

Redes Complexas: Internet, Web e outras aplicações em computação

-Bloco #7-

2º semestre de 2008

”The structure of scientific collaboration networks”

Virgílio A. F. Almeida

Outubro de 2008



Departamento de Ciência da Computação
Universidade Federal de Minas Gerais

Por que é importante estudar redes de colaboração

1. Por que alguns grupos são bem sucedidos? Network science!
2. Grandes projetos são colaborativos: ciência, empreendimentos, financeiros, culturais, etc
3. Em 2008, Nature publicou 700 artigos, apenas 6 de autores individuais, o resto grupos.
4. Várias observações a partir dos grafos das redes de colaborações:
 1. em pesquisa: descobriu-se os grupos mais citados tem “right mix o specialism and diversity, and familiarity and freshness.”
 2. “collaboration between-institution had a higher impact than did those within institutions.”
 3. Preferential attachment
 4. First-mover Advantage (Mark Newman 2008)
5. Estudo de colaboração sob o enfoque de “network science”.

The structure of scientific collaboration networks

Revisão do Artigo:

M.E.J. Newman, Proc. Natl. Acad. Sci. USA 98 (2001),
404: *The structure of scientific collaboration networks*

Redes de colaboração científica

1. **Pál Erdős (1913-1996)**
2. Oliver Sacks: "A mathematical genius of the first order, Paul Erdős was totally obsessed with his subject - he thought and wrote mathematics for nineteen hours a day until the day he died. He traveled constantly, living out of a plastic bag, and had no interest in food, sex, companionship, art - all that is usually indispensable to a human life."

-- **The Man Who Loved Only Numbers** (Paul Hoffman, 1998)



Redes de colaboração científica

1. Erdős Number:

1. Erdős publicou **> 1,400 papers** com **> 500 co-autores** durante sua vida

1. Publicou **2 papers por mes, dos 20 anos aos 83**, quando morreu.

2. Principais contribuições na matemática: graph theory, number theory and prime number theory, ...

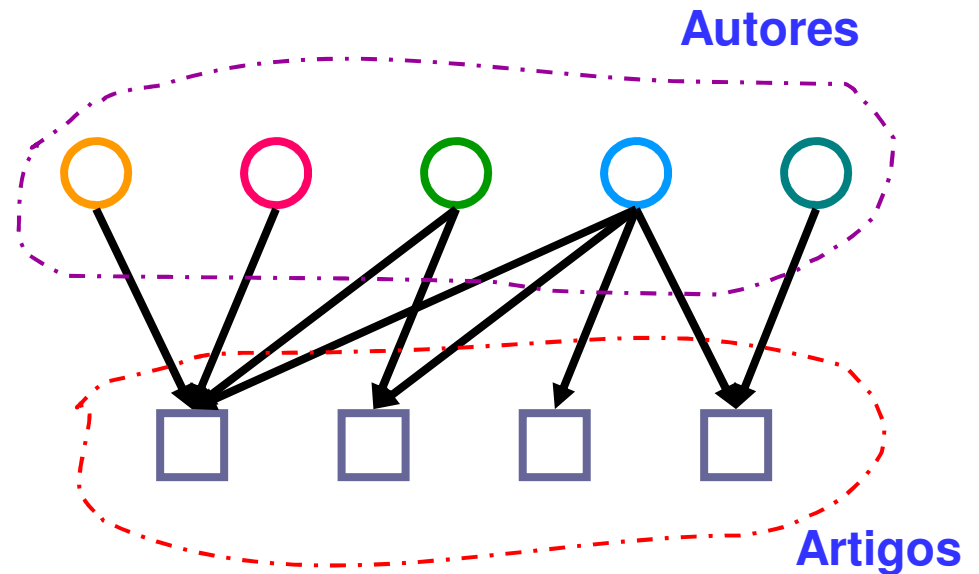
2. Erdős Number de João é 1:

P. Erdős – C. K. Chui – João

2. Erdős teve **small-world network** de colaboração na pesquisa matemática.

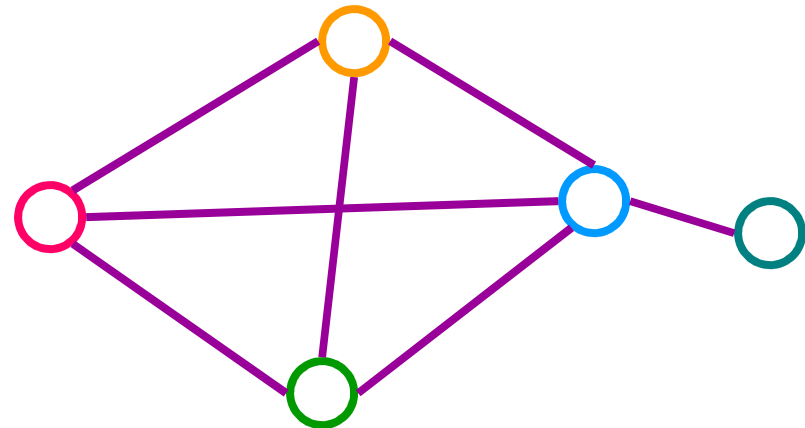
Estrutura da rede de colaboração

- Rede bi-partida



- Projeção de um modo

- Dois nodos do primeiro grupo estão conectados se há arestas conectando esses nodos ao mesmo nodo do segundo grupo.



