

Processamento e Transferência da Informação por Meio Eletrônico
PPGCI - IBICT
A Internet como Objeto, Fonte e Campo de Estudo
PGED - IBqM

Jacqueline Leta
Fábio Castro Gouveia

Indicadores Webmétricos:

Dados coletados por *Log* ou *Page Tagging*

- ▶ Quem está visitando seu site: por IP (*Log*) ou por Cookie (*Page Tagging*).
- ▶ Qual o caminho percorrido pelo visitante: Indica tendências e interesse do visitante e ajuda na avaliação da estrutura.
- ▶ Quanto tempo o visitante gasta em cada página: Um padrão de muito tempo dispendido pode significar conteúdo interessante, mas também pode ser um conteúdo confuso.
- ▶ Onde o visitante sai do seu site: Pode ser uma página lógica ou apenas uma onde o visitante terminou sua visita.

Indicadores Webmétricos:
Métricas na Web 2.0

▶ Web 1.0

▶ Presença do Usuário

- ▶ Visitantes
- ▶ Visitas Únicas
- ▶ Páginas Vistas
- ▶ Tempo de Navegação

▶ Web 2.0

▶ Ações do Usuário

- ▶ Participação
- ▶ Contribuição
- ▶ Vinculação
- ▶ Interesse



Indicadores Webmétricos:
Métricas na Web 2.0

- ▶ Votação
- ▶ Tagging (Rotulagem)
- ▶ Comentários
- ▶ Weblogs
- ▶ Links Sociais
- ▶ Redes Sociais
- ▶ Redes de Compartilhamento de Mídia
- ▶ Wikis



Analisando outros *sites*

- ▶ **Compete.com e Alexa.com**
 - ▶ Ranking dos 1.000.000 de sites mais visitados no mundo;
 - ▶ Confiabilidade cai abaixo do número 100.000;
 - ▶ Ranking depende de *toolbar* mas também faz medidas por outras fontes;
 - ▶ *Sites* menos internacionais são prejudicados no ranking.
- ▶ **Archive.org**
 - ▶ Arquivo das páginas da Internet;
 - ▶ WayBackMachine.

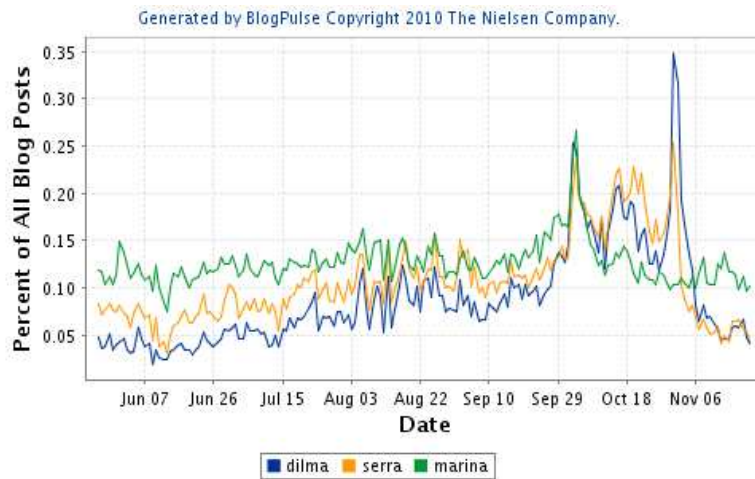


Métricas da Web 2.0

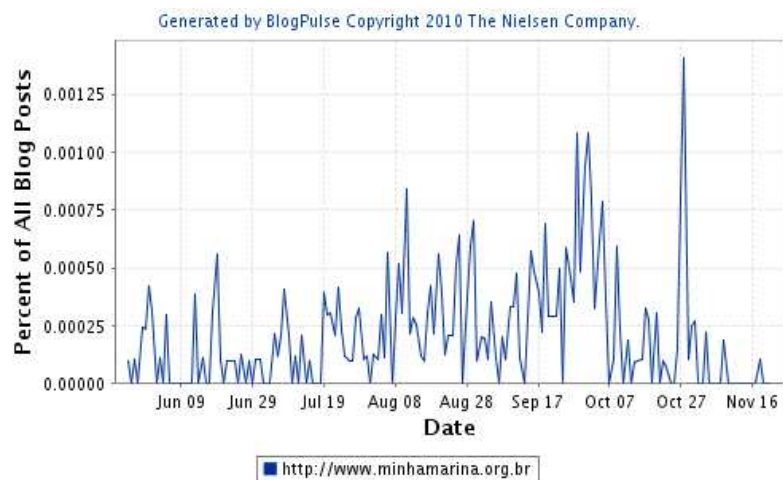
- ▶ **Acompanhamento de termos/links postados em Blogs**
 - ▶ BlogPulse (www.blogpulse.com) – extinto ;-(
- ▶ **Acompanhamento de termos buscados no Google**
 - ▶ Google Trends (www.google.com/trends)
 - ▶ Google Insights
 - ▶ Trends for Websites



BlogPulse



BlogPulse



Processamento e Transferência da Informação por Meio Eletrônico
PPGCI - IBICT
A Internet como Objeto, Fonte e Campo de Estudo
PGED - IBqM

Jacqueline Leta
Fábio Castro Gouveia

O que é Webometria

- ▶ Ciência que faz estudos quantitativos de conteúdo e estrutura da Internet com a aplicação de métodos informétricos (Almind & Ingwersen, 1997)
 - ▶ Netometrics (Bossy, 1995): “Suplementa a bibliometria e a cientometria ao observar a ‘ciência em ação’ na Internet”
 - ▶ Webometry (Abraham, 1996)
 - ▶ Internetometrics (Almind & Ingwersen, 1996)
 - ▶ Webometrics (Almind & Ingwersen, 1997)
 - ▶ Cybermetrics (periódico de 1997 por Isidro Aguillo)
- ▶ Ciência com cerca de quinze anos de existência que tem o link como unidade central de estudo



