



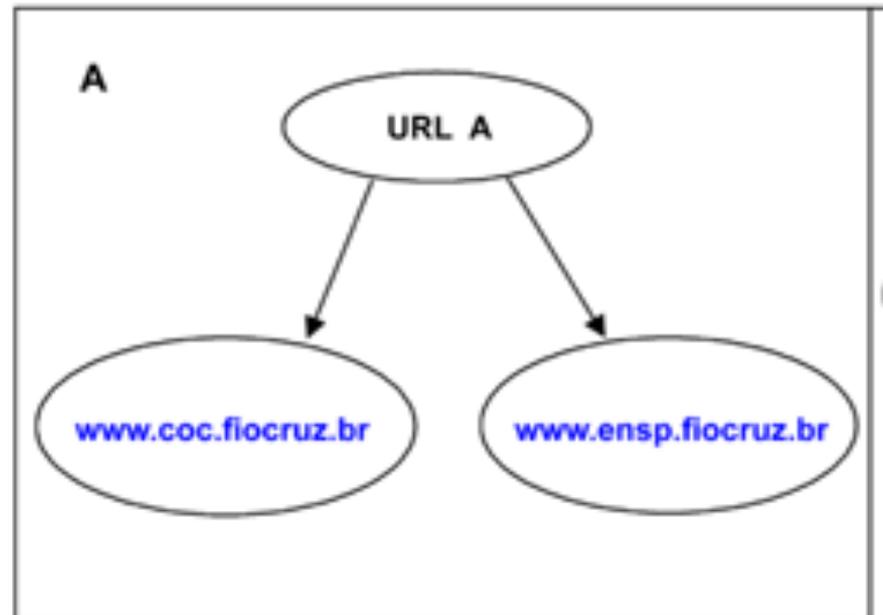
Examining the robustness of web co-link analysis

Liwen Vaughan e Juan Tang and Jian Du

Anne Clinio e Rafaela Giodano

Co-link

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



Multidimensional scaling (MDS)

"Conjunto de técnicas estatísticas utilizadas para a visualização da informação, explorando as similaridades ou "dissimilaridades" dos dados".

Sobre a pesquisa

- Objetivo: Avaliar a robustez da análise de co-links para a inteligência de negócios (business intelligence)
- Metodologia: Multidimensional Scalling (MDS)
- Unidades de pesquisa: indústrias chinesas
 - Tecnologia da informação/eletrônicos
 - Indústria química
- Dois momentos:
 - Inverno de 2007
 - Outono de 2008

Por que a China?

- Grande interesse em business intelligence. Há, inclusive, uma organização nacional - a Sociedade de Inteligência Competitiva da China.
- Ausência de estudos sobre métodos de pesquisa em business intelligence. Ainda mais, com uso de fontes web.
- Motivação é verificar se as análises de hiperlinks usados no Ocidente seriam válidos no contexto chinês.
- País com maior número de usuários web desde 2008, quando ultrapassou os EUA.

Passo a passo

1º - Identificação das empresas mais relevantes

TI	Química
<p data-bbox="218 591 807 725">Ranking das top 100 empresas</p> <p data-bbox="218 811 942 1016">Fonte: Ministério da Indústria e Tecnologia de Informação</p> <p data-bbox="218 1102 513 1158">Ano: 2006</p>	<p data-bbox="983 591 1644 733">Ranking China's top 50 chemical enterprises</p> <p data-bbox="983 839 1609 901">Fonte: SRI Consulting</p> <p data-bbox="983 1093 1273 1149">Ano: 2004</p>

Passo a passo

2º - Localização dos sites das empresas pesquisadas

TI	Química
<p data-bbox="218 691 915 819">Buscaram manualmente o site de cada empresa.</p> <p data-bbox="218 911 852 962">Sites foram checados.</p>	<p data-bbox="981 591 1653 819">Endereços dos sites estavam disponíveis no ranking.</p> <p data-bbox="981 925 1599 976">Sites foram checados</p>

Passo a passo

3º - Seleção de unidades de pesquisa

TI	Química
60 empresas.	30 empresas.

OBS: A retirada de algumas empresas do estudo foi necessária para a geração de um mapa MDS que efetivamente mostrasse o panorama de cada setor.

Passo a passo

4º - Coleta de dados - Escolha da search engine

Buscador	Dificuldade
Baidu MSN China Yahoo China Google China Google	Sem filtro de internal links
Yahoo	Não realizava pesquisa de co-links

- **Buscadores utilizados:** 1) MSN Life Search e 2) Yahoo!

Passo a passo

5º - Coleta de dados - Query syntax (MSN e Yahoo!)