Estrutura de redes de colaboração científica

1. Investiga-se a estrutura das redes de colaboração científica (i.e.: por que a estrutura é tão importante???) Estrutura afeta a função
2. Dois cientistas são considerados conectados se eles são co-autores de um artigo,
 - Uma janela de 5 anos, de 1995 a 1999 inclusive
3. A maior rede tem mais de um milhão de pessoas
4. Primeira reconstrução detalhada de uma rede real de colaboração.

Science citation index

1,000 Most Cited Physicists, 1981-June 1997

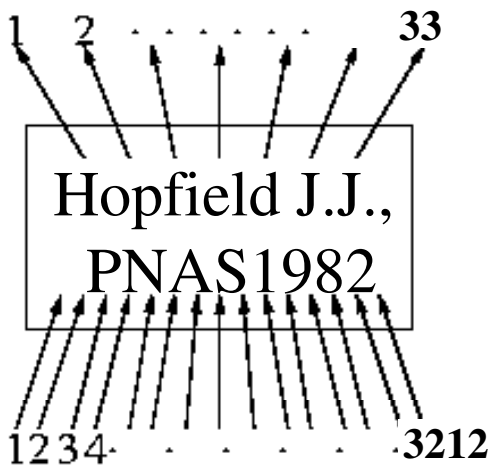
Out of over 500,000 Examined

(see <http://www.sst.nrel.gov>)

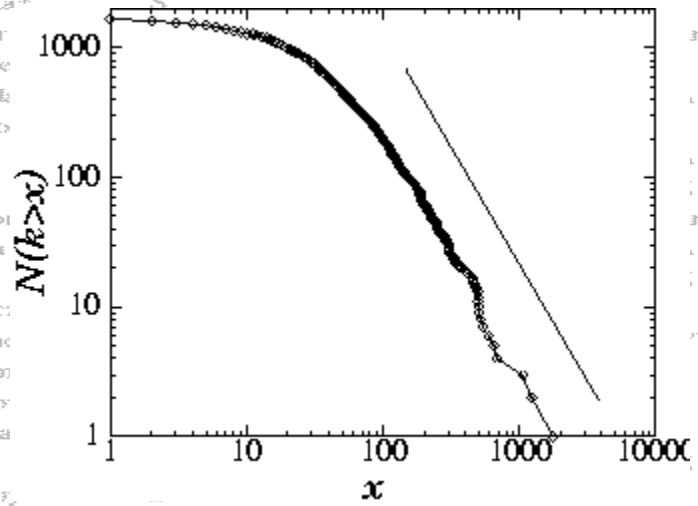
Author name	Institute	Country	Field	avg. cites	total art.	total cites	rank by total cit.
Witten	E Princeton (U)	USA, NJ	High-energy (T)	168	138	23235	1
Gossard	AC UCSB (U)	USA, CA	Semi				2
Cava	RJ Bell Labs (D)	USA, NJ	Supr				3
Ballogg	B Bell Labs (D)	USA, NJ	Supr				4
Ploog	K Max Planck (NL)	Germany	Semi				5
Ellis	J Euro Nuclear Cent.	Switzerland	Astr				6
Fisk	Z Florida State (U)	USA, FL	Solid				7
Cardona	M Max Planck (NL)	Germany	Semi				8
Nanopoulos	DV Texas A&M (U)	USA, TX	High				9
Heeger	AJ UCSB (U)	USA, CA	Poly				10
Lee*	PA						11
Suzuki*	T						12
Anderson	NJ		Solid				13
Suzuki*							14
Freeman	IL		Solid				15
Tanaka*							16
Muller			nd Supr				17
Schnee			Supr				18
Chen			Optics (E)	60	162	9668	19
Morko			Semiconductors (E)	20	477	9668	19
Miller			Semiconductors (E)	67	144	9652	21
Chu			Suprcond (E)	44	273	9453	22
Bednorz			nd Superconductivity (E)	116	85	9311	23
Cohen			Solid State (T)	33	284	9311	23
Meng			Suprconductivity (E)	86	108	9300	25
Waszc			Suprconductivity (E)	57	162	9170	26
Shirane			Suprconductivity (E)	3	269	8841	27
Wieg			Semiconductors (E)	85	104	8822	28
Vando			Magnetism (E)	67	129	8686	29
Uchida				28	301	8520	30
Hor			Suprconductivity (E)	72	119	8512	31
Mu			Astronomy (E)	111	76	8439	32
Birgeneau	RJ MIT (U)	USA, MA	Superconductivity (E)	41	286	8375	33
Jorgensen	JD Argonne (NL)	USA, IL	Superconductivity (E)	67	208	8298	34
Hinks	DG Argonne (NL)	USA, IL	Suprconductivity (E)	37	223	8283	35

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

Os aspectos quantitativos mais interessantes:

1. A distância entre cientistas
2. O número de colaboradores de cientistas.
3. Número de artigos/papers que os cientistas escrevem
4. O grau de *clustering*
 - Probabilidade que dois colaboradores de um cientista sejam também colaboradores entre si.