Terminologias mais frequentes

- ▶ *Webmetrics* ou *Web Metrics*
 - ▶ Geralmente se refere a análise de *logs* ou por *page tagging*
- ▶ *Cybermetrics*

“O estudo dos aspectos quantitativos da construção e uso dos recursos de informação, estruturas e tecnologias da Internet como um todo a partir de abordagens informétricas e bibliométricas”
(Björneborn & Ingwersen, 2004)

 - ▶ Grupos de discussão e listas de e-mail
 - ▶ Medida relativas à tecnologia, topologia e tráfego na Internet
 - ▶ Web (Webometria)
- ▶ *Webometrics*



Relação entre Informetria e Webometia segundo Ingwersen & Björneborn (2002)

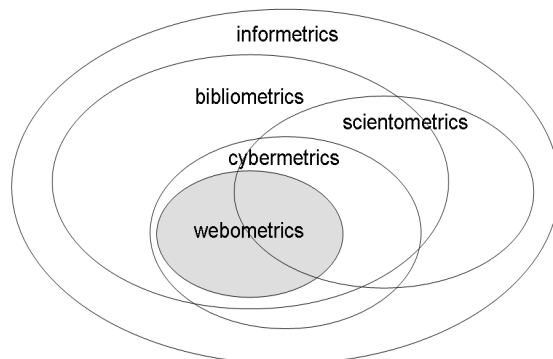
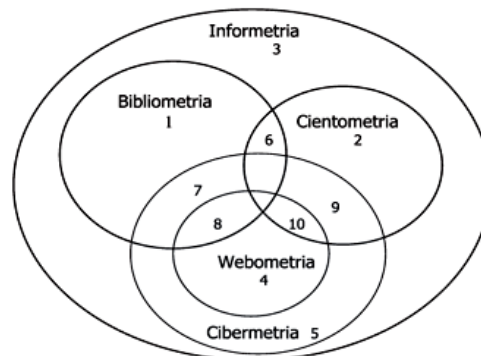


Diagrama de inserção da Webometria (Vanti, 2005)

FIGURA 2
Inter-relação entre os subcampos das métricas dentro da biblioteconomia e da ciência da informação, segundo Vanti



Limitações: Mecanismos de Busca

- ▶ Mecanismos de busca não indexam todas as páginas da Internet;
- ▶ Número limitado de resultados: 250 a 1000;
- ▶ Existência de diferentes endereços URL para um mesmo sítio;
- ▶ Mecanismos de busca podem gerar resultados diferentes;
- ▶ Sítios em subdiretórios ou que compartilham o domínio dificultam a análise;
- ▶ Resultados podem variar ao longo do tempo;
- ▶ Resultados podem variar dependendo do número de resultados por página solicitados.

Limitações: Webometria e Bibliometria

- ▶ Sítios não são artigos ou revistas, mas pela Webometria podemos levantar a visibilidade e impacto de instituições na Web;
- ▶ *Links* não são iguais a citações, mas os métodos de análise podem ser compartilhados;
- ▶ *Links* mantém a Internet viva e também são dinâmicos;
- ▶ *Links* afetam o *ranking* das páginas em mecanismos de busca (têm valor comercial);
- ▶ Teoria sobre a motivação de *links* é necessária.



Limitações e Desafios: estrutura e indexação

- ▶ Comunicação e estrutura da Web
 - ▶ A estrutura da Web é caótica em virtude das diferentes arquiteturas de informação de cada sítio
 - ▶ A Internet hoje em dia representa um meio para dar visibilidade a uma instituição, pesquisa ou pesquisador
- ▶ Mecanismos de busca
 - ▶ A Web cresce mais rápido do que pode ser indexada e alguns sítios fazem uso de métodos para manipular o seu *ranking*
 - ▶ Sítios dinâmicos podem ser de difícil indexação aumentando a estimativa do tamanho do sítio ou não sendo indexáveis ("Web Invisível")
 - ▶ Os mecanismos de busca não são perfeitos pois a cobertura não é completa, mas permitem desenvolver uma série de estudos Webométricos



Tipos de Estudos Webométricos

- ▶ Fator de Impacto na *Web* – Web IF
 - ▶ Análise de *Co-Link*
 - ▶ Análise de Interlink
-
- ▶

Cláusulas ou *strings* de busca

- ▶ **WIF externo – eWIF (visibilidade/tamanho do sítio)**
 - ▶ (A) link:URL n – domain:URL n ou linkdomain:URL n – domain:URL n
 - ▶ (B) domain:URL n
 - ▶ WIF externo = A/B
 - ▶ **Co-link (co-inlink)**
 - ▶ link:URL n link:URL n
 - ▶ linkdomain:URL n linkdomain:URL n
 - ▶ **Interlink**
 - ▶ link:URL n domain:URL n
(todos os *links* provenientes das páginas do domínio que apontam para a página inicial de outro domínio)
 - ▶ linkdomain:URL n domain:URL n
(todos os *links* provenientes das páginas do domínio que apontam para outro domínio)
-
- ▶

Fator de Impacto na Web externo
(External WIF – eWIF)

domain:URLn

Lista de páginas do domínio URLn cadastradas na base de dados do mecanismo de busca

(Ex: domain:www.url1.com)

linkdomain: URLn -domain:URLn

Lista de páginas cadastradas que têm *link* para o URLn e que não são deste domínio URLn

(Ex: linkdomain:www.url1.com -domain:www.url1.com)

▶

Fator de Impacto na Web externo
(External WIF – eWIF*)

Nº de páginas externas que têm *links*
para determinado domínio
(*linkdomain:URLn - domain:URLn*)

—————
(*domain:URLn*)

Nº de páginas do domínio

- ▶ eWIF é maior quando o número de páginas com *links* recebido é alto e o número de páginas do domínio é baixo

* Segundo Ingwersen (1998) e modificado por Thelwall (2001)

▶

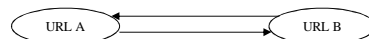
Diferentes tipologias de *links*

- ▶ *Inlink* e *outlink*: estão associados à perspectiva da URL que recebe e da URL que fornece o *link*.



