

# Os Aspéctos Teóricos e Metodológicos da Cienciometria

## Webometria: aplicações e desafios

1º Encontro Brasileiro de Bibliometria e Cienciometria  
14 a 16 de setembro de 2008

Fábio Castro Gouveia

MV-COC-Fiocruz

[fgouveia@coc.fiocruz.br](mailto:fgouveia@coc.fiocruz.br)

15 de setembro de 2008

# Webometria: aplicações e desafios

- O que é Webometria
- Ferramentas Webométricas (Mec. Busca)
- Tipos de estudos Webométricos
- Exemplos de aplicação
- Limitações e Desafios

# 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 cienciometria 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 pouco mais de dez anos de existência e tem o *link* como unidade central de estudo

# O que é Webometria

## 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*

# Ferramentas Webométricas

- *Webcrawler* especializado
  - Algoritmo de Busca Público
  - Desenvolvido com propósitos Webométricos
  - Base de dados deve ser criada especificamente para cada pesquisa
- Mecanismos de busca (Google / Altavista)
  - Algoritmo de Busca Proprietário
  - Prioridade em fornecer resultados satisfatórios e não necessariamente exatos
  - Ampla base de dados disponível

# Ferramentas Webométricas

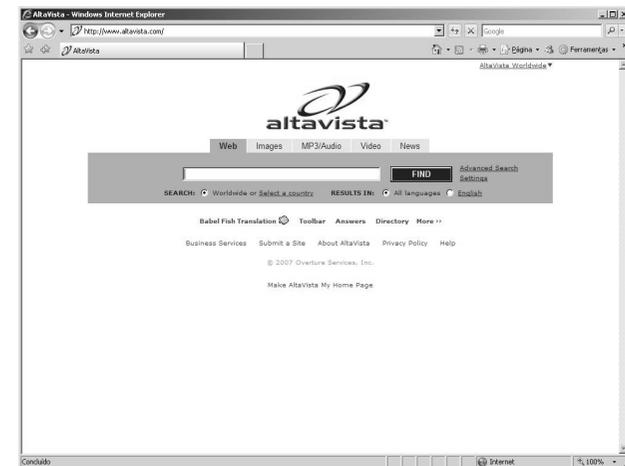
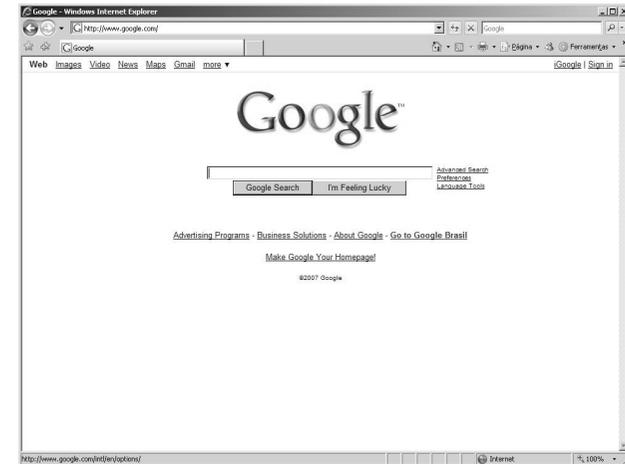
## Mecanismos de Busca

- Mecanismos de Busca e Diretórios direcionam cerca de 85% do tráfego na Internet (Hu et al., 2001)
- Legitimação do conteúdo nos Diretórios e Mecanismos de Busca, e o PageRank (Brin & Page, 1998)
  - Diretórios = Humanos
  - Metadatos = Produtor de Conteúdo
  - Page Rank = Rede de Links
- Indexação das páginas na Internet: Google 79%; Yahoo/Altavista 69%; MSN 62%; Ask/Teoma 58%. Sobreposição de 28,85% (Gulli & Signorin, 2005)
- Mecanismos de busca permitem levantamentos amplos de relações entre sítios na Web (Wilkinson et al., 2003)
- Somente os mecanismos de busca viabilizam determinados estudos Webométricos (Kim et al., 2006)

# Ferramentas Webométricas

## Mecanismos de Busca

- A escolha por um determinado mecanismo de busca para coleta dos dados, geralmente, leva em consideração dois aspectos
  - o percentual de páginas indexadas
  - os operadores booleanos permitidos para a consulta.
- Os booleanos permitem que sejam realizadas buscas avançadas e sua complexidade varia em cada mecanismo.



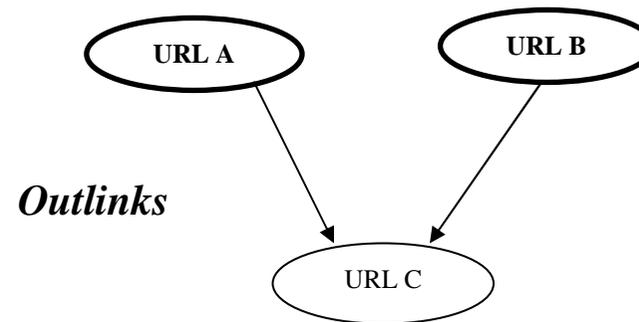
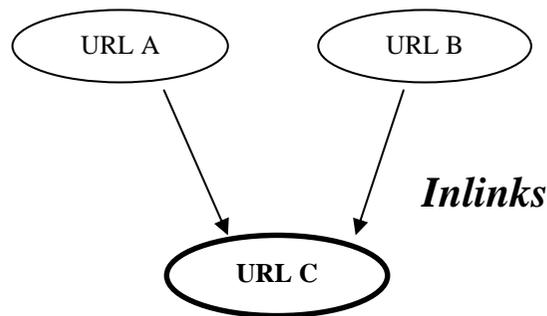
# Tipos de Estudos Webométricos

- Fator de Impacto na *Web* – Web IF
  - Ingwersen (1998)
  - eWIF - modificado por Thelwall (2001)
- Análise de *Co-Link*
  - Larson (1996)
- Análise de Interlink
  - Thelwall (2002)

# Tipos de Estudos Webométricos

Diferentes tipologias de *links*

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

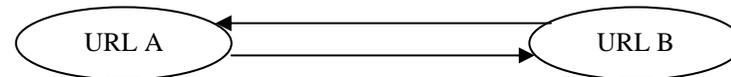


# Tipos de Estudos Webométricos

## Diferentes tipologias de *links*

### ● *Interlinks* (Thelwall, 2002)

- são os *links* trocados entre duas URLs.