Bases de Dados

- **Databases of Scientific Articles – mostrando co-autores:**
 1. **Los Alamos e-Print Archives:** preprints (1992 -)
 2. **Medline:** biomedical research articles (1961 -)
 3. **Stanford Public Information Retrieval System (SPIRES):** high-energy physics articles (1974 -)
 4. **Network Computer Science Technical Reference Library (NCSTRL):** computer science articles (10 years records)

- Computados entre 10,000 e 2 milhões de artigos para um período de 5 anos → propriedades: **small-world** e **scale-free** (com **power-law** degree distributions)

Resultados:

Table 1. Summary of results of the analysis of seven scientific collaboration networks

	Los Alamos e-Print Archive						
	MEDLINE	Complete	astro-ph	cond-mat	hep-th	SPIRES	NCSTRL
Total papers	2,163,923	98,502	22,029	22,016	19,085	66,652	13,169
Total authors	1,520,251	52,909	16,706	16,726	8,361	56,627	11,994
First initial only	1,090,584	45,685	14,303	15,451	7,676	47,445	10,998
Mean papers per author	6.4 (6)	5.1 (2)	4.8 (2)	3.65 (7)	4.8 (1)	11.6 (5)	2.55 (5)
Mean authors per paper	3.754 (2)	2.530 (7)	3.35 (2)	2.66 (1)	1.99 (1)	8.96 (18)	2.22 (1)
Collaborators per author	18.1 (1.3)	9.7 (2)	15.1 (3)	5.86 (9)	3.87 (5)	173 (6)	3.59 (5)
Cutoff z_c	5,800 (1,800)	52.9 (4.7)	49.0 (4.3)	15.7 (2.4)	9.4 (1.3)	1,200 (300)	10.7 (1.6)
Exponent τ	2.5 (1)	1.3 (1)	0.91 (10)	1.1 (2)	1.1 (2)	1.03 (7)	1.3 (2)
Size of giant component	1,395,693	44,337	14,845	13,861	5,835	49,002	6,396
First initial only	1,019,418	39,709	12,874	13,324	5,593	43,089	6,706
As a percentage	92.6 (4)%	85.4 (8)%	89.4 (3)	84.6 (8)%	71.4 (8)%	88.7 (1.1)%	57.2 (1.9)%
Second largest component	49	18	19	16	24	69	42
Mean distance	4.6 (2)	5.9 (2)	4.66 (7)	6.4 (1)	6.91 (6)	4.0 (1)	9.7 (4)
Maximum distance	24	20	14	18	19	19	31
Clustering coefficient C	0.066 (7)	0.43 (1)	0.414 (6)	0.348 (6)	0.327 (2)	0.726 (8)	0.496 (6)

Numbers in parentheses are standard errors on the least significant figures.

Número de autores

Table 1. Summary of results of the analysis of seven scientific collaboration networks

	Los Alamos e-Print Archive						
	MEDLINE	Complete	astro-ph	cond-mat	hep-th	SPIRES	NCSTRL
Total papers	2,163,923	98,502	22,029	22,016	19,085	66,652	13,169
Total authors	1,520,251	52,909	16,706	16,726	8,361	56,627	11,994
First initial only	1,090,584	45,685	14,303	15,451	7,676	47,445	10,998
Mean papers per author	6.4 (6)	5.1 (2)	4.8 (2)	3.65 (7)	4.8 (1)	11.6 (5)	2.55 (5)
Mean authors per paper	3.754 (2)	2.530 (7)	3.35 (2)	2.66 (1)	1.99 (1)	8.96 (18)	2.22 (1)
Collaborators per author	18.1 (1.3)	9.7 (2)	15.1 (3)	5.86 (9)	3.87 (5)	173 (6)	3.59 (5)
Cutoff z_c	5,800 (1,800)	52.9 (4.7)	49.0 (4.3)	15.7 (2.4)	9.4 (1.3)	1,200 (300)	10.7 (1.6)
Exponent τ	2.5 (1)	1.3 (1)	0.91 (10)	1.1 (2)	1.1 (2)	1.03 (7)	1.3 (2)
Size of giant component	1,395,693	44,337	14,845	13,861	5,835	49,002	6,396
First initial only	1,019,418	39,709	12,874	13,324	5,593	43,089	6,706
As a percentage	92.6 (4)%	85.4 (8)%	89.4 (3)	84.6 (8)%	71.4 (8)%	88.7 (1.1)%	57.2 (1.9)%
Second largest component	49	18	19	16	24	69	42
Mean distance	4.6 (2)	5.9 (2)	4.66 (7)	6.4 (1)	6.91 (6)	4.0 (1)	9.7 (4)
Maximum distance	24	20	14	18	19	19	31
Clustering coefficient C	0.066 (7)	0.43 (1)	0.414 (6)	0.348 (6)	0.327 (2)	0.726 (8)	0.496 (6)

Numbers in parentheses are standard errors on the least significant figures.

Número de autores

- Dois problemas

1. Dois autores podem ter o mesmo nome
2. Autores podem identificar-se eles próprios em diferentes maneiras em diferentes artigos:

- Exemplos:

- **408** : **V Almeida**, Azer Bestavros, Mark Crovella, and Adriana Oliveira. *Characterizing reference locality in the WWW*. In Proceedings of the Fourth International Conference on Parallel and Distributed Information Systems (PDIS96), December 1996.
- **101**: (A hierarchical characterization of a live streaming media workload - E Veloso, Virgílio **Almeida**, W Meira, A Bestavros, S Jin - Proceedings of the 2nd ACM SIGCOMM Workshop on Internet ..., 2002 - portal.acm.org ... Eveline Veloso Virgílio **Almeida** Wagner Meira
- **163**: Capacity Planning for Web Services: metrics, models, and methods DA Menasce, VAF **Almeida** - 2001 - Prentice Hall PTR Upper Saddle River, NJ, USA

Número de autores

1. Estimativa de Erro

- Todas análises foram feitas duas vezes
 - Na primeira vez, todas iniciais foram usadas
 - Na segunda vez, apenas primeiras iniciais foram usadas