Diferentes tipologias de *links*

- ▶ *Interlinks* (Thelwall, 2002): são os *links* trocados entre duas URLs.

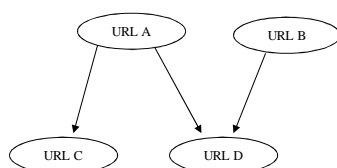


- ▶ *Co-links* ou *co-citations* (Herrero-Solana & Morales-del-Castillo, 2004): ocorrem quando uma URL aponta simultaneamente para duas ou mais URLs.



Diferentes tipologias de *links*

- ▶ *Co-outlinks e co-inlinks* (BJÖRNEBORN & INGWERSEN, 2004)



Exemplo de matriz para coleta de *co-links*

linkdomain:URL1 linkdomain:URL2

URL1 = www.ufrj.br e URL2 = www.ibict.br

linkdomain:www.ufrj.br linkdomain:www.ibict.br

	URL1	URL2	URL3
URL1	Σ da linha	linkdomain:URL1 linkdomain:URL2	linkdomain:URL1 linkdomain:URL3
URL2	linkdomain:URL2 linkdomain:URL1	Σ da linha	linkdomain:URL2 linkdomain:URL3
URL3	linkdomain:URL3 linkdomain:URL1	linkdomain:URL3 linkdomain:URL2	Σ da linha

	URL1	URL2	URL3
URL1	790	220	570
URL2	220	580	360
URL3	570	360	930

Exemplo de matriz de *co-link*

	XPLT-AR	CASA-BR	CDCC-BR	CVIV-BR	ESTC-BR	GOEL-BR	MVDA-BR	MAST-BR	MCTP-BR	MIMC-CL	MLKA-CO	MEPM-CO	SINA-MX	XPLR-MX	MLUZ-MX	PPLT-MX	UNIV-MX	NINO-VE
XPLT-AR	107	10	0	8	10	0	0	1	0	2	12	3	14	14	2	16	15	0
CASA-BR	10	331	18	20	132	18	19	34	24	0	10	0	10	8	2	12	12	0
CDCC-BR	0	18	114	3	37	11	6	21	12	0	2	0	0	0	3	3	0	0
CVIV-BR	8	20	3	160	20	10	19	20	11	0	8	0	8	8	2	8	15	0
ESTC-BR	10	133	36	20	395	38	18	39	28	1	15	0	11	13	2	12	19	0
GOEL-BR	0	18	11	10	38	199	26	48	41	0	0	0	0	0	0	0	7	0
MVDA-BR	0	19	6	19	18	26	160	46	19	0	0	0	0	0	0	0	7	0
MAST-BR	1	34	21	20	39	48	46	256	29	1	2	2	1	1	1	2	8	0
MCTP-BR	0	25	11	11	28	41	19	29	179	0	3	0	1	0	0	2	9	0
MIMC-CL	2	0	0	0	1	0	0	1	0	319	10	81	0	3	4	78	76	66
MLKA-CO	12	10	2	8	15	0	0	2	3	10	399	14	24	27	105	30	131	6
MEPM-CO	3	0	0	0	0	0	0	2	0	79	14	333	1	3	4	81	79	66
SINA-MX	14	10	0	8	11	0	0	1	1	0	24	1	189	30	14	36	39	0
XPLR-MX	14	8	0	8	13	0	0	1	0	3	27	3	30	206	9	46	44	0
MLUZ-MX	2	2	0	2	2	0	0	1	0	4	105	4	14	9	412	57	210	0
PPLT-MX	16	12	3	8	12	0	0	2	2	79	30	80	36	46	57	638	192	65
UNIV-MX	15	12	3	15	19	7	7	8	9	75	131	78	39	44	210	190	932	67
NINO-VE	0	0	0	0	0	0	0	0	0	65	6	67	0	0	65	66	270	0

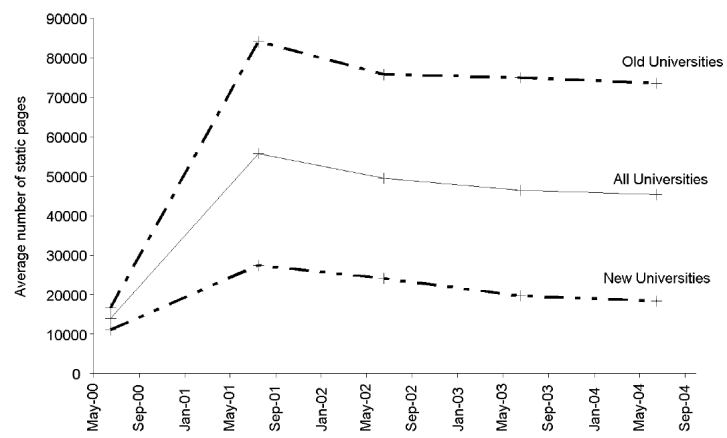
Análise de trabalhos desenvolvidos

- ▶ Estudos longitudinais
 - ▶ Payne & Thelwall (2007)
- ▶ Web Impact Factor
 - ▶ Thelwall (2001)
 - ▶ Gouveia (2007)
 - ▶ Webometrics.info
- ▶ Co-link analysis (co-inlink)
 - ▶ Larson (1996)
 - ▶ Herrero-Solana & Morales-del-Castillo (2004)
 - ▶ Gouveia & Kurtenbach (2008)
 - ▶ Vaughan & You (2006)
- ▶ Interlink analysis
 - ▶ Musgrove, Binns, Page-Kennedy & Thelwall (2003)
 - ▶ Thelwall & Zuccala (2008)
 - ▶ Vanti (2010)

Estudos longitudinais

Payne & Thelwall (2007)

A longitudinal study of academic webs: Growth and stabilisation
Scientometrics, Vol. 71, No. 3 (2007) 523–539



Web Impact Factor

Thelwall (2001)

Results from a Web Impact Factor crawler
Journal of Documentation (2001) 57(2), 177-191