### ● *Co-links* (Larson, 1996) ou *co-citations* (Herrero-Solana & Morales-del-Castillo, 2004)

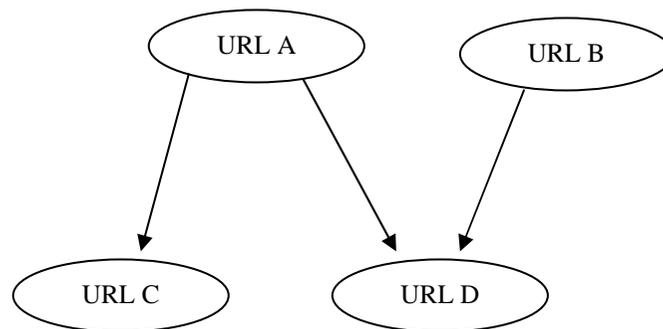
- ocorrem quando uma URL aponta simultaneamente para duas ou mais URLs.

# Tipos de Estudos Webométricos

## Diferentes tipologias de *links*

### ● *Co-outlinks e co-inlinks*

- Björneborn & Ingwersen (2004)



# Tipos de Estudos Webométricos

## Cláusulas ou *strings* de busca

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

# Exemplos

Fator de Impacto na Web externo

*External Web Impact Factor*

*eWIF*

AltaVista Search: host:www.invivo.fiocruz.br - Windows Internet Explorer

http://www.altavista.com/web/results?itag=ody&q=host%3Awww.invivo.fiocruz.br&kgs=0&ks=0

AltaVista Search: host:www.invivo.fiocruz.br

altavista Web Images MP3/Audio Video News Family Filter: off Help

host:www.invivo.fiocruz.br **FIND** [Advanced Search](#) [Settings](#)

SEARCH:  Worldwide or [Select a country](#) RESULTS IN:  All languages  English

AltaVista found 379 results

[Invivo](#)  
Fale com a Fiocruz. Busca. no Portal. neste Site ...  
[www.invivo.fiocruz.br](http://www.invivo.fiocruz.br) • [Translate](#)

[quiz01](#)  
[www.invivo.fiocruz.br/quiz.html](http://www.invivo.fiocruz.br/quiz.html) • [Translate](#)

[The Cell](#)  
Today Malpighi is considered the precursor of embriology ... 1 2 3 4 5 6 7 8 9 10 11 12. History of the microscope Cell Theory The Cell, workshop of life ...  
[www.invivo.fiocruz.br/cell/teoria\\_01.htm](http://www.invivo.fiocruz.br/cell/teoria_01.htm)

[Exposição Dengue](#)  
português english ...  
[www.invivo.fiocruz.br/dengue/home\\_dengue.htm](http://www.invivo.fiocruz.br/dengue/home_dengue.htm) • [Translate](#)

[Exposição Dengue](#)  
Cartoon published in. the Chanteclair, Paris, 1911. Public Calamity. At the beginning of last century, when Oswaldo Cruz was violently struggling to ... 1 2 3 ...  
[www.invivo.fiocruz.br/dengue/introducao\\_ig.htm](http://www.invivo.fiocruz.br/dengue/introducao_ig.htm)

[Invivo ::: Seção: Ciência](#)  
Bob Esponja, Meninas Superpoderosas, Shrek, A Noiva-Cadáver... Você já parou para pensar em como esses personagens ganham vida? ...  
[www.invivo.fiocruz.br/cgi/cgilua.exe/sys/start.htm?sid=9](http://www.invivo.fiocruz.br/cgi/cgilua.exe/sys/start.htm?sid=9) • [Translate](#)

[Invivo ::: Seção: Saúde](#)

Concluído Internet 100%

# Exemplo de dados de eWIF

Nome da Instituição	País	URL	Abbr.	Host	Link NOT Host	eWIF
Science Museum	Reino Unido	www.sciencemuseum.org.uk	SCIM-UK	16300	48000	2,9448
Deutsches Museum	Alemanha	www.deutsches-museum.de	DMUS-DE	2840	18900	6,6549
Ars Electronica Center	Áustria	www.aec.at	AREC-AT	44400	14900	0,3356
Ciudad de las Artes y las Ciencias de Valencia	Espanha	www.cac.es	CACV-ES	1430	14700	10,2797
Eden Project	Reino Unido	www.edenproject.com	EPRJ-UK	1430	9770	6,8322
Cité des Sciences et de l'Industrie	França	www.cite-sciences.fr	CSID-FR	9540	9690	1,0157
National Museum of Emerging Science and Innovation	Japão	www.miraikan.jst.go.jp	NMES-JP	542	9210	16,9926
Ontario Science Centre	Canadá	www.ontariosciencecentre.ca	OTSC-CA	6070	7390	1,2175
Casa de las Ciencias	Espanha	www.casaciencias.org	CCLC-ES	1520	7060	4,6447
Museo Nazionale della Scienza e della Tecnologia	Itália	www.museosciienza.org	MNST-IT	4600	6530	1,4196
Museum of Science and Industry in Manchester	Reino Unido	www.msims.org.uk	MSIM-UK	665	5760	8,6617
Maritime Museum Rotterdam	Holanda	www.maritiemmuseum.nl	MMRT-NL	1440	4860	3,3750
Palais de la Découverte	França	www.palais-decouverte.fr	PLDC-FR	763	4770	6,2516
Bloomfield Science Museum Jerusalem	Israel	www.mada.org.il	BSMU-IL	917	4410	4,8092
Questacon	Austrália	www.questacon.edu.au	QTCN-AU	1440	4400	3,0556
Heureka - Finnish Science Centre	Finlândia	www.heureka.fi	HFSC-FI	1300	4360	3,3538
Parque de las Ciencias	Espanha	www.parqueciencias.com	PQCI-ES	470	3750	7,9787
New York Hall of Science	EUA	www.nyhallsci.org	NYHS-US	662	3020	4,5619
W5 Online	Reino Unido	www.w5online.co.uk	WSOL-UK	108	2960	27,4074
Experimentarium	Dinamarca	www.experimentarium.dk	EXPM-DK	15000	2850	0,1900
Science Museum of Virginia	EUA	www.smv.org	SMVG-US	945	2800	2,9630
Deutsches Hygiene-Museum	Alemanha	www.dhmd.de	DHMD-DE	852	2600	3,0516
Explore-At-Bristol	Reino Unido	www.at-bristol.org.uk	ATBR-UK	1690	2350	1,3905
Technorama Swiss Science Center	Suiça	www.technorama.ch	TSSC-CH	2870	2290	0,7979
National Space Centre	Reino Unido	www.spacecentre.co.uk	NSPC-UK	396	2280	5,7576