→ *Upper and lower bounds* no número de autores

→ Estimativa de erro para outras quantidades

Número médio de artigos por autor e autores por artigo

Table 1. Summary of results of the analysis of seven scientific collaboration networks

	Los Alamos e-Print Archive						
	MEDLINE	Complete	astro-ph	cond-mat	hep-th	SPIRES	NCSTRL
Total papers	2,163,923	98,502	22,029	22,016	19,085	66,652	13,169
Total authors	1,520,251	52,909	16,706	16,726	8,361	56,627	11,994
First initial only	1,090,584	45,685	14,303	15,451	7,676	47,445	10,998
Mean papers per author	6.4 (6)	5.1 (2)	4.8 (2)	3.65 (7)	4.8 (1)	11.6 (5)	2.55 (5)
Mean authors per paper	3.754 (2)	2.530 (7)	3.35 (2)	2.66 (1)	1.99 (1)	8.96 (18)	2.22 (1)
Collaborators per author	18.1 (1.3)	9.7 (2)	15.1 (3)	5.86 (9)	3.87 (5)	173 (6)	3.59 (5)
Cutoff z_c	5,800 (1,800)	52.9 (4.7)	49.0 (4.3)	15.7 (2.4)	9.4 (1.3)	1,200 (300)	10.7 (1.6)
Exponent τ	2.5 (1)	1.3 (1)	0.91 (10)	1.1 (2)	1.1 (2)	1.03 (7)	1.3 (2)
Size of giant component	1,395,693	44,337	14,845	13,861	5,835	49,002	6,396
First initial only	1,019,418	39,709	12,874	13,324	5,593	43,089	6,706
As a percentage	92.6 (4)%	85.4 (8)%	89.4 (3)	84.6 (8)%	71.4 (8)%	88.7 (1.1)%	57.2 (1.9)%
Second largest component	49	18	19	16	24	69	42
Mean distance	4.6 (2)	5.9 (2)	4.66 (7)	6.4 (1)	6.91 (6)	4.0 (1)	9.7 (4)
Maximum distance	24	20	14	18	19	19	31
Clustering coefficient C	0.066 (7)	0.43 (1)	0.414 (6)	0.348 (6)	0.327 (2)	0.726 (8)	0.496 (6)

Numbers in parentheses are standard errors on the least significant figures.

Número médio de artigos por autor e autores por artigo

1. Tipicamente

- 4 artigos por autor
- 3 autores por artigo

2. Excessões que chamam atenção

- 2 autores por artigo em hep-th & comp. sci.
- 9 autores por artigo em SPIRES

Número de colaboradores

Table 1. Summary of results of the analysis of seven scientific collaboration networks

	MEDLINE	Los Alamos e-Print Archive				SPIRES	NCSTRL
		Complete	astro-ph	cond-mat	hep-th		
Total papers	2,163,923	98,502	22,029	22,016	19,085	66,652	13,169
Total authors	1,520,251	52,909	16,706	16,726	8,361	56,627	11,994
First initial only	1,090,584	45,685	14,303	15,451	7,676	47,445	10,998
Mean papers per author	6.4 (6)	5.1 (2)	4.8 (2)	3.65 (7)	4.8 (1)	11.6 (5)	2.55 (5)
Mean authors per paper	3.754 (2)	2.530 (7)	3.35 (2)	2.66 (1)	1.99 (1)	8.96 (18)	2.22 (1)
Collaborators per author	18.1 (1.3)	9.7 (2)	15.1 (3)	5.86 (9)	3.87 (5)	173 (6)	3.59 (5)
Cutoff z_c	5,800 (1,800)	52.9 (4.7)	49.0 (4.3)	15.7 (2.4)	9.4 (1.3)	1,200 (300)	10.7 (1.6)
Exponent τ	2.5 (1)	1.3 (1)	0.91 (10)	1.1 (2)	1.1 (2)	1.03 (7)	1.3 (2)
Size of giant component	1,395,693	44,337	14,845	13,861	5,835	49,002	6,396
First initial only	1,019,418	39,709	12,874	13,324	5,593	43,089	6,706
As a percentage	92.6 (4)%	85.4 (8)%	89.4 (3)	84.6 (8)%	71.4 (8)%	88.7 (1.1)%	57.2 (1.9)%
Second largest component	49	18	19	16	24	69	42
Mean distance	4.6 (2)	5.9 (2)	4.66 (7)	6.4 (1)	6.91 (6)	4.0 (1)	9.7 (4)
Maximum distance	24	20	14	18	19	19	31
Clustering coefficient C	0.066 (7)	0.43 (1)	0.414 (6)	0.348 (6)	0.327 (2)	0.726 (8)	0.496 (6)

Numbers in parentheses are standard errors on the least significant figures.