(link:www.abc.com) (link:www.xyz.com)

(linkdomain:www.abc.com) (linkdomain:www.xyz.com)

- Ao todo, obtiveram 4 conjuntos de dados
 - "link" com todas as páginas
 - "link" com apenas as páginas em língua chinesa
 - "linkdomain" com todas as páginas
 - "linkdomain" com apenas as págs em língua chinesa

OBS: Utilizaram dois buscadores por questão de disponibilidade de cada momento da pesquisa. Atentos às diferenças

Passo a passo

6º - Processamento de dados

- Normalização
 - Normalizaram co-links coletados para obter medida relativa do número de co-links através de **índice Jaccard**.
- Método de Escalonamento multidimensional (MDS) cria um **mapa** que permite perceber a posição da empresa no mercado.
 - Liderança
 - Principais competidores

Resultados

Sobre querys

- Nesta pesquisa, os mapas criados a partir da query "linkdomain" apresentaram os clusters com maior clareza do que os gerados pela query "link"
- Porém, estudos prévios (Vaughan and You, 2005a) apontam que a query "link" refletiria melhor a realidade das indústrias
- Estudos complementares qualitativos são necessários para entender o padrão de co-link, específico da realidade chinesa.

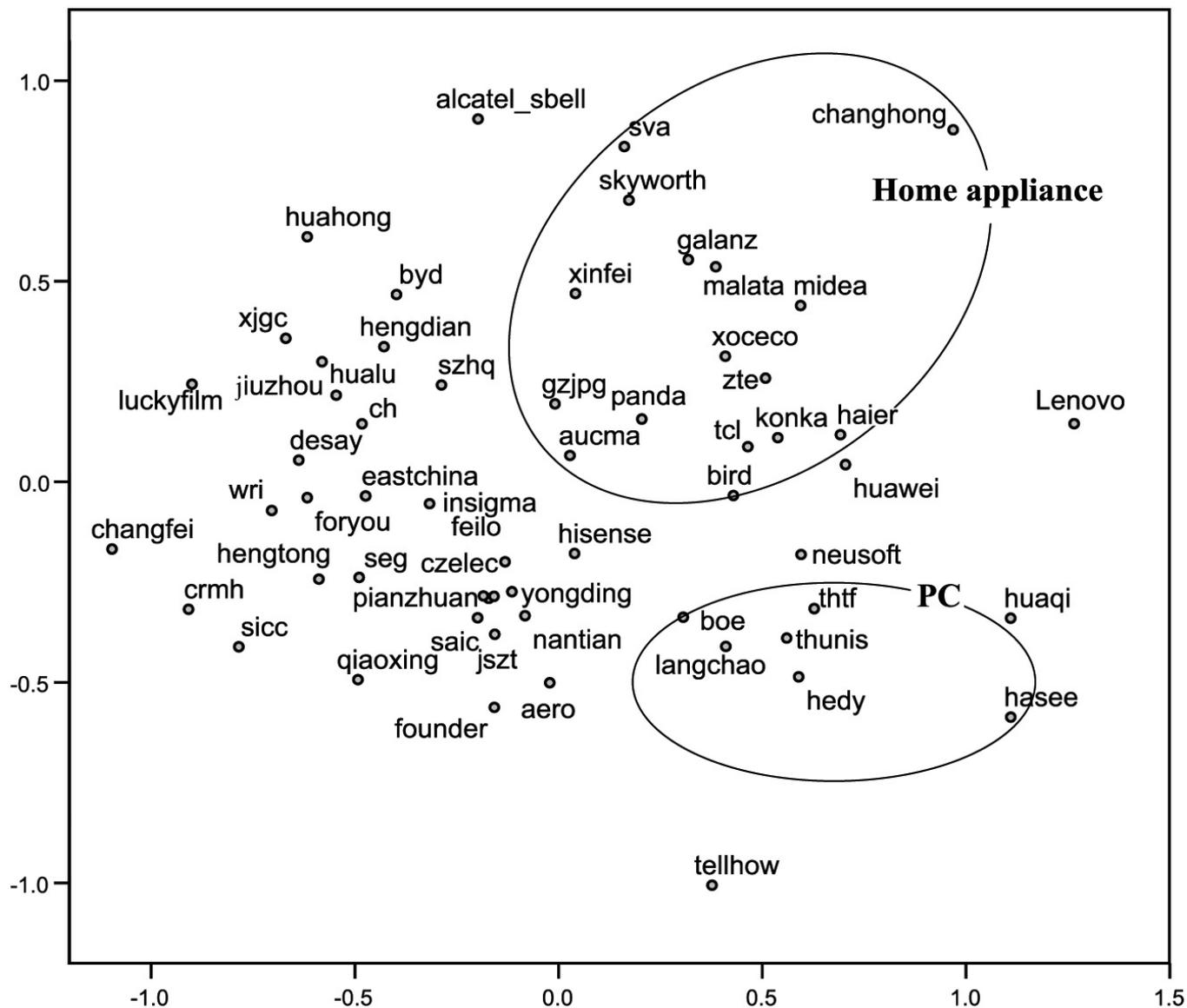
Resultados

Sobre páginas internacionais ou páginas chinesas

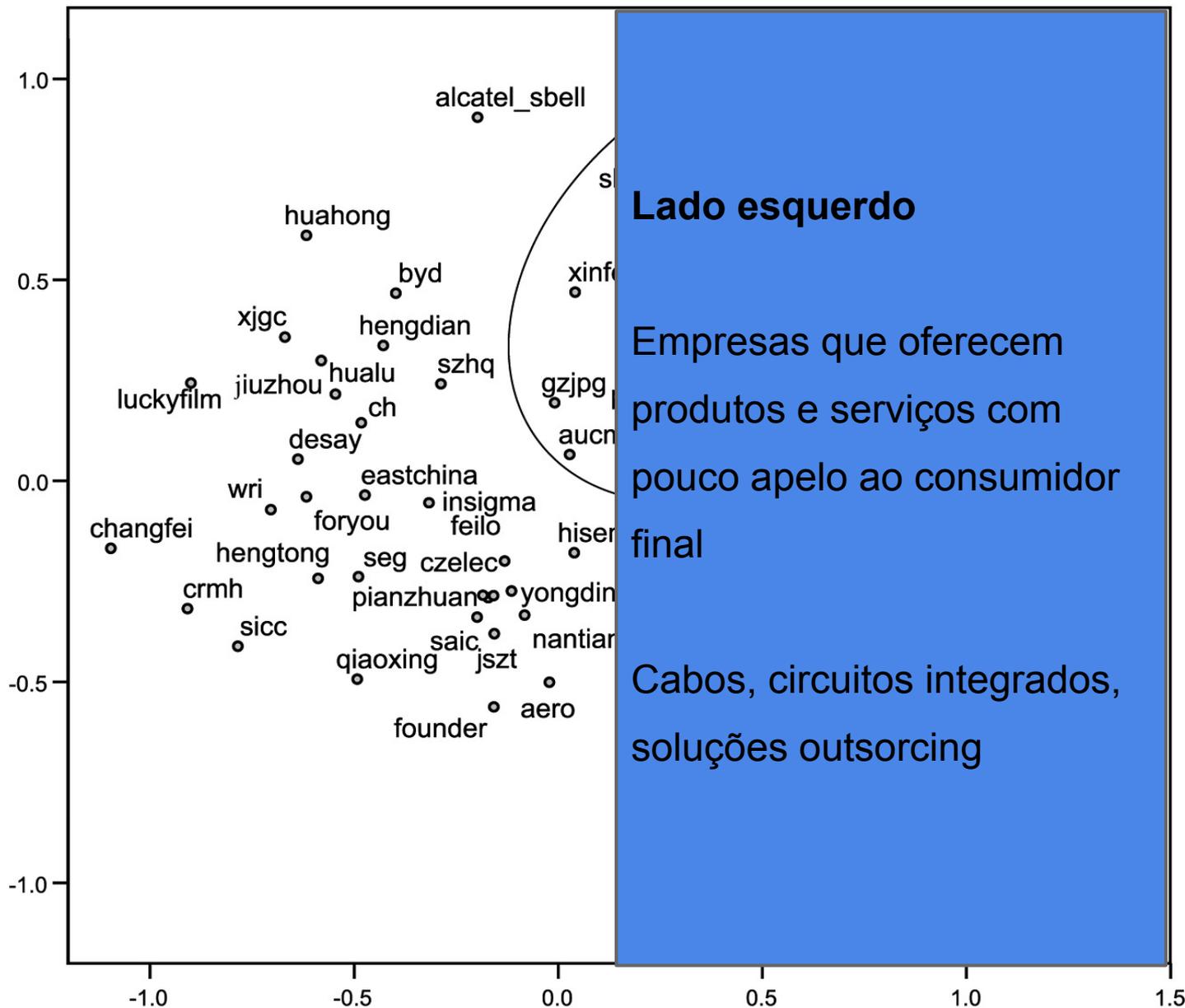
- Os mapas foram similares, mas o diagrama que representa apenas os sites chineses foi considerado melhor em ambas as indústrias.
- Possível explicação: Os sites chineses não atraem tantos links internacionais.