# Exemplo comparativo de eWIF para diferentes associações de museus e centros de ciência

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

# Exemplos

Análise de *Co-link*  
*Co-link Analysis*

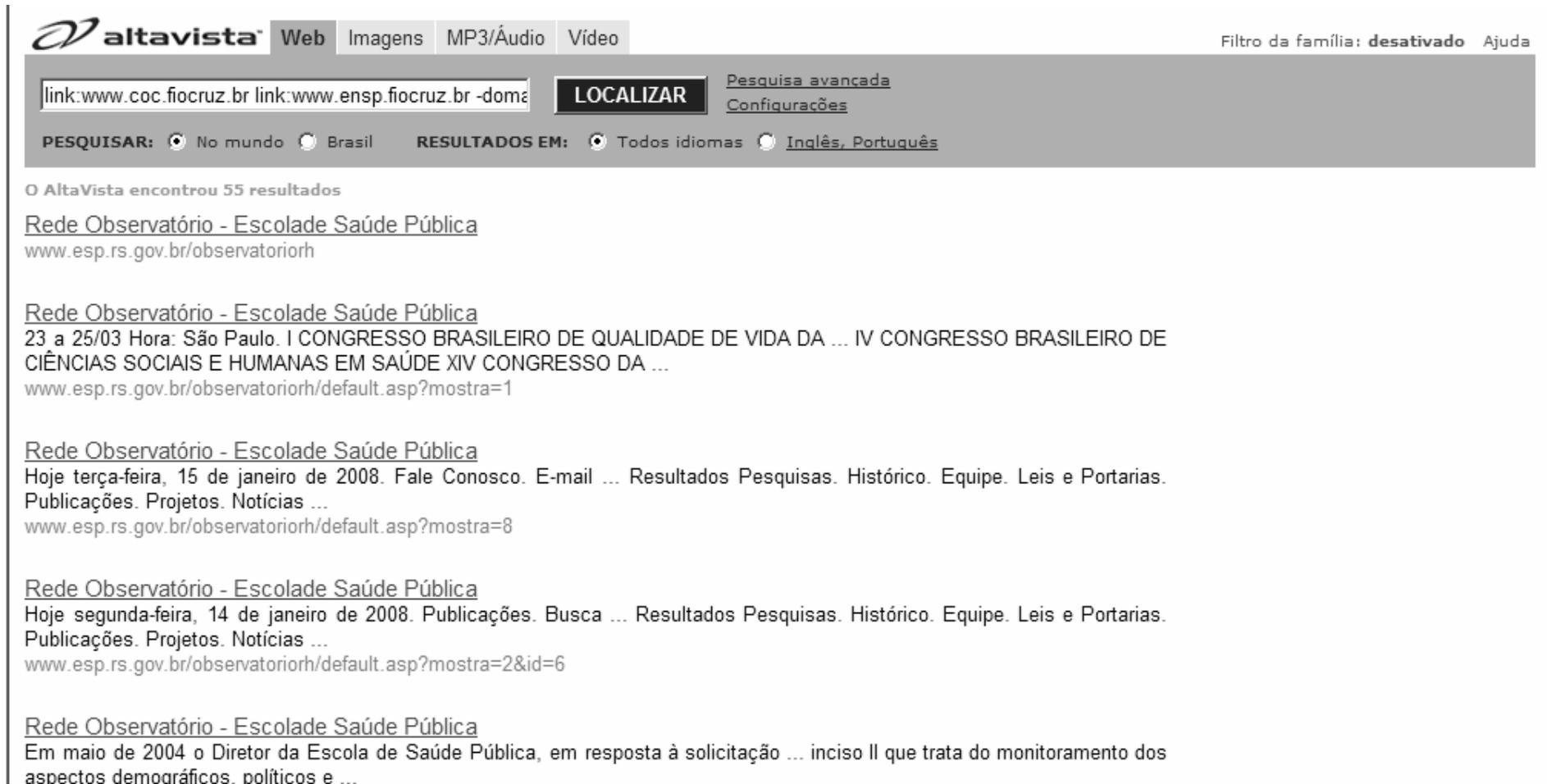
# Exemplo de matriz para coleta de *co-links*

<i>Link:URL1 AND Link:URL2</i>			
URL1 = www.website.com.br e URL2 = www.outro.com.br			
<i>Link:www.website.com.br AND Link:www.outro.com.br</i>			