Número de colaboradores

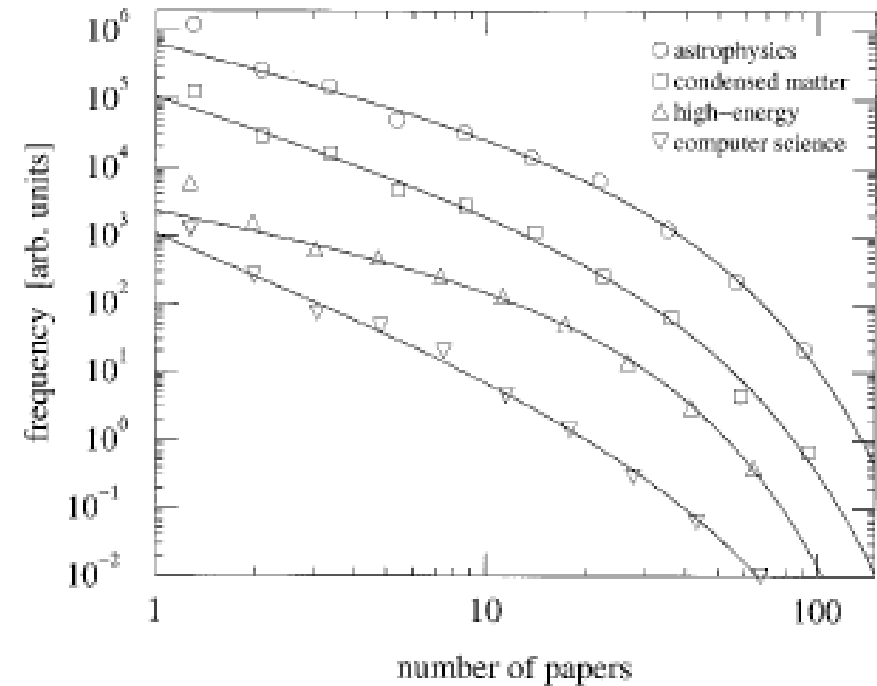
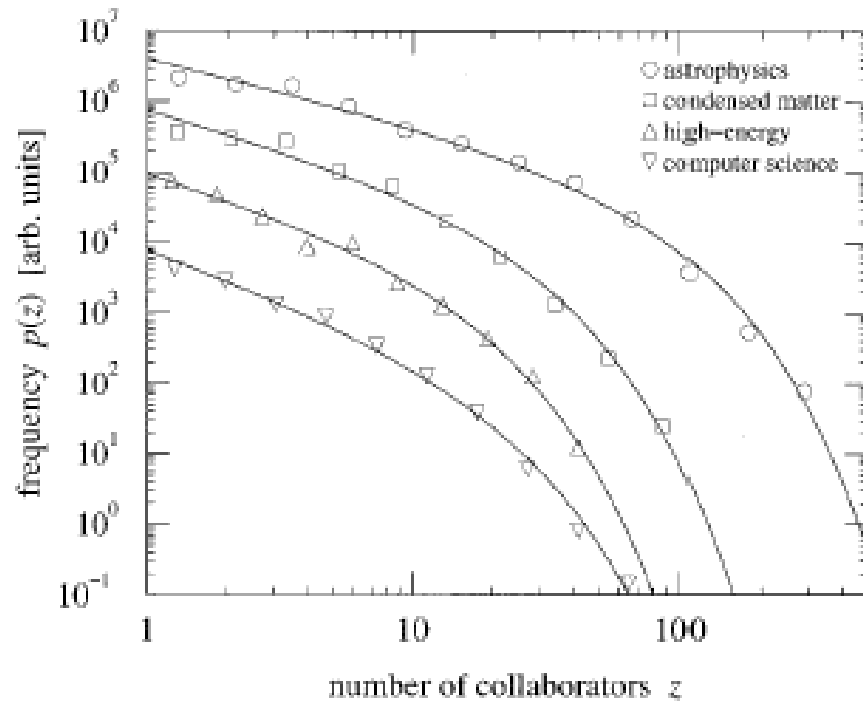
- O número total médio de pessoas com as quais um cientista colabora durante um período de estudo:
 1. Na ordem de 4, para bases de dados puramente teóricas
 2. Mais alto para bases de dados parcial ou totalmente experimentais
 3. 173 em "high-energy physics" de caráter experimental

Número de colaboradores

1. Os dados não seguem uma forma "power-law" perfeitamente
2. Mas os dados "fit" uma "power-law" com um exponencial "cutoff":