Table 2: *Results of the WIF calculations*

Domain	Pages	Self-links	Self-Link WIF	External Links	External Link WIF	Research Rank
uce.ac.uk	3349	17821	5.32	64	0.0191	5
coventry.ac.uk	6253	22896	3.66	44	0.0070	4
warwick.ac.uk	65373	376946	5.77	405	0.0062	1
bham.ac.uk	84235	284849	3.38	485	0.0058	2
wlv.ac.uk	38163	127420	3.34	204	0.0053	6
aston.ac.uk	51381	229664	4.47	143	0.0028	3



Thelwall (2001)

Results from a Web Impact Factor crawler
Journal of Documentation (2001) 57(2), 177-191

- ▶ **Cálculo do número de páginas de um sítio**
 - ▶ Páginas de apoio para ensino
 - ▶ Informações gerais sobre a universidade
 - ▶ Informações para alunos interessados
 - ▶ Páginas experimentais
 - ▶ Espelhos de conteúdos de outros sítios
 - ▶ Manuais online
 - ▶ Estatísticas de eventos ou acessos a páginas na Web
 - ▶ Páginas criadas por estudantes para execução de tarefas
 - ▶ Páginas pessoais
- ▶ WIFs podem ser calculados filtrando as páginas não relevantes;
- ▶ Não foi observada relação entre WIF e produção acadêmica.



Resultados de eWIF para associações de museus e centros de ciência

Associação	Máximo	Mínimo	Média	Mediana
ASTC	116,5385	0,0095	7,6387	3,2785
ECSITE	92,4490	0,0433	4,8479	1,8574
NSCF	18,4783	0,0748	3,0674	1,8613
ASPAC	94,5455	0,0429	13,5067	4,2511
RedPOP	15,5952	0,0528	1,6118	0,8714
ABCMC	9,9737	0,0151	1,5188	0,8707



WR Ranking Web of World universities: T...

Ranking Web of World Universities

July 10

home world countries world rank rank by country european rank latin american rank

> home > top Latin America

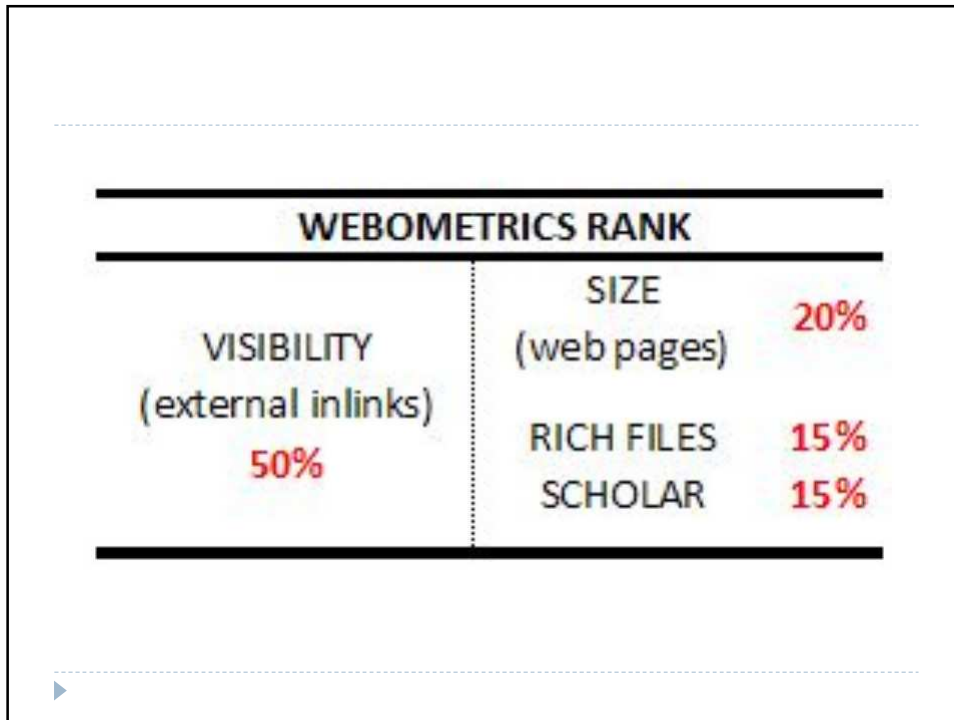
Rank Data

About Us
 About the Ranking
 Top 12000 Universities
 Premier League (Excel Files)
 Top USA & Canada
 Top Latin America
 Top Europe
 Top Cent. & East. Europe
 Top Asia
 Top South East Asia
 Top South Asia
 Top Arab World
 Top Oceania
 Top Africa
 Top Sub-Saharan Africa
 Country Scoreboard
 Best Practices
 Notes (*/**)

Top Latin America

Universities 1 to 100 of 100

CONTINENT RANK	UNIVERSITY	COUNTRY	WORLD RANK	POSITION			
				SIZE	VISIBILITY	RICH FILES	SCHOLAR
1	Universidad Nacional Autónoma de México **		70	171	88	129	3
2	Universidade de São Paulo ***		122	84	107	79	734
3	Universidad de Chile		199	282	231	174	177
4	Universidade Estadual de Campinas ***		239	175	222	190	734
5	Universidad de Buenos Aires		274	342	307	237	340
6	Universidade Federal de Santa Catarina ***		377	317	442	280	734
7	Universidade Federal do Rio de Janeiro ***		386	537	371	303	734
8	Tecnológico de Monterrey *		460	449	406	442	1,161
9	Universidade Federal de Minas Gerais ***		470	441	601	352	734

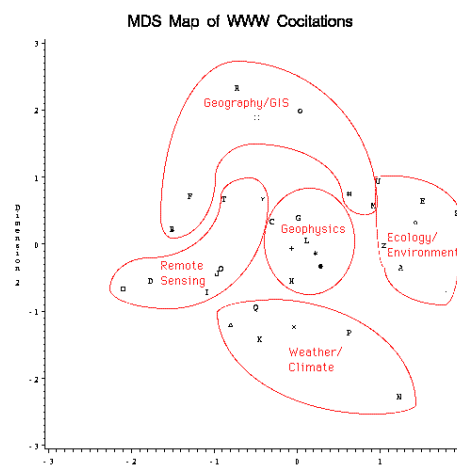


Co-link analysis (co-inlink)

Larson (1996)

Bibliometrics of the World Wide Web: An Exploratory Analysis of the Intellectual Structure of Cyberspace.