	URL1	URL2	URL3
URL1	$\Sigma$ da coluna	Link:URL1 AND Link:URL2	Link:URL1 AND Link:URL3
URL2	Link:URL2 AND Link:URL1	$\Sigma$ da coluna	Link:URL2 AND Link:URL3
URL3	Link:URL3 AND Link:URL1	Link:URL3 AND Link:URL2	$\Sigma$ da coluna

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

# Exemplo de levantamento de *co-link*



The image shows a screenshot of an AltaVista search engine interface. At the top left is the AltaVista logo. To its right are navigation tabs for 'Web', 'Imagens', 'MP3/Áudio', and 'Vídeo'. On the far right, it says 'Filtro da família: desativado' and 'Ajuda'. Below the logo is a search bar containing the text 'link:www.coc.fiocruz.br link:www.ensp.fiocruz.br -doma'. To the right of the search bar is a 'LOCALIZAR' button. Further right are links for 'Pesquisa avançada' and 'Configurações'. Below the search bar, there are radio buttons for 'PESQUISAR:' with options 'No mundo' (selected) and 'Brasil', and 'RESULTADOS EM:' with options 'Todos idiomas' (selected) and 'Inglês, Português'. The search results section begins with the text 'O AltaVista encontrou 55 resultados'. The first result is titled 'Rede Observatório - Escolade Saúde Pública' with the URL 'www.esp.rs.gov.br/observatoriorh'. The second result is also titled 'Rede Observatório - Escolade Saúde Pública' and includes the text '23 a 25/03 Hora: São Paulo. I CONGRESSO BRASILEIRO DE QUALIDADE DE VIDA DA ... IV CONGRESSO BRASILEIRO DE CIÊNCIAS SOCIAIS E HUMANAS EM SAÚDE XIV CONGRESSO DA ...' and the URL 'www.esp.rs.gov.br/observatoriorh/default.asp?mostra=1'. The third result is titled 'Rede Observatório - Escolade Saúde Pública' and includes the text 'Hoje terça-feira, 15 de janeiro de 2008. Fale Conosco. E-mail ... Resultados Pesquisas. Histórico. Equipe. Leis e Portarias. Publicações. Projetos. Notícias ...' and the URL 'www.esp.rs.gov.br/observatoriorh/default.asp?mostra=8'. The fourth result is titled 'Rede Observatório - Escolade Saúde Pública' and includes the text 'Hoje segunda-feira, 14 de janeiro de 2008. Publicações. Busca ... Resultados Pesquisas. Histórico. Equipe. Leis e Portarias. Publicações. Projetos. Notícias ...' and the URL 'www.esp.rs.gov.br/observatoriorh/default.asp?mostra=2&id=6'. The fifth result is titled 'Rede Observatório - Escolade Saúde Pública' and includes the text 'Em maio de 2004 o Diretor da Escola de Saúde Pública, em resposta à solicitação ... inciso II que trata do monitoramento dos aspectos demográficos, políticos e ...'.

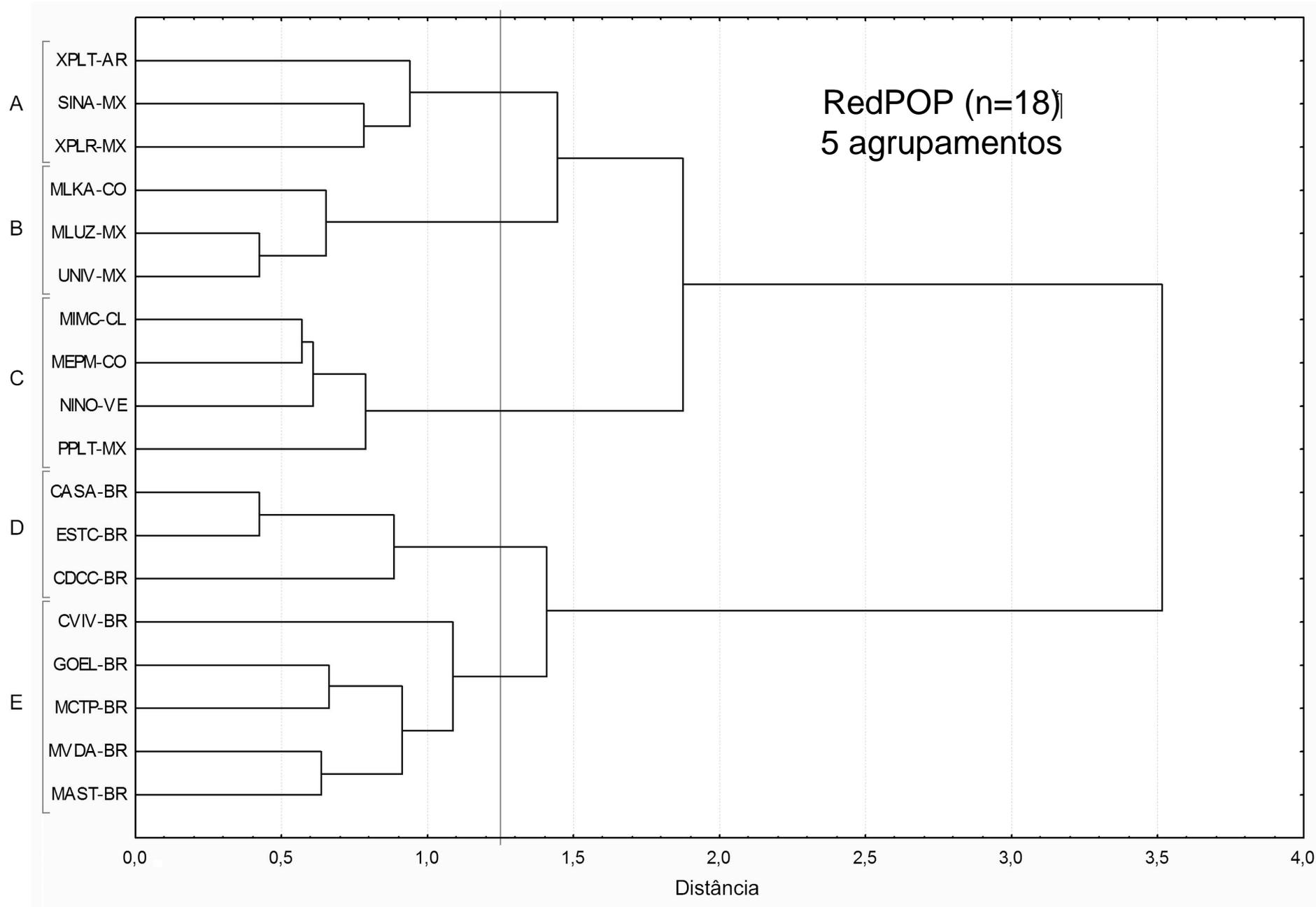
# *Red de Popularización de la Ciencia y la Tecnología para América Latina y el Caribe (RedPOP)*