$$P(z) \sim z^{-\tau} e^{-z/z_c}$$

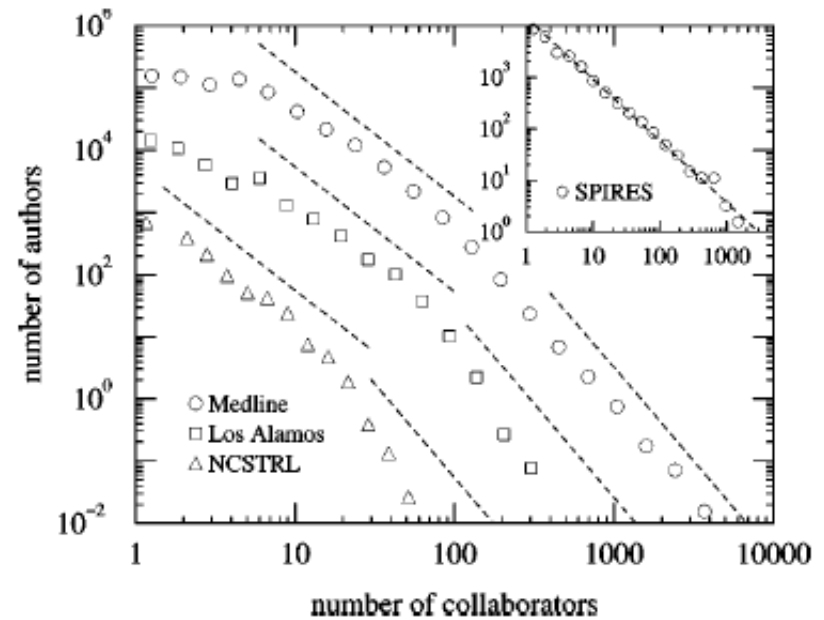
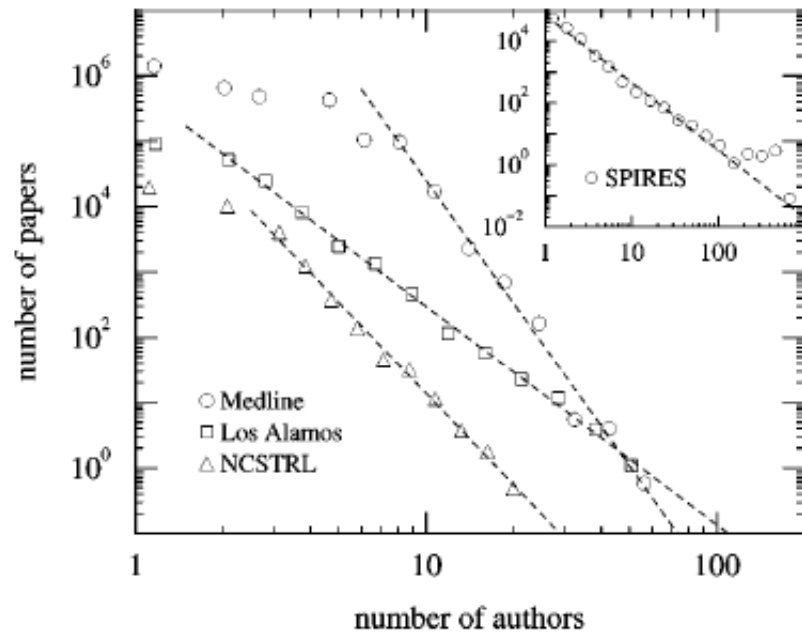
Número de colaboradores



As linhas sólidas são mínimos quadrados ajustados a:

$$P(z) \sim z^{-\tau} e^{-z/z_c}$$

Número de artigos e colaboradores



As linhas sólidas são mínimos quadrados ajustados a:

$$P(z) \sim z^{-\tau} e^{-z/z_c}$$

Número de colaboradores

1. Uma forma de função comum em física e sugere que uma distribuição subjacente que segue uma power-law, mas com restrições impostas ao valor máximo de z .
2. Explicação possível: a janela finita de 5 anos nos dados usados.

Número de colaboradores: grafo ponderado

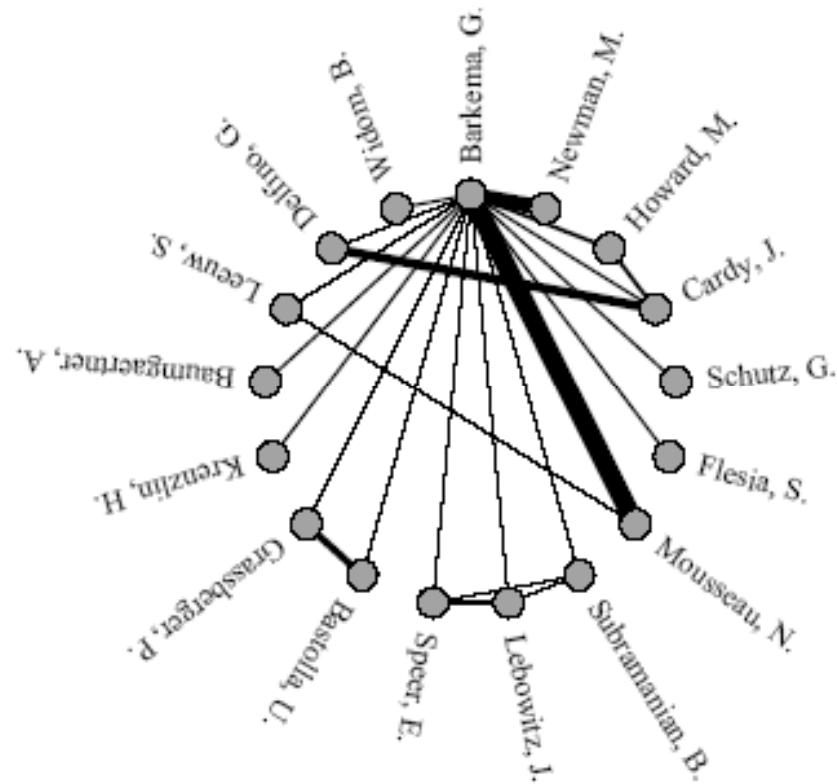


FIG. 10. Gerard Barkema and his collaborators, with lines representing collaborations whose thickness is proportional to our estimate, Eq. (4), of the strength of the corresponding tie.

Número de colaboradores

1. O valor do "cutoff", z_c
 1. Da ordem de 10, para maioria das bases teóricas
 2. Muito maior no SPIRES e MEDLINE
 3. Explicação:
 - Colaborações experimentais muito grandes no SPIRES
 - MEDLINE interessante
 - Aparecem os nomes de diretores de laboratórios em cada artigo do laboratório???