<https://sherlock.ischool.berkeley.edu/asis96/asis96.html>



Herrero-Solana & Morales-del-Castillo (2004)

Mapas "geopolíticos" de internet: aplicación de las nuevas técnicas de representación de la información

Ci. Inf., Brasilia, v. 33, n. 3, p.69-75, set./dez. 2004

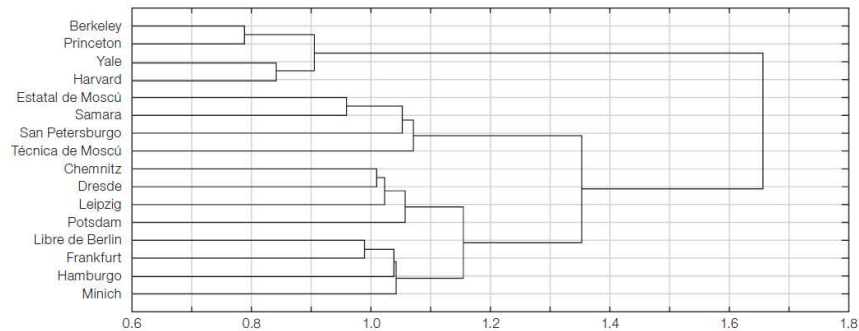
	BERK	YALE	PRIN	HARV	MSU	SPBU	BMST	SAMA	CHEM	LEIP	DRES	POTS	BERLIN	FRAN	HAMB	MUNC
BERK	364962	32068	33971	46355	1183	610	131	43	4125	2000	2635	1224	8456	5330	6248	3706
YALE	32068	211793	16488	31542	712	129	18	14	1246	1063	1158	619	3651	2430	2517	2203
PRIN	33971	16488	163141	23726	426	203	22	15	1536	1068	1243	664	3271	2285	2215	1715
HARV	46355	31542	23726	335381	1020	336	30	35	1680	1640	1535	1000	6250	3723	3745	3571
MSU	1183	712	426	1020	36151	1398	879	852	95	215	175	108	362	194	299	235
SPBU	610	129	203	336	1398	81042	411	349	29	44	42	30	113	76	114	74
BMST	131	18	22	30	879	411	104704	301	8	4	11	5	6	7	9	4
SAMA	43	14	15	35	852	349	301	8143	4	6	5	4	7	7	14	8
CHEM	4125	1246	1536	1680	95	29	8	4	92075	3980	4313	1701	3449	2260	2014	1527
LEIP	2000	1063	1068	1640	215	44	4	6	3980	114569	3982	2500	3101	2405	2640	2466
DRES	2635	1158	1243	1535	175	42	11	5	4313	3982	107023	2142	3087	2279	2405	2152
POTS	1224	619	664	1000	108	30	5	4	1701	2500	2142	66973	1969	1462	1531	1317
BERLIN	8456	3651	3271	6250	362	113	6	7	3449	3101	3087	1969	99957	5142	5189	4412
FRAN	5330	2430	2285	3723	194	76	7	7	2260	2405	2279	1462	5142	66451	3914	3383
HAMB	6848	2517	2215	3745	299	114	9	14	2014	2640	2405	1531	5189	3914	189381	3636
MUNC	3706	2203	1715	3571	235	74	4	8	1527	2466	2152	1317	4412	3383	3636	111302



Herrero-Solana & Morales-del-Castillo (2004)

Mapas "geopolíticos" de internet: aplicación de las nuevas técnicas de representación de la información

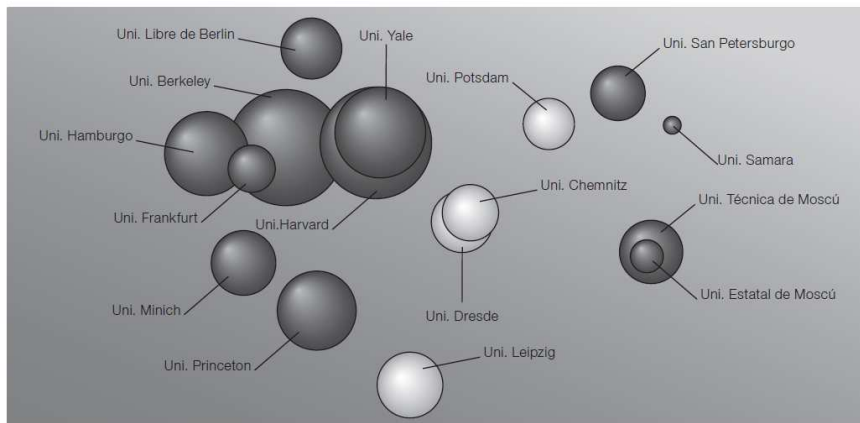
Ci. Inf., Brasilia, v. 33, n. 3, p.69-75, set./dez. 2004



Herrero-Solana & Morales-del-Castillo (2004)

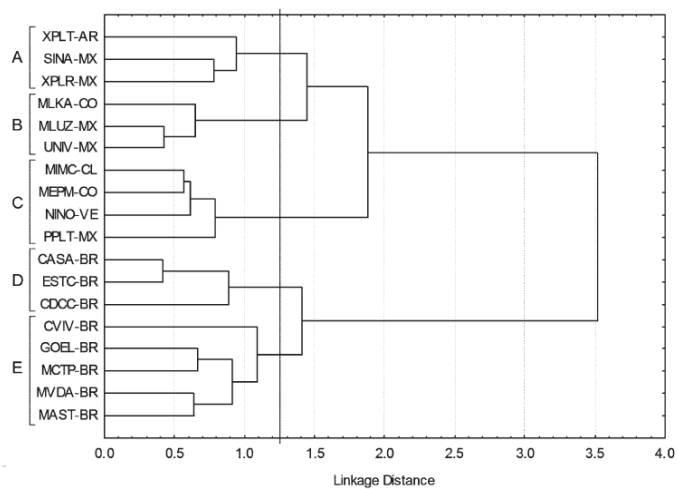
Mapas "geopolíticos" de internet: aplicación de las nuevas técnicas de representación de la información