1ª etapa: Mapa MDS - TI

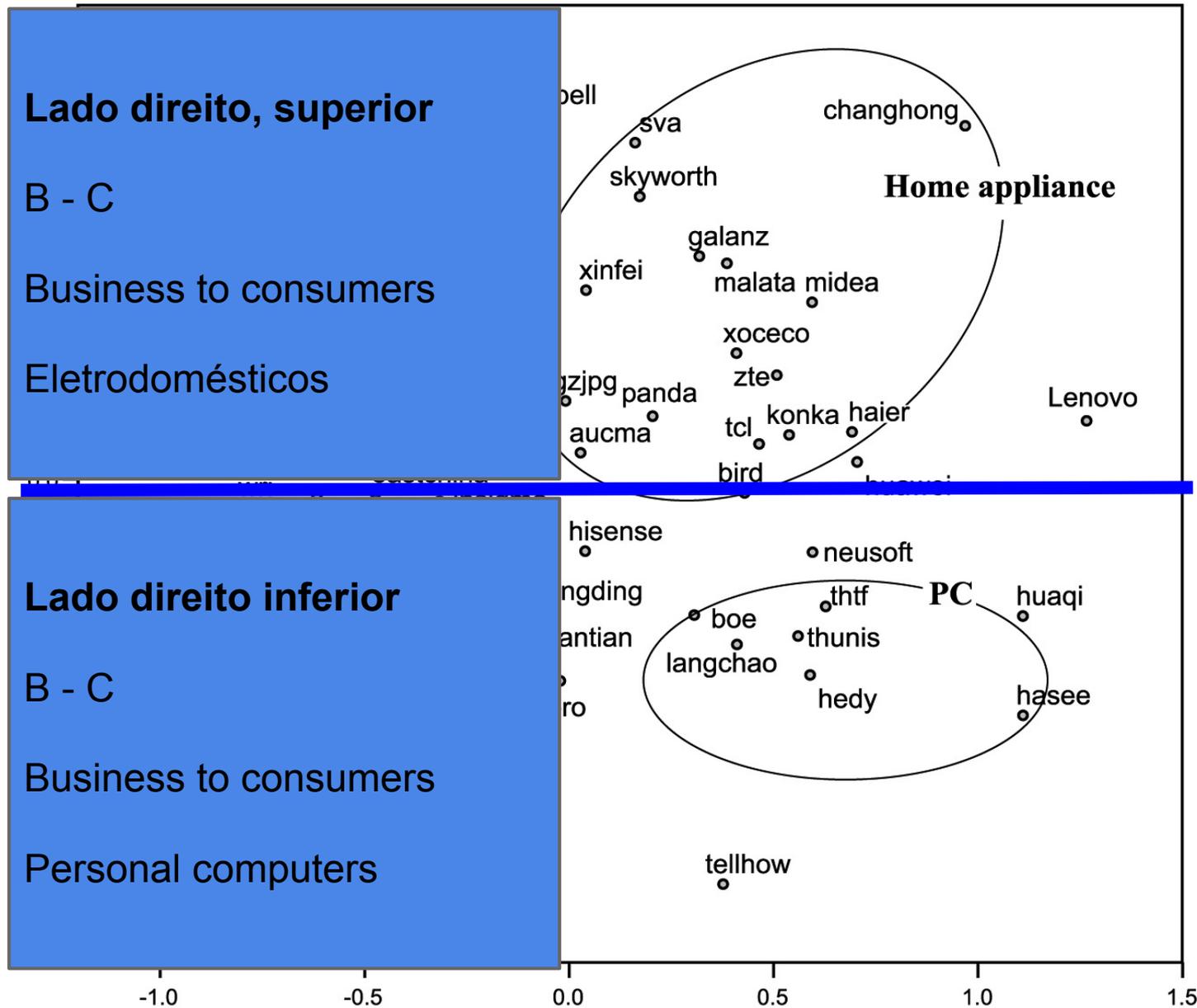
stress value: 0,05



1ª etapa: Mapa MDS - TI



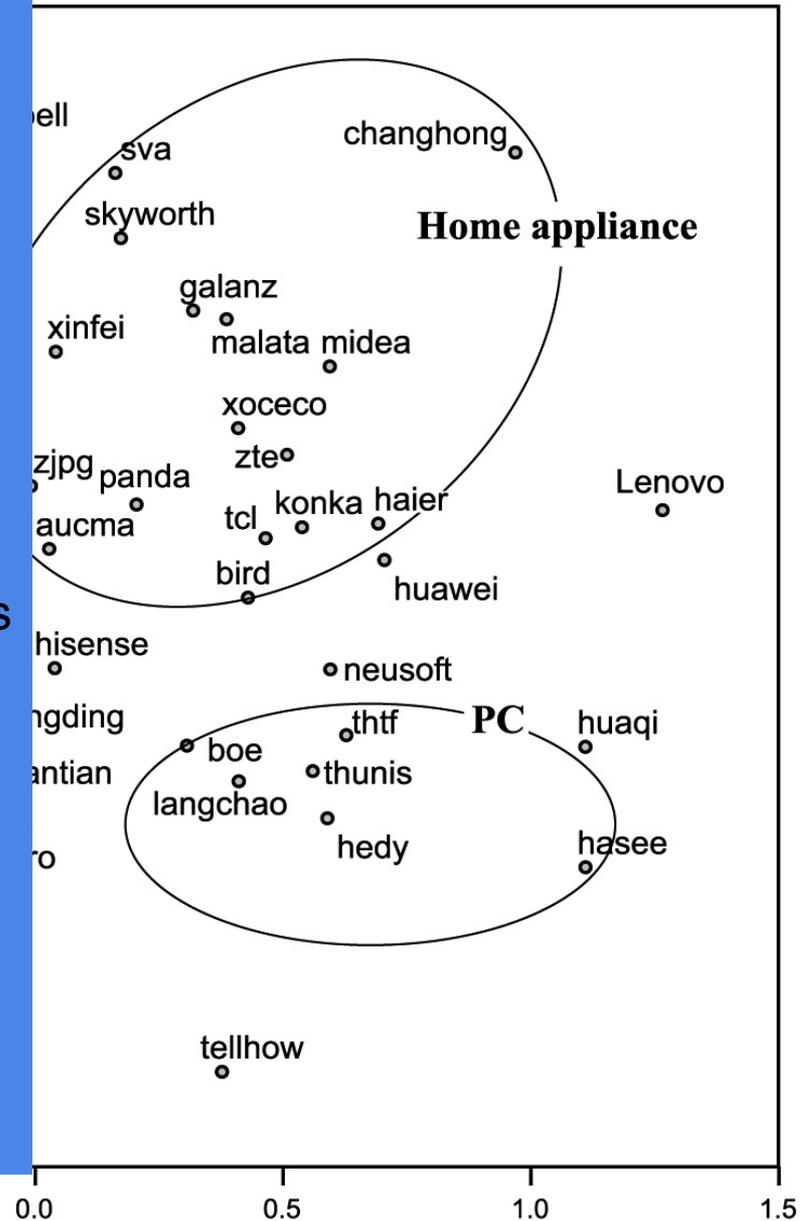