Número de colaboradores

1. O expoente α (obs: power law \rightarrow mice & elephants)
 1. Em todas "hard sciences" próximo de 1
 2. No MEDLINE 2.5 (similar ao www)
3. Para $\alpha < 2$
 1. As propriedades médias da rede são dominadas por poucos elefantes, i.e., "Great people"
4. For $\alpha > 2$
 1. As propriedades médias da rede são dominadas pela classe de ratos, i.e., "little people"

O componente gigante

Table 1. Summary of results of the analysis of seven scientific collaboration networks

	Los Alamos e-Print Archive						
	MEDLINE	Complete	astro-ph	cond-mat	hep-th	SPIRES	NCSTRL
Total papers	2,163,923	98,502	22,029	22,016	19,085	66,652	13,169
Total authors	1,520,251	52,909	16,706	16,726	8,361	56,627	11,994
First initial only	1,090,584	45,685	14,303	15,451	7,676	47,445	10,998
Mean papers per author	6.4 (6)	5.1 (2)	4.8 (2)	3.65 (7)	4.8 (1)	11.6 (5)	2.55 (5)
Mean authors per paper	3.754 (2)	2.530 (7)	3.35 (2)	2.66 (1)	1.99 (1)	8.96 (18)	2.22 (1)
Collaborators per author	18.1 (1.3)	9.7 (2)	15.1 (3)	5.86 (9)	3.87 (5)	173 (6)	3.59 (5)
Cutoff z_c	5,800 (1,800)	52.9 (4.7)	49.0 (4.3)	15.7 (2.4)	9.4 (1.3)	1,200 (300)	10.7 (1.6)
Exponent τ	2.5 (1)	1.3 (1)	0.91 (10)	1.1 (2)	1.1 (2)	1.03 (7)	1.3 (2)
Size of giant component	1,395,693	44,337	14,845	13,861	5,835	49,002	6,396
First initial only	1,019,418	39,709	12,874	13,324	5,593	43,089	6,706
As a percentage	92.6 (4)%	85.4 (8)%	89.4 (3)	84.6 (8)%	71.4 (8)%	88.7 (1.1)%	57.2 (1.9)%
Second largest component	49	18	19	16	24	69	42
Mean distance	4.6 (2)	5.9 (2)	4.66 (7)	6.4 (1)	6.91 (6)	4.0 (1)	9.7 (4)
Maximum distance	24	20	14	18	19	19	31
Clustering coefficient C	0.066 (7)	0.43 (1)	0.414 (6)	0.348 (6)	0.327 (2)	0.726 (8)	0.496 (6)

Numbers in parentheses are standard errors on the least significant figures.

O componente gigante

1. Um grande número de pesquisadores, todos conectados mutuamente.
2. Na maioria das bases, o maior desses grupos abrange 80-90% de todos autores.
3. Em hep-th (i.e., high energy physics) e "comp. sci", os CGs são menores, mas ainda assim superiores a 50%
4. O segundo maior grupo é bem menor.

O componente gigante

■ Observações:

1. As redes de colaboração científica são altamente conectadas.
2. Não há sinais de fragmentação

Gráu médio de separação

Table 1. Summary of results of the analysis of seven scientific collaboration networks

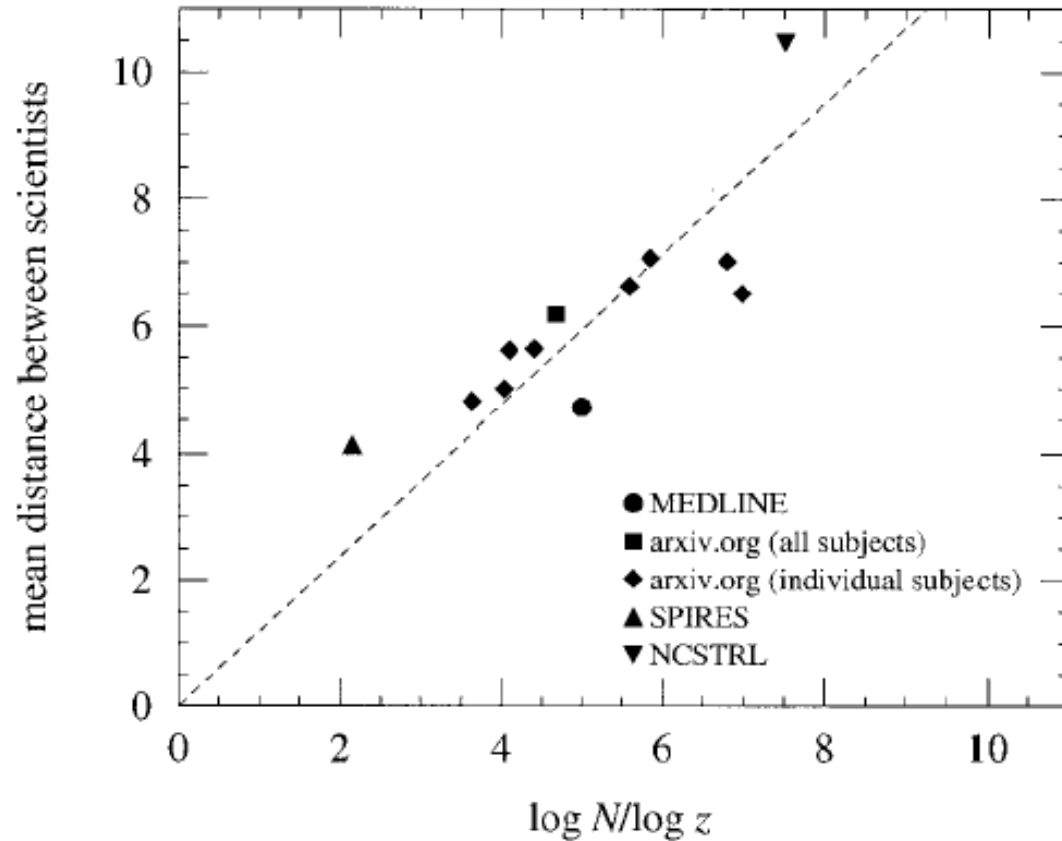
	MEDLINE	Los Alamos e-Print Archive				SPIRES	NCSTRL
		Complete	astro-ph	cond-mat	hep-th		
Total papers	2,163,923	98,502	22,029	22,016	19,085	66,652	13,169
Total authors	1,520,251	52,909	16,706	16,726	8,361	56,627	11,994
First initial only	1,090,584	45,685	14,303	15,451	7,676	47,445	10,998
Mean papers per author	6.4 (6)	5.1 (2)	4.8 (2)	3.65 (7)	4.8 (1)	11.6 (5)	2.55 (5)
Mean authors per paper	3.754 (2)	2.530 (7)	3.35 (2)	2.66 (1)	1.99 (1)	8.96 (18)	2.22 (1)
Collaborators per author	18.1 (1.3)	9.7 (2)	15.1 (3)	5.86 (9)	3.87 (5)	173 (6)	3.59 (5)
Cutoff z_c	5,800 (1,800)	52.9 (4.7)	49.0 (4.3)	15.7 (2.4)	9.4 (1.3)	1,200 (300)	10.7 (1.6)
Exponent τ	2.5 (1)	1.3 (1)	0.91 (10)	1.1 (2)	1.1 (2)	1.03 (7)	1.3 (2)
Size of giant component	1,395,693	44,337	14,845	13,861	5,835	49,002	6,396
First initial only	1,019,418	39,709	12,874	13,324	5,593	43,089	6,706
As a percentage	92.6 (4)%	85.4 (8)%	89.4 (3)	84.6 (8)%	71.4 (8)%	88.7 (1.1)%	57.2 (1.9)%
Second largest component	49	18	19	16	24	69	42
Mean distance	4.6 (2)	5.9 (2)	4.66 (7)	6.4 (1)	6.91 (6)	4.0 (1)	9.7 (4)
Maximum distance	24	20	14	18	19	19	31
Clustering coefficient C	0.066 (7)	0.43 (1)	0.414 (6)	0.348 (6)	0.327 (2)	0.726 (8)	0.496 (6)