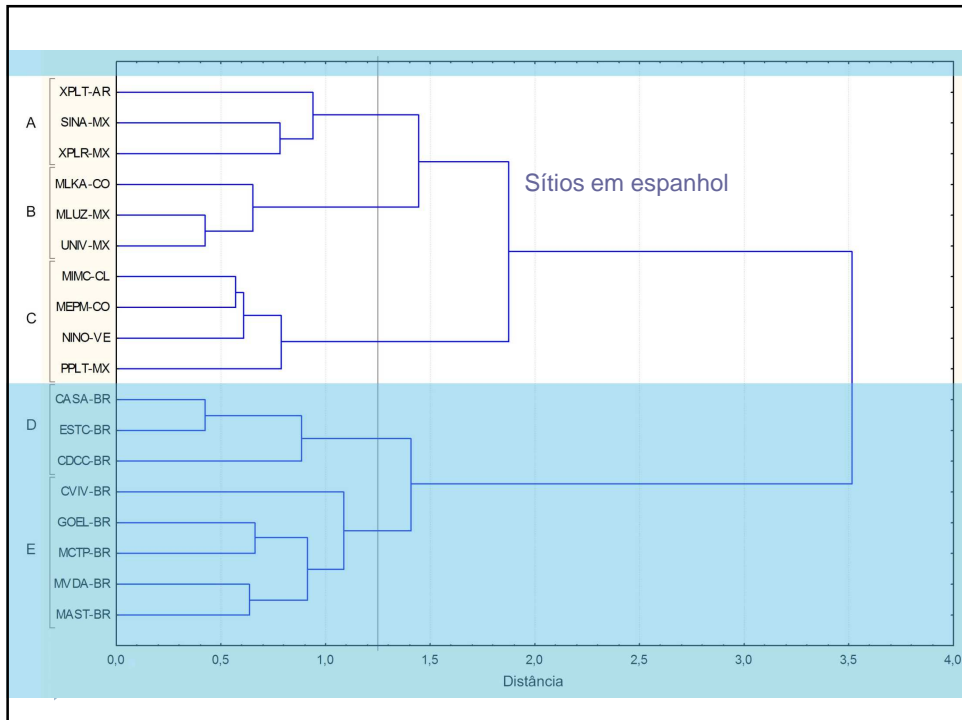
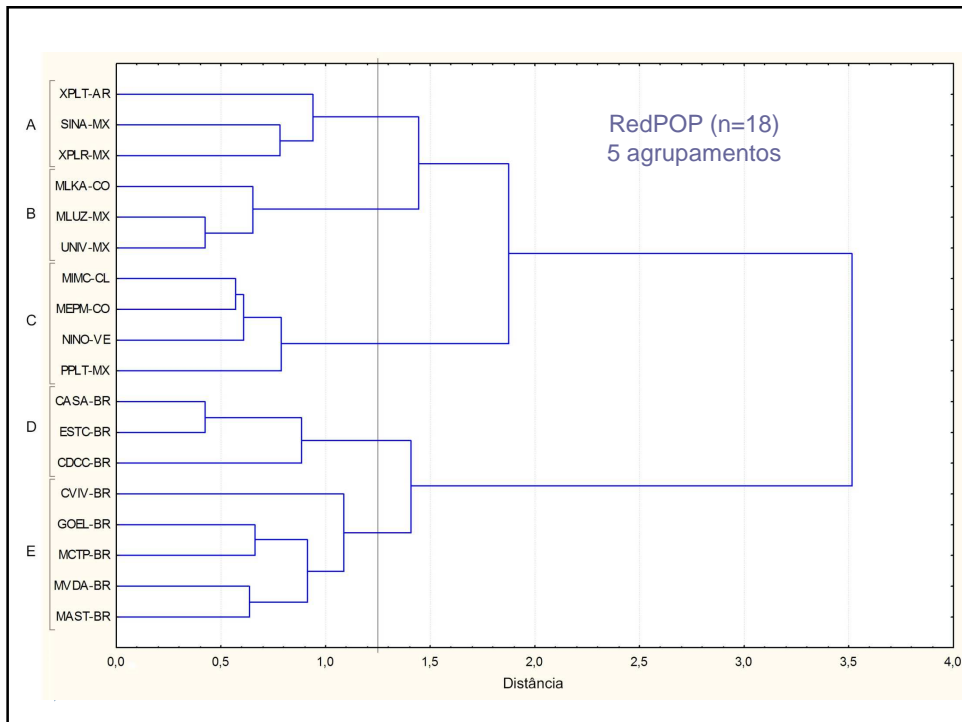
Ci. Inf., Brasília, v. 33, n. 3, p.69-75, set./dez. 2004