1ª etapa: Mapa MDS - TI



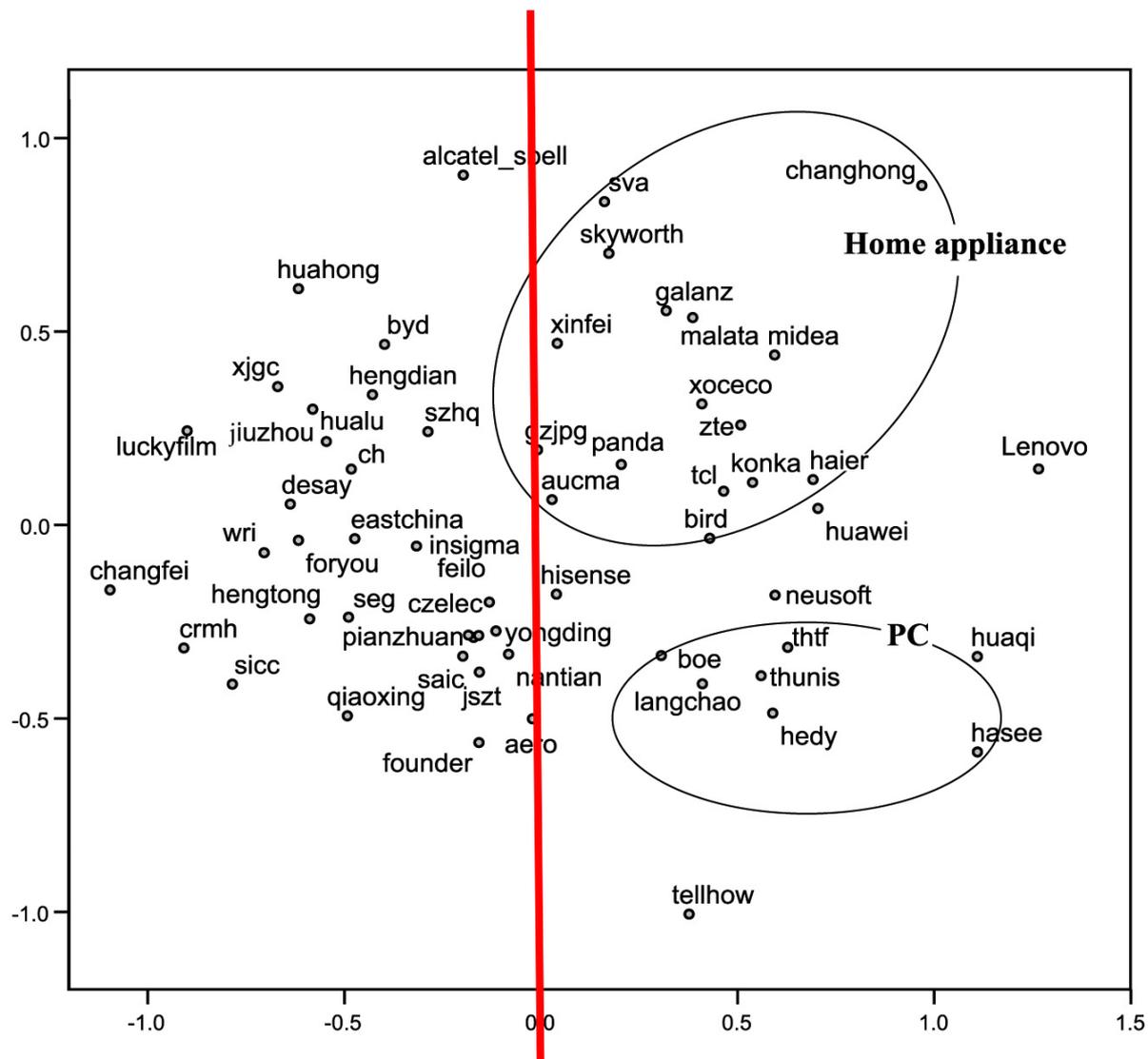
1ª etapa: Mapa MDS - TI

Lado direito

- Clusters mais definidos
- Companhias chamam mais atenção de consumidores (mais links).