Numbers in parentheses are standard errors on the least significant figures.

Gráu médio de separação

1. A distância típica entre dois cientistas está por volta de 6.
2. Forte correlação entre a distância média e $\log N$.
3. O diâmetro é da ordem de 20

Distância média de separação



- Demonstração empírica da variação logarítmica com o tamanho da rede para uma rede social.

Clustering

Table 1. Summary of results of the analysis of seven scientific collaboration networks

	Los Alamos e-Print Archive						
	MEDLINE	Complete	astro-ph	cond-mat	hep-th	SPIRES	NCSTRL
Total papers	2,163,923	98,502	22,029	22,016	19,085	66,652	13,169
Total authors	1,520,251	52,909	16,706	16,726	8,361	56,627	11,994
First initial only	1,090,584	45,685	14,303	15,451	7,676	47,445	10,998
Mean papers per author	6.4 (6)	5.1 (2)	4.8 (2)	3.65 (7)	4.8 (1)	11.6 (5)	2.55 (5)
Mean authors per paper	3.754 (2)	2.530 (7)	3.35 (2)	2.66 (1)	1.99 (1)	8.96 (18)	2.22 (1)
Collaborators per author	18.1 (1.3)	9.7 (2)	15.1 (3)	5.86 (9)	3.87 (5)	173 (6)	3.59 (5)
Cutoff z_c	5,800 (1,800)	52.9 (4.7)	49.0 (4.3)	15.7 (2.4)	9.4 (1.3)	1,200 (300)	10.7 (1.6)
Exponent τ	2.5 (1)	1.3 (1)	0.91 (10)	1.1 (2)	1.1 (2)	1.03 (7)	1.3 (2)
Size of giant component	1,395,693	44,337	14,845	13,861	5,835	49,002	6,396
First initial only	1,019,418	39,709	12,874	13,324	5,593	43,089	6,706
As a percentage	92.6 (4)%	85.4 (8)%	89.4 (3)	84.6 (8)%	71.4 (8)%	88.7 (1.1)%	57.2 (1.9)%
Second largest component	49	18	19	16	24	69	42
Mean distance	4.6 (2)	5.9 (2)	4.66 (7)	6.4 (1)	6.91 (5)	4.0 (1)	9.7 (4)
Maximum distance	24	20	14	18	19	19	31
Clustering coefficient C	0.066 (7)	0.43 (1)	0.414 (6)	0.348 (6)	0.327 (2)	0.726 (8)	0.496 (6)

Numbers in parentheses are standard errors on the least significant figures.

Clustering

1. Tipicamente, dois cientistas tem uma probabilidade da ordem de 30% de colaborar se ambos já tiverem colaborado previamente com um terceiro cientista.
2. MEDLINE novamente exceção ($C=0.066$)
 - Explicação (?)
 - Organização de laboratórios "Top-down" → "tree-like"
 - redes de colaboração → baixo "clustering coefficient"

Conclusões

- Em todos casos as comunidades parecem constituir um "small-world"
- As redes são altamente "clustered"
- Diferenças estatísticas significantes entre diferentes comunidades:
 - Biomédicas tem "clustering" muito menor que outras áreas, e o expoente da curva de colaboradores é maior que 2, indicando que a rede é dominada por muitas pessoas com poucos colaboradores.