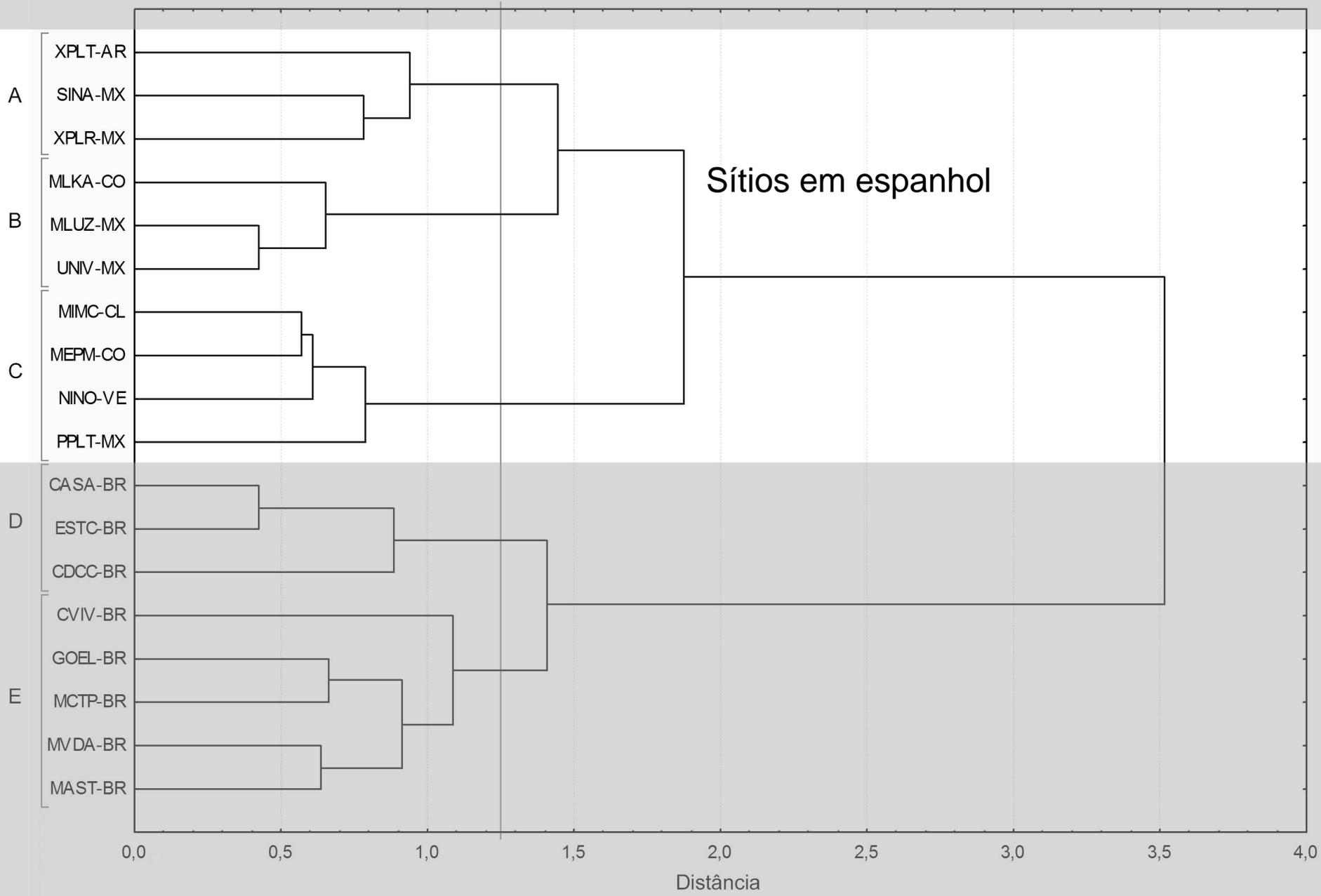
RedPOP - *Red de Popularización de la Ciencia y la Tecnología para América Latina y el Caribe*