- Sites chineses são orientados para o consumidor final.



1ª etapa: Mapa MDS - TI



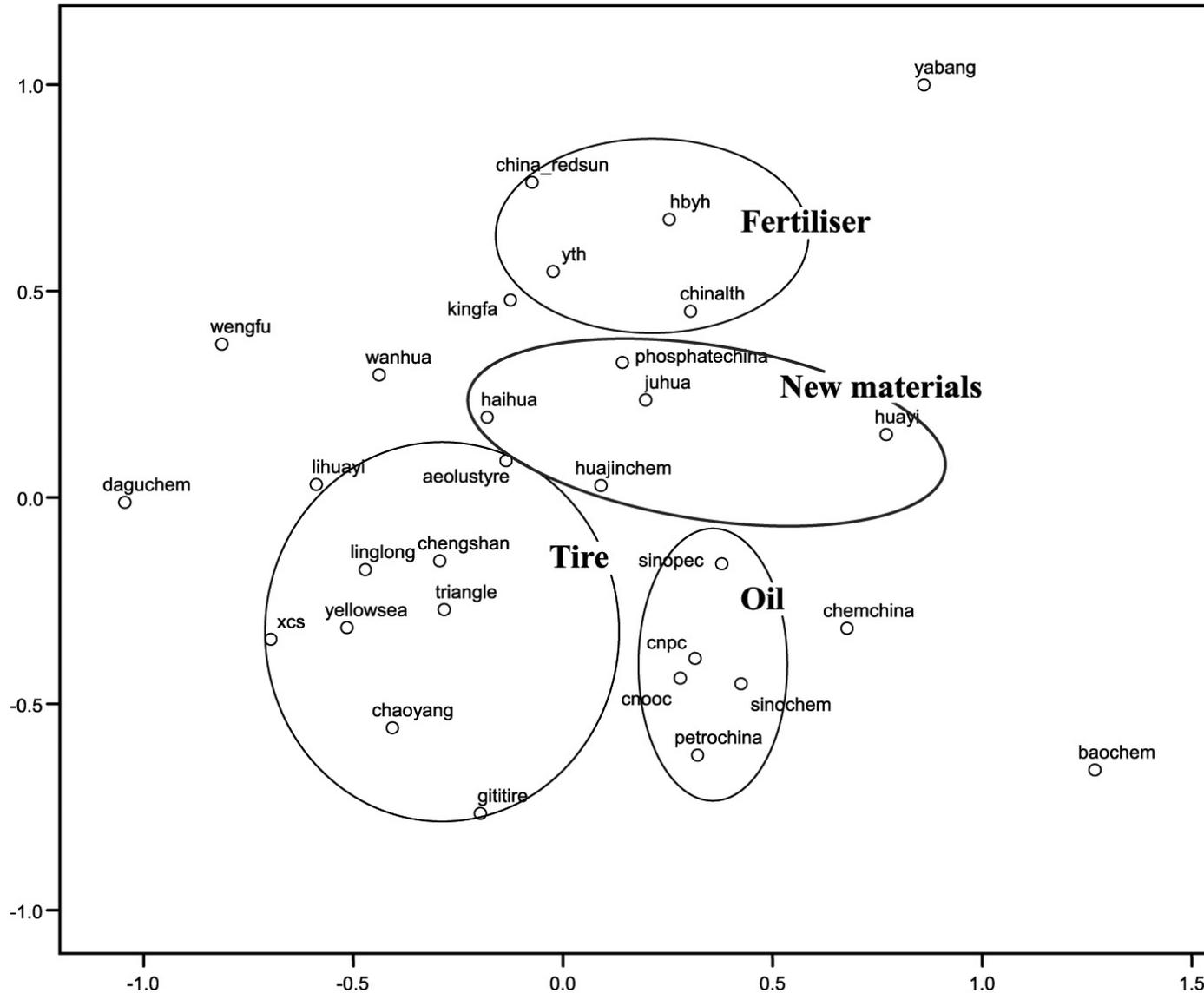
Média de inlinks:

