
Webometrics: What is it Good For?

Mike Thelwall

Statistical Cybermetrics Research Gp.
University of Wolverhampton, UK

Overview

- Common Webometrics Techniques
 - What they might be useful for
 - Problems and pitfalls
 - Case studies
 1. EU and European integration
 2. NetReAct: EU life science researcher mobility
 - The future for Webometrics?
 - In Scientometrics
 - In library and information science
 - In social science
-

Webometrics: *Contents*