# Matriz de *co-link*

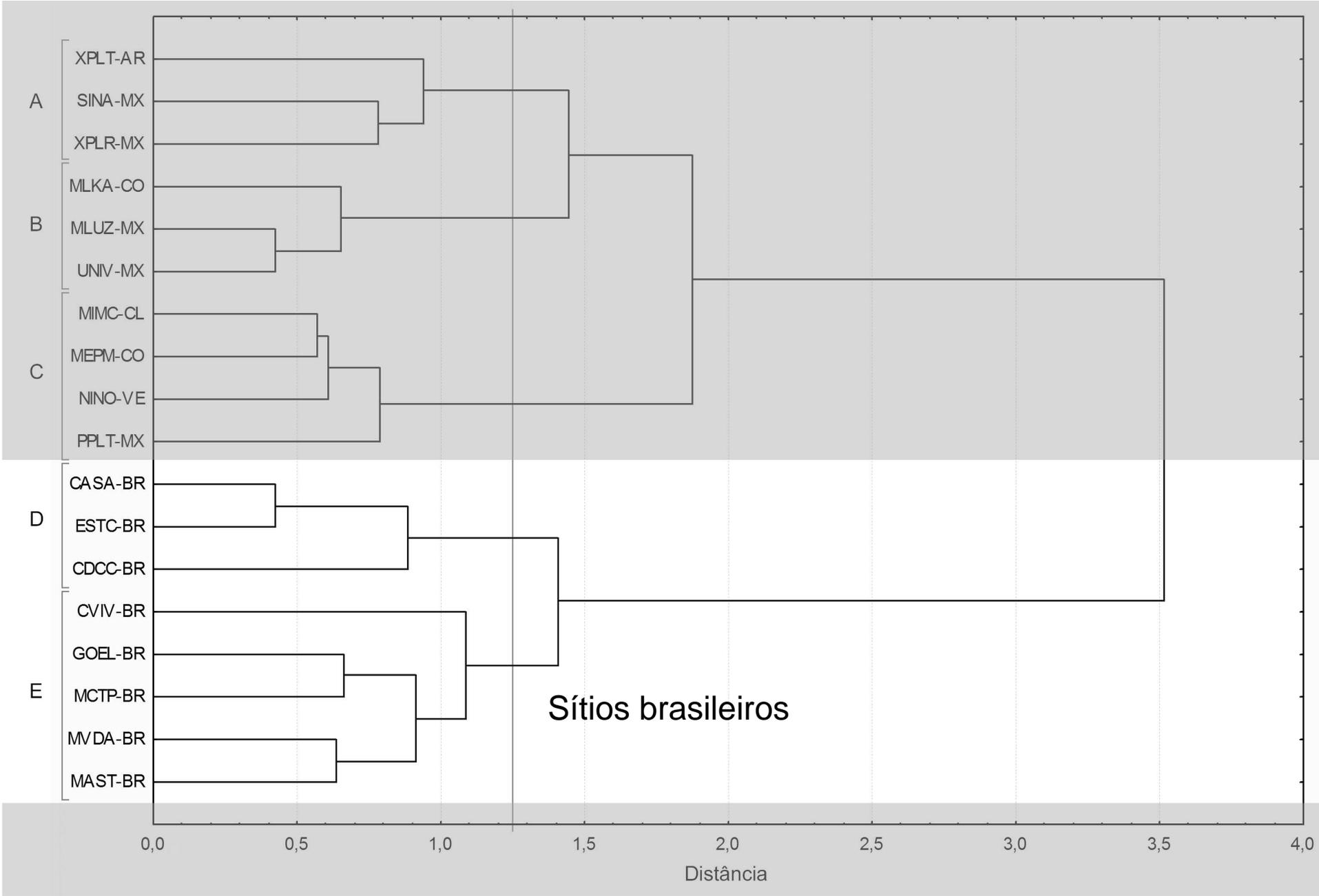
	XPLT-AR	CASA-BR	CDCC-BR	CVIV-BR	ESTC-BR	GOEL-BR	MVDA-BR	MAST-BR	MCTP-BR	MIMC-CL	MLKA-CO	MEPM-CO	SINA-MX	XPLR-MX	MLUZ-MX	PPLT-MX	UNIV-MX	NINO-VE
XPLT-AR	107	10	0	8	10	0	0	1	0	2	12	3	14	14	2	16	15	0
CASA-BR	10	331	18	20	132	18	19	34	24	0	10	0	10	8	2	12	12	0
CDCC-BR	0	18	114	3	37	11	6	21	12	0	2	0	0	0	0	3	3	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	0	65	66	270



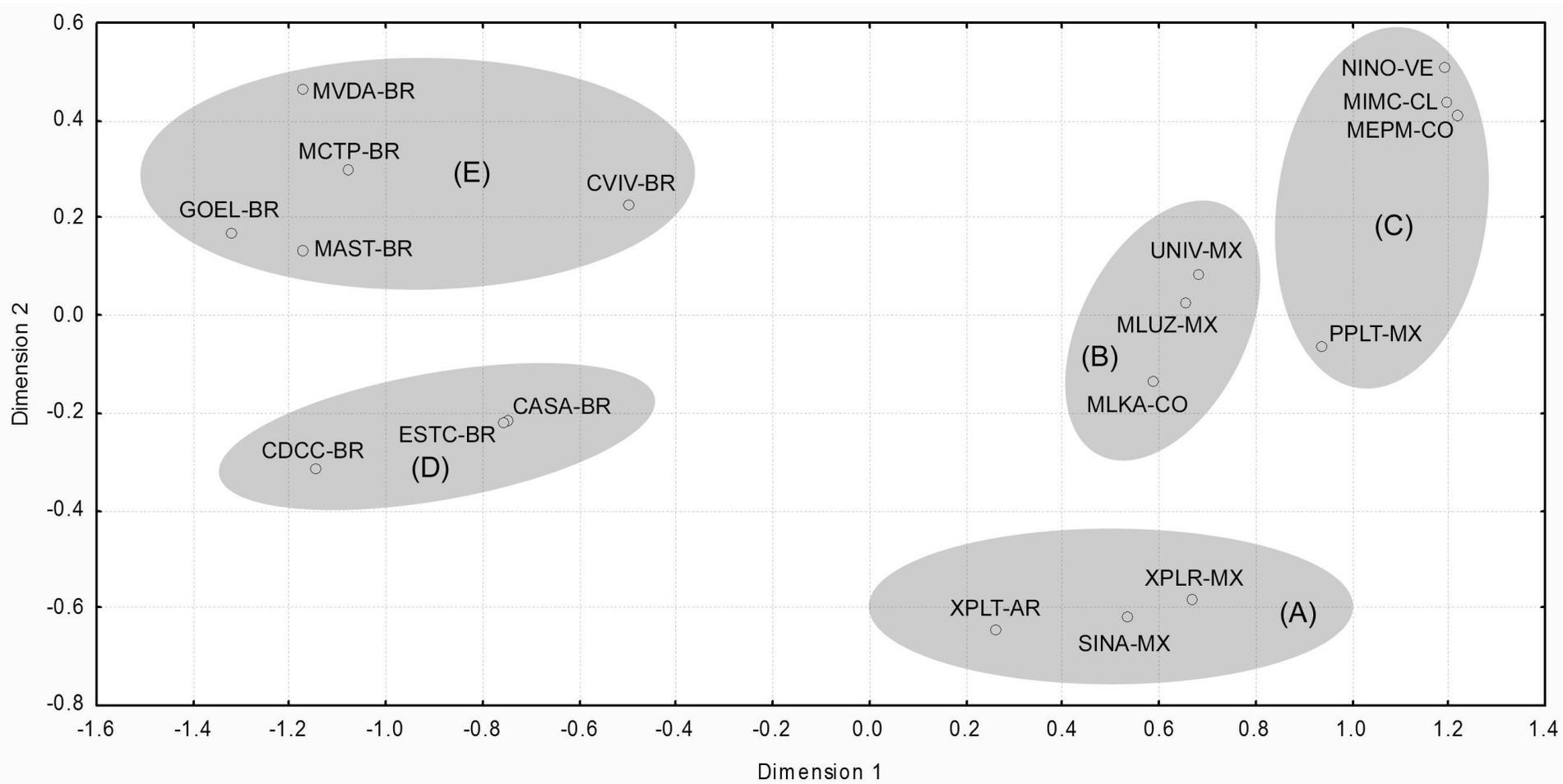
Gouveia & Kurtenbach (2009)