963

8579

1ª etapa: Mapa MDS - Química

stress value: 0,04

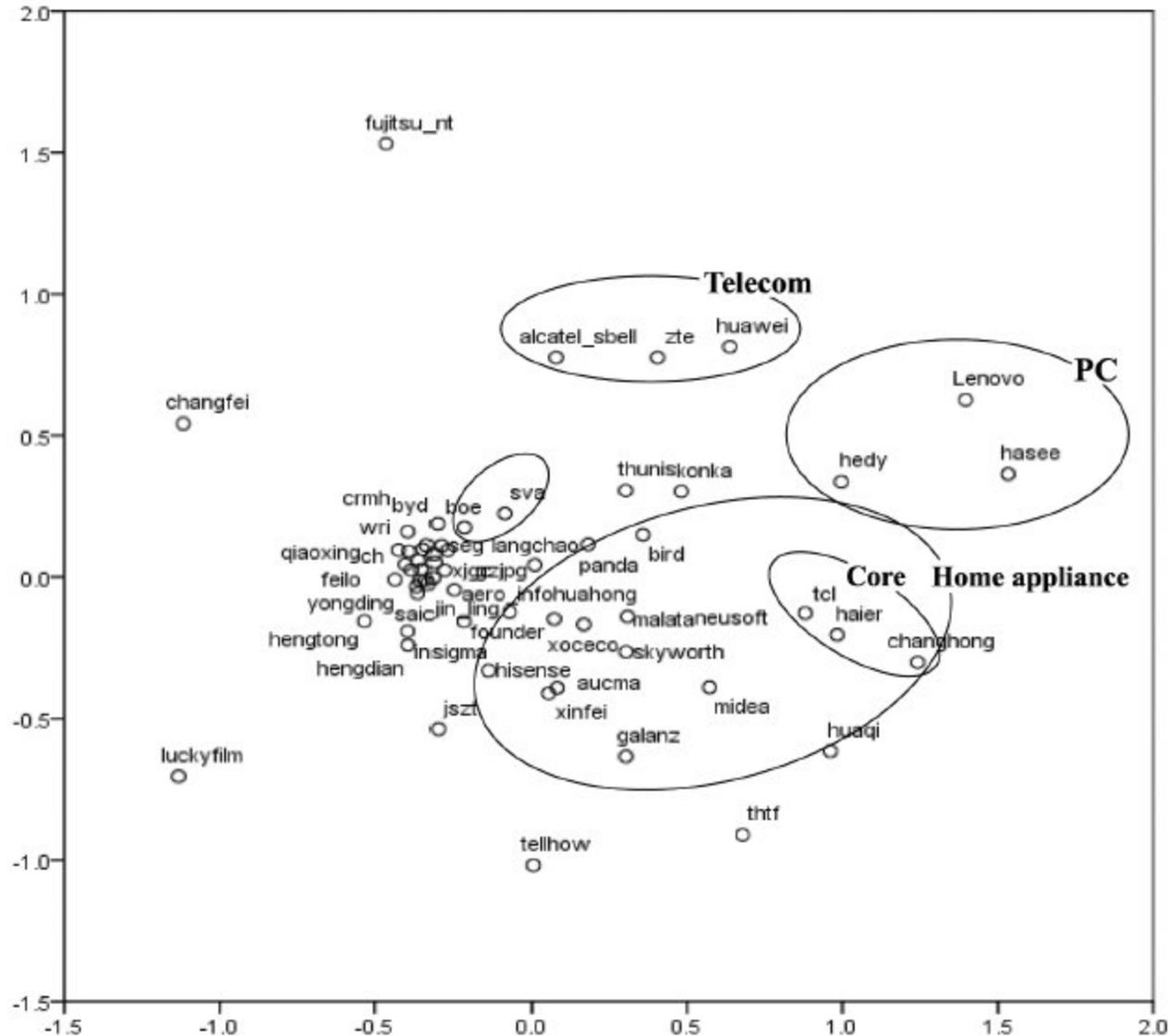


1ª etapa: Mapa MDS - Química

- Mapa apresenta quatro setores.
- Pneus e fertilizantes bem posicionados
- Clusters de novos materiais seria mais estreito se não fosse a empresa Huay.
- ChemChina está fora dos clusters porque atua em diversos segmentos.

2^a etapa: Mapa MDS - TI

stress value: 0,056



2ª etapa: Mapa MDS - TI

Mapa repete informações coletadas na 1ª etapa.

- Repetiram-se os clusters de eletrodomésticos e PC.
- Três empresas no core do setor de eletrodomésticos.

