness of Journal Ranking Schemes as a Tool for Locating Information*. PLoS ONE **3**, e1683 (2008)
- Radicchi F., Fortunato, S., Castellano, C.: *Universality of citation distributions: towards an objective measure of scientific impact*. Proc. Natl. Acad. Sci. USA **105**, 17268–17272 (2008)
- Bommarito II, M.J., Katz, D.: *Properties of the United States Code Citation Network*. arXiv:0911.1751 (2009)

Tsallis distribution

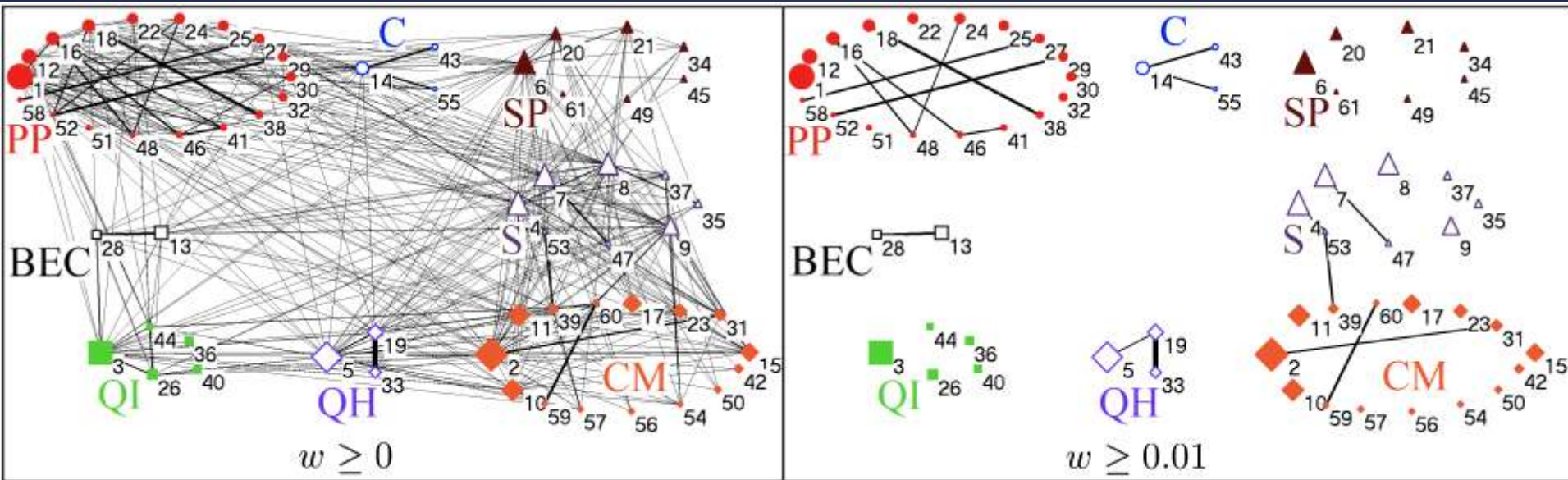
- Wallace, M.L., Larivière, V., Gingras, Y.: *Modeling a Century of Citation Distributions*. J. Informet. **3**, 296–303 (2009)
- Anastasiadis, A.D., de Albuquerque, M.P., de Albuquerque, M.P., Mussi, D.B.: *Tsallis q -exponential describes the distribution of scientific citations: a new characterization of the impact*. Scientometrics **83**, 205–218 (2009)

Modified Bessel function

- van Raan, A.F.J.: *Two-step competition process leads to quasi power-law income distributions. Application to scientific publications and citation distributions*. Physica A **298**, 530–536 (2001)
- van Raan, A.F.J.: *Competition amongst scientists for publication status: toward a model of scientific publication and citation distributions*. Scientometrics **51**, 347–357 (2001)