Gouveia & Kurtenbach (2009)

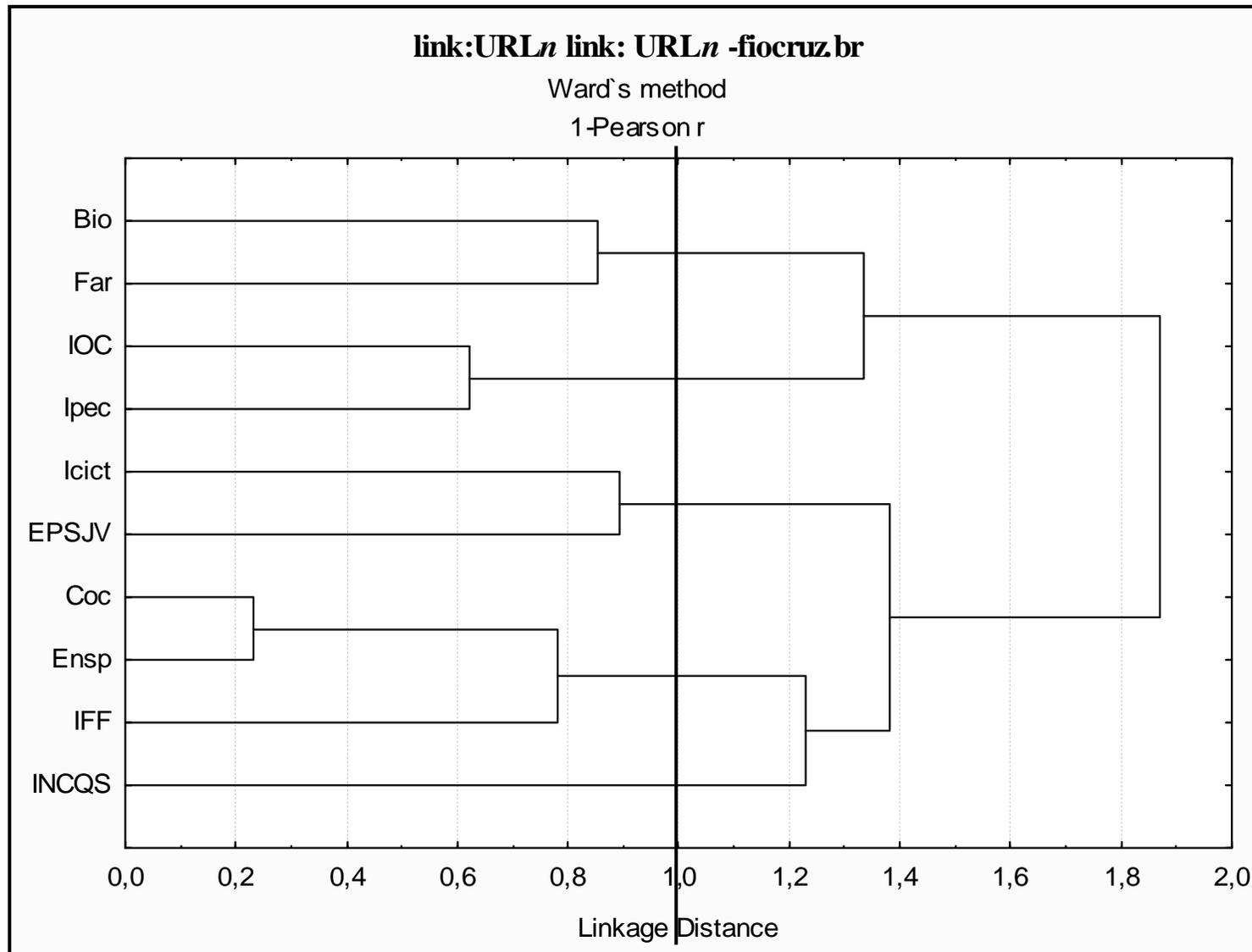


Gouveia & Kurtenbach (2009)



Gouveia & Kurtenbach (2009)

# Co-link de sítios de unidades da Fiocruz



Lang, Gouveia & Leta

## Analise de *co-link*

Quais os fatores mais importantes?