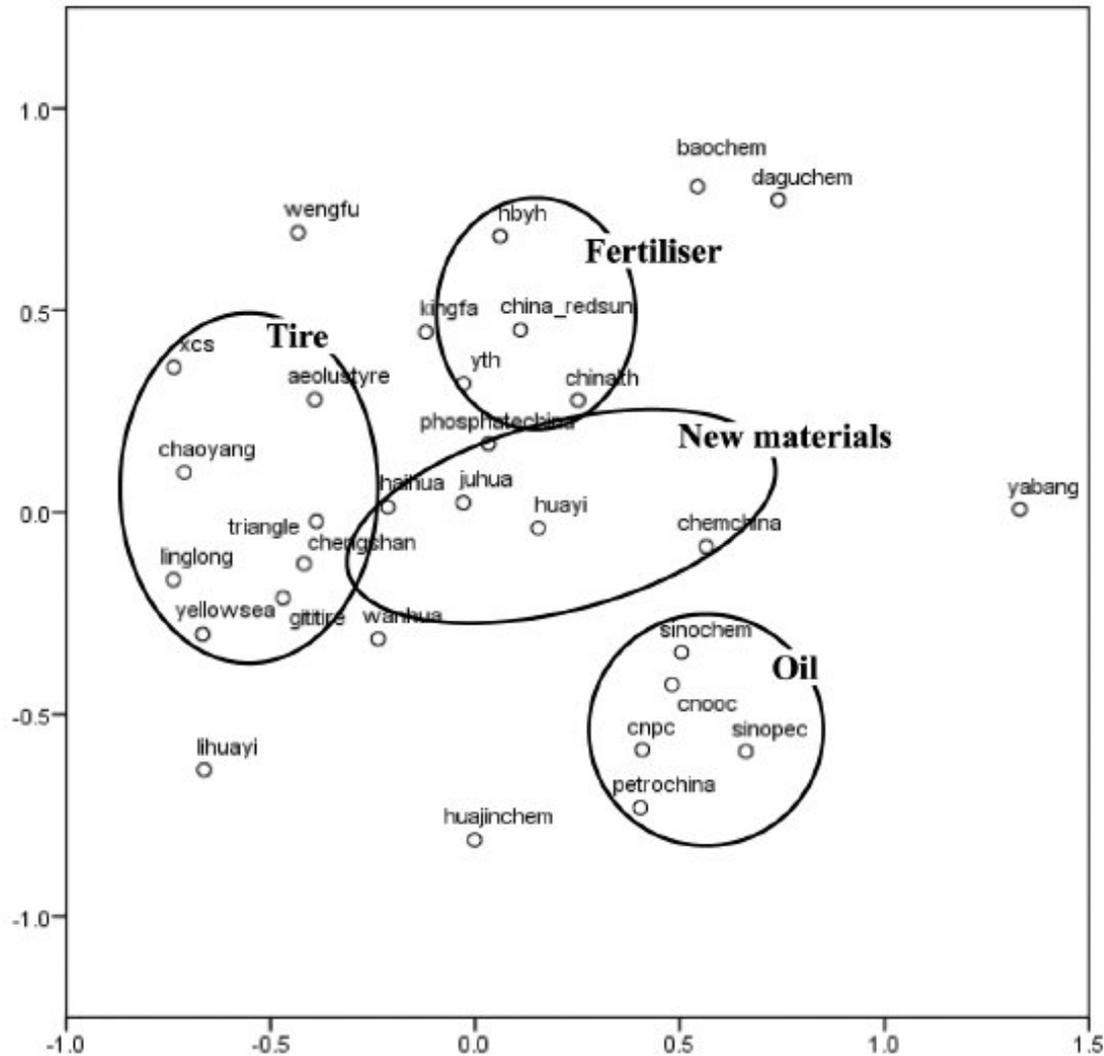
Mudanças no mercado são identificadas:

- Surge claramente o cluster de telecom.
- Aproximação de empresas de telas de LCD.
- Cluster do PC apresenta apenas empresas core

Para pesquisadores, o uso de dois buscadores não afetou os resultados, o que comprovaria a robustez do método.

2ª etapa: Mapa MDS - Química

stress value: 0,056



2ª etapa: Mapa MDS - Química

- Quatro seções distintas, sem grandes diferenças da primeira etapa.
- Destaca-se a entrada da ChemChina no cluster de materiais.

Conclusões da pesquisa na China

- Dados geraram mapas consistentes com as situações competitivas das indústrias pesquisadas.
- No setor de TI, as mudanças no mapa refletem as transformações do mercado entre as coletas de dados.
- Percebe-se que os sites de empresas chinesas são altamente orientados ao consumidor.
- Organizações que comercializam produtos mais "próximos" dos consumidores atraíram mais links do que aqueles voltados para empresas.

Conclusões gerais

- Os resultados indicam que o método de análise de co-links é bastante robusto.
- Na prática, a análise de co-links permite obter informações sobre o mercado.
- Vislumbra-se sua utilização para análise de indústrias na maioria dos países.
- No nível teórico, contribui para o conhecimento do fenômeno de link na web, útil para ciência da informação como um todo, não restrito ao mundo dos negócios.



Examining the robustness of web co-link analysis

Liwen Vaughan e Juan Tang and Jian Du

Anne Clinio e Rafaela Giodano