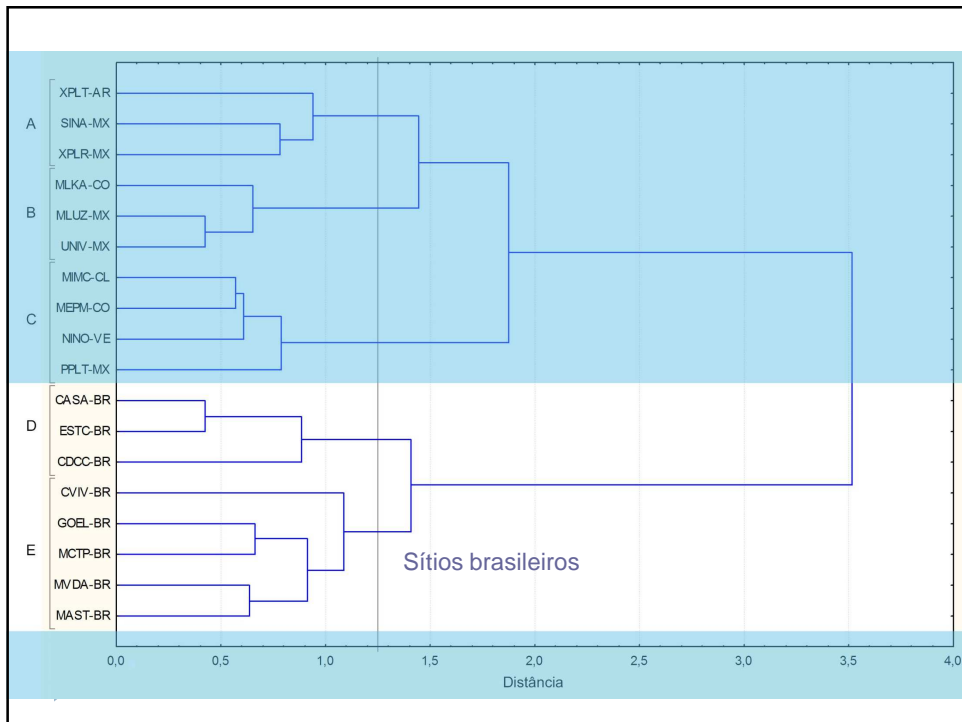


Gouveia & Kurtenbach (2009)

Mapping the web relations of science centres and museums from Latin America
Scientometrics, Vol. 79, No. 3 (2009) 491-505







Gouveia & Kurtenbach (2009)

Mapping the web relations of science centres and museums from Latin America
Scientometrics, Vol. 79, No. 3 (2009) 491–505

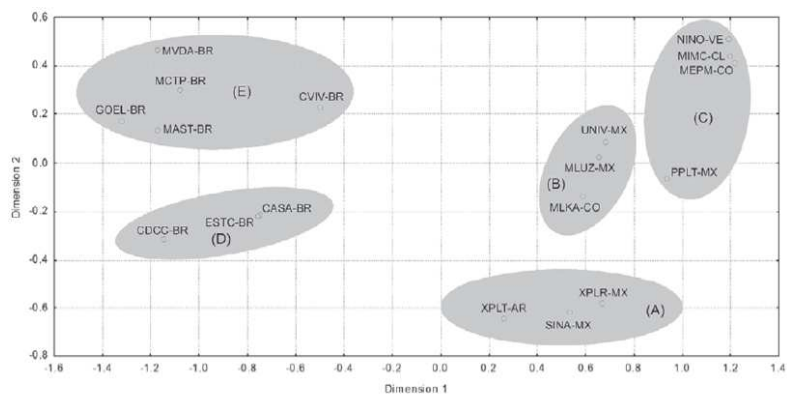
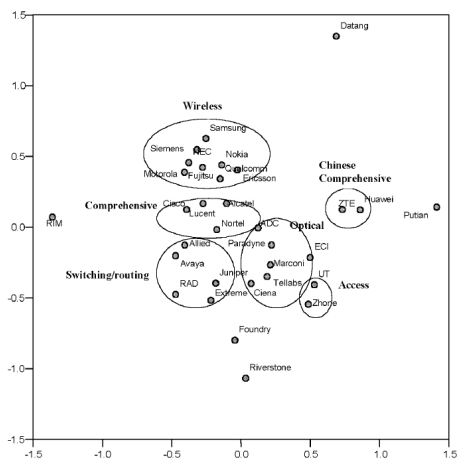


Figure 3. MDS 2D plot using distance matrix from Web co-link analysis of 18 Latin American science centre and museum websites

Vaughan & You (2006)

Comparing business competition positions based on web co-link data: The global market vs. the Chinese market

Scientometrics, Vol. 68, No. 3 (2006) 611-628



Vaughan & You (2006)

Comparing business competition positions based on web co-link data: The global market vs. the Chinese market

Scientometrics, Vol. 68, No. 3 (2006) 611-628

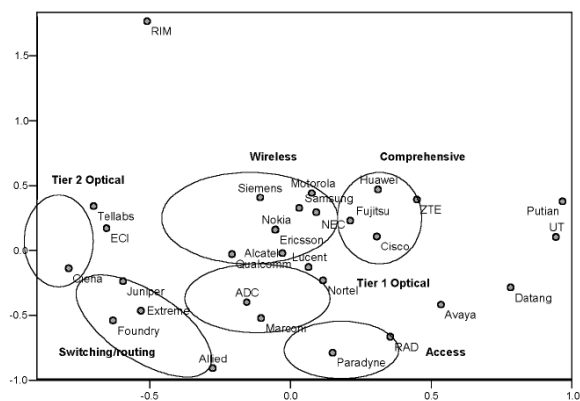
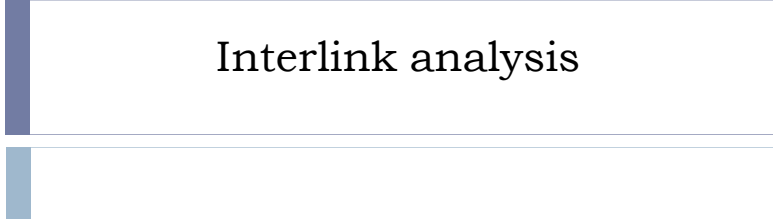


Figure 2. MDS mapping result based on Yahoo! China data

Análise de co-link

Quais os fatores mais importantes?