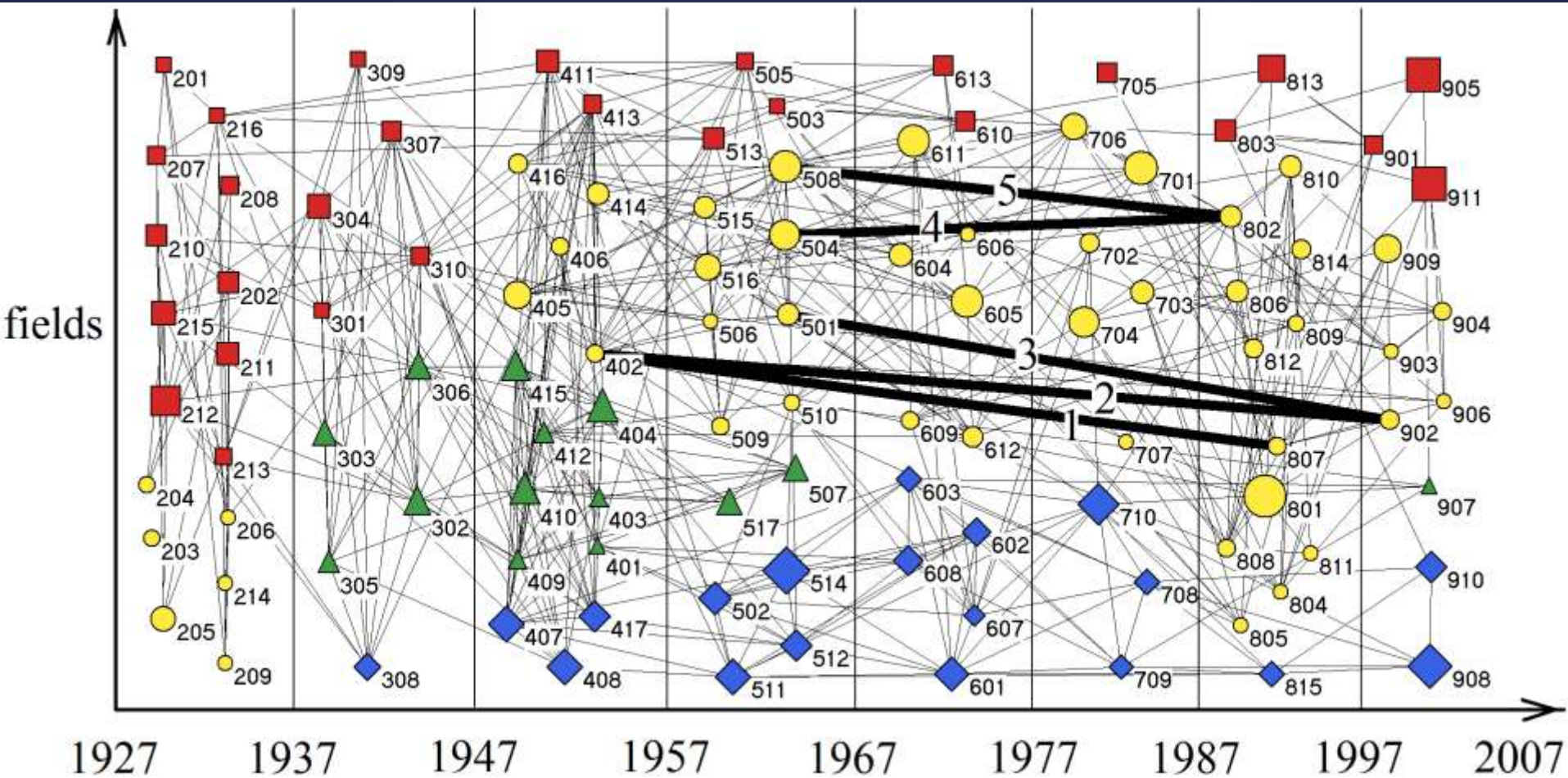
Community structure

P. Chen, S. Redner, *Community structure of the Physical Review Citation Networks*, J. Informet. **4**, 278-290 (2010)



Community structure

P. Chen, S. Redner, *Community structure of the Physical Review Citation Networks*, *J. Informet.* **4**, 278-290 (2010)



Dynamical models

Price's model

D. de Solla Price, *Networks of scientific papers*, *Science* **149**, 510 (1965)

D. de Solla Price, *A general theory of bibliometric and other cumulative advantage processes*, *J. Amer. Soc. Inform. Sci.* **27**, 292 (1976)