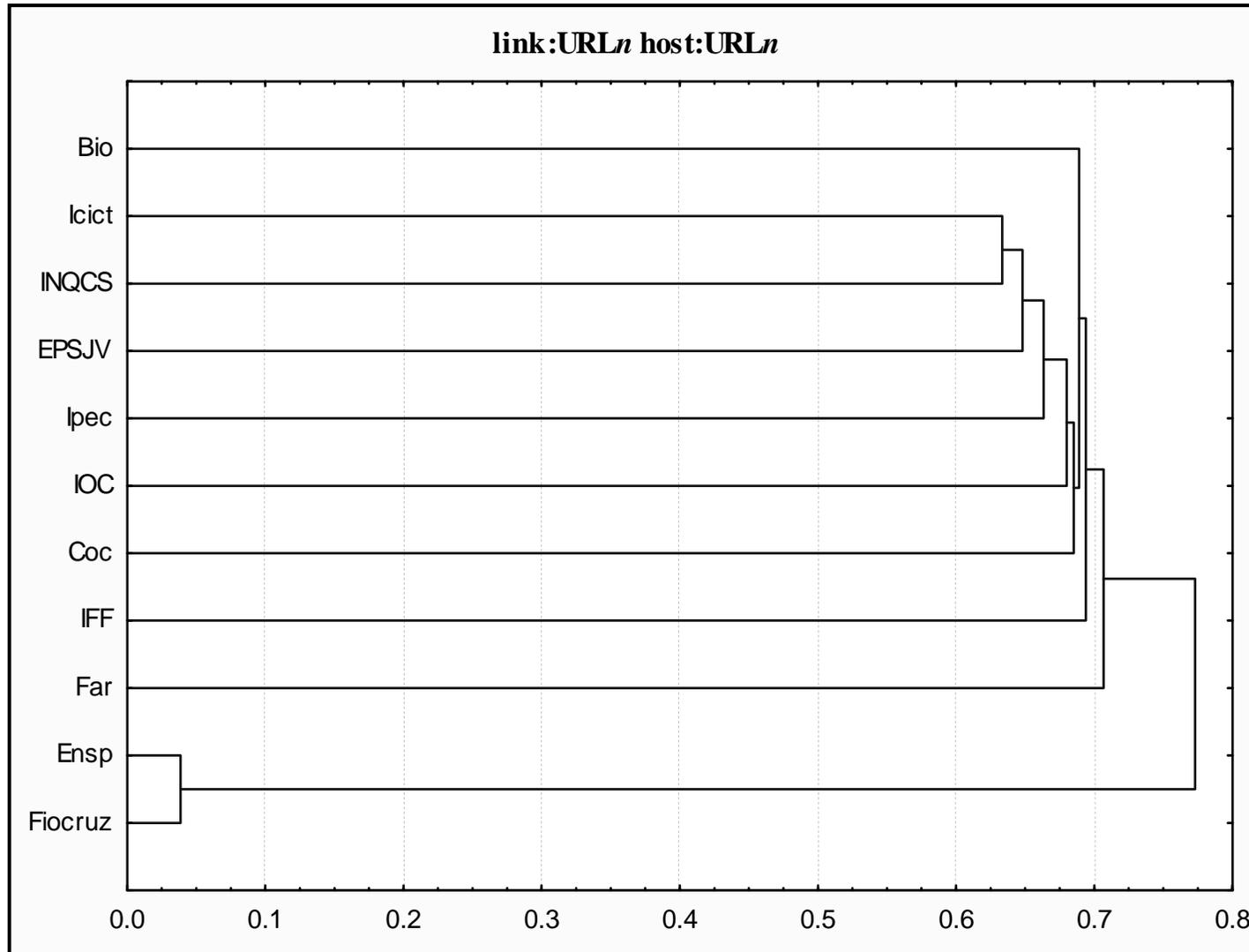
- Idioma comum
- Idiomas disponíveis no sítio
- Distância geográfica
- Relação política
- Proximidade temática
- Associações e redes
- Projetos de cooperação

# Exemplos

Análise de Interlink

*Interlink analysis*

# *Interlinking* de unidades da Fiocruz



# Exemplos

*Ranking de Universidades na Web*  
*Universities Web Ranking*

# Ranking Web of Latin American Universities

July '08

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

> home > top Brazil

- Data
- About Us
- About the Ranking
- Top 200 Universities
- Top Brazil
- Top Andean Pact
- Top Southern Cone
- Top Central America-Caribe
- Best Practices
- Notes

- Catalogue
- Universities by country
- R&D Centres by country

- Information
- Methodology

## Top Brazil

Universities 1 to 100 of 100

RANK	UNIVERSITY	COUNTRY	WORLD RANK
1	Universidade de Sao Paulo		113
2	Universidade Estadual de Campinas		212
3	Universidade Federal do Rio de Janeiro		330
4	Universidade Federal de Santa Catarina Brasil		381
5	Universidade Federal do Rio Grande do Sul		388
6	Universidade Federal de Minas Gerais		539
7	Universidade de Brasilia		543
8	Universidade Estadual Paulista		571
9	Universidade Federal do Parana		604
10	Pontificia Universidade Católica do Rio de Janeiro		615
11	Universidade Federal da Bahia		692

# Limitações e Desafios

- 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 e Desafios

## ● Os sítios e os *links*

- 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* mantêm 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

## ● 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

# Os Aspéctos Teóricos e Metodológicos da Cienciometria

## Webometria: aplicações e desafios

1º Encontro Brasileiro de Bibliometria e Cienciometria  
14 a 16 de setembro de 2008

Fábio Castro Gouveia

MV-COC-Fiocruz

[fgouveia@coc.fiocruz.br](mailto:fgouveia@coc.fiocruz.br)

15 de setembro de 2008