- ▶ Idioma comum;
 - ▶ Idiomas disponíveis no sítio;
 - ▶ Distância geográfica;
 - ▶ Relação política;
 - ▶ Proximidade temática;
 - ▶ Associações e redes;
 - ▶ Projetos de cooperação.
-



Interlink analysis

Musgrove, Binns, Page-Kennedy & Thelwall (2003)

A method for identifying clusters in sets of interlinking Web spaces
 Scientometrics, Vol. 58, No. 3 (2003) 657-672

To	France	Italy	Germany	Spain	Greece	Norway	Holland	Portugal	Switzerland	Belgium	Denmark	Austria	Sweden	UK	Finland
France (f)	0	1624	5589	1665	287	645	2401	445	1727	2363	676	1016	965	12186	600
Italy (i)	1926	0	5038	1625	376	831	2025	510	1055	1040	570	1315	1139	10720	777
Germany (g)	6583	5327	0	4237	999	3884	12382	1401	6865	3109	2072	10544	5218	38375	2922
Spain (e)	2837	2701	8249	0	439	1265	3140	1055	2147	1845	1032	1956	1850	16477	1254
Greece (m)	345	508	2158	834	0	228	679	136	169	304	385	468	447	3058	216
Norway (n)	883	619	3377	796	138	0	1582	210	479	505	980	692	2556	9991	1169
Holland (h)	1455	1416	7747	1106	277	757	0	435	838	1623	607	2989	1213	13475	802
Portugal (p)	640	724	2411	1506	175	483	1010	0	307	479	341	572	520	6371	309
Switzerland (c)	1404	730	4899	477	103	337	2077	179	0	604	208	1206	594	3857	378
Belgium (b)	1154	650	1908	619	212	255	1557	204	470	0	248	615	394	3915	297
Denmark (d)	521	379	4984	324	99	640	3027	143	261	374	0	347	902	3264	512
Austria (a)	1108	1747	10929	1398	360	740	2130	386	1396	1152	1049	0	1196	8780	908
Sweden (s)	781	761	3493	607	152	1764	1863	235	464	655	1017	966	0	15078	1607
UK (u)	5517	5752	22289	5753	1438	3289	9459	1621	4203	6314	2206	4535	5236	0	3004
Finland (l)	682	799	3363	865	1334	1591	200	500	463	770	928	2574	6396	0	0



Musgrove, Binns, Page-Kennedy & Thelwall (2003)

A method for identifying clusters in sets of interlinking Web spaces
 Scientometrics, Vol. 58, No. 3 (2003) 657-672

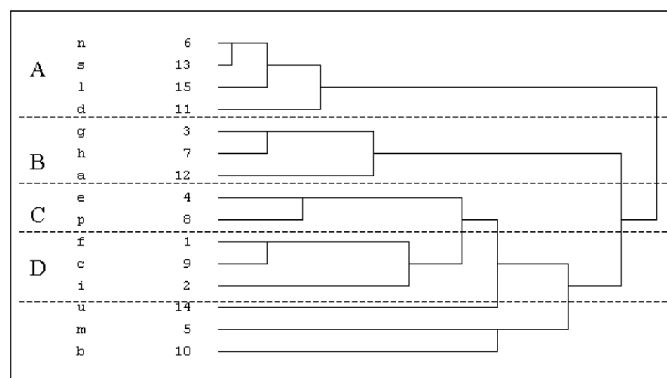


Figure 3. Dendrogram for inbound links



Musgrove, Binns, Page-Kennedy & Thelwall (2003)

A method for identifying clusters in sets of interlinking Web spaces
Scientometrics, Vol. 58, No. 3 (2003) 657-672

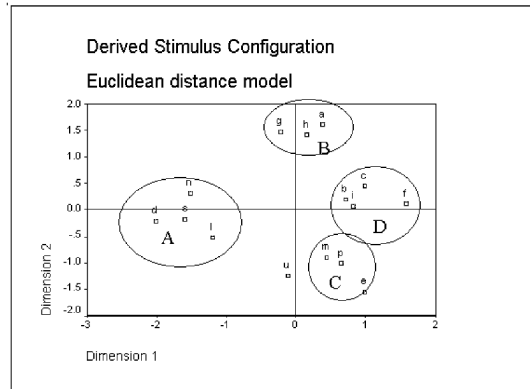


Figure 4. Multidimensional scaling results for inbound links

Thelwall & Zuccala (2008)

A university-centred European Union link analysis
Scientometrics, Vol. 75, No. 3 (2008) 407-420

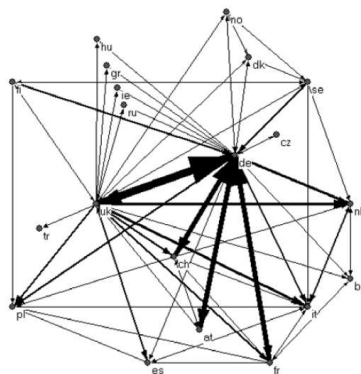


Figure 3. (Links to universities; core and periphery shape) European link network with the width of arrows proportional to the number of pages in the source country domain linking to university web sites in the target country domain. Links below 10% of the maximum are not shown; unlinked countries are not shown

Vanti (2010)

A presença das universidades públicas federais brasileiras na Web
XI ENANCIB

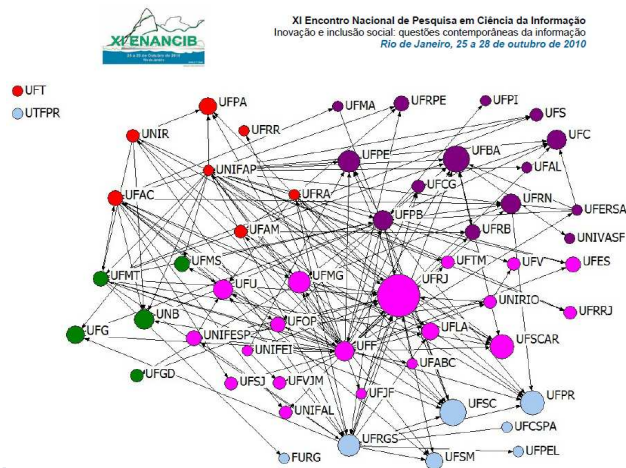


Figura 6 – Rede social das IFES brasileiras.
Fonte: Elaboração das autoras.

Processamento e Transferência da Informação por Meio Eletrônico
PPGCI - IBICT
A Internet como Objeto, Fonte e Campo de Estudo
PGED - IBqM

Jacqueline Leta
Fábio Castro Gouveia