Idea: popular papers become more popular (**cumulative advantage**)

H. A. Simon, *On a class of skew distribution functions*, *Biometrika* **42**, 425 (1955)

Price's model

Directed graph of n vertices

Mean in-degree/out-degree m

$$\sum_k kP(k) = m$$

Principle: probability that a newly appearing paper cites a previous paper is proportional to the number k of citations of the paper

Problem: what happens if a paper has no citations?

Solution: proportionality to $k+k_0$, where k_0 is constant

Price took $k_0 = 1$ (publication of the paper is a sort of first citation)

Price's model

$$P(k) = \frac{k(k-1)\dots 1}{(k+2+1/m)\dots(3+1/m)} P(0) = (1+1/m)B(k+1, 2+1/m)$$

$$B(a, b) = \frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)} \quad \text{Beta function!}$$

$$\Gamma(\alpha) = \int_0^{\infty} x^{\alpha-1} e^{-x} dx$$

For large a and fixed b : $B(a, b) \rightarrow a^{-b}$

$$P(k) = (1+1/m)B(k+1, 2+1/m) \rightarrow k^{-(2+1/m)} \quad \text{for large } k$$

For generic k_0 :

$$P(k) = \frac{m+1}{m(k_0+1)+1} \frac{B(k+k_0, 2+1/m)}{B(k_0, 2+1/m)} \rightarrow k^{-(2+1/m)}$$

Exponent of the power law tail is independent of k_0

Static models

Model by Karrer and Newman

B. Karrer, M. E. J. Newman, *Random Acyclic Networks*, Phys. Rev. Lett. **102**, 128701 (2009)

B. Karrer, M. E. J. Newman, *Random graph models for directed acyclic networks*, Phys. Rev. E **80**, 046110 (2009)

Directed and acyclic graphs

Basic ideas:

- 1) nodes are ordered based on their age: $i > j$ means that node i has been introduced after node j
- 2) nodes can only set links to (“cite”) older nodes

The in- and out-degrees of all nodes are fixed from the beginning!

Useful quantity:

$$\lambda_i = \sum_{j=1}^{i-1} k_j^{\text{in}} - \sum_{j=1}^i k_j^{\text{out}}$$

Consistency relations:

$$\lambda_i \geq 0 \text{ for } i = 2 \dots n - 1 \quad \lambda_1 = \lambda_n = 0$$

Expected number of links between i and j :

$$P_{ij} = k_i^{\text{in}} k_j^{\text{out}} \frac{\prod_{l=i+1}^{j-1} \lambda_l}{\prod_{l=i+1}^j (\lambda_l + k_l^{\text{out}})}$$

Dynamics on citation networks

Diffusion processes on directed networks

PageRank: mixed process combining diffusion and teleportation!

$$P_i = \frac{q}{N} + (1 - q) \sum_j \frac{a_{ji}}{k_j^{\text{out}}} P_j$$

Just as it happened for Google and the Web, PageRank may lead to better estimates of prestige than purely local measures like the mere citation count!

- ◆ Chen et al. used PageRank to rank scientific papers published on APS journals [P. Chen et al., J. Informetrics **1**, 8 (2007)]
- ◆ Walker et al. proposed a modification of PageRank (*CiteRank*) which accounts for the age of the papers [D. Walker et al., J. Stat. Mech. P06010 (2007)]
- ◆ Rosvall and Bergstrom applied PageRank to citation networks of journals (*Eigenfactor*) [C. Bergstrom, C&RL News, **68**, no. 5 (2007)]
- ◆ Radicchi et al. applied PageRank to citation networks of authors (*SARA score*) [F. Radicchi et al., Phys. Rev. E **80**, 056103 (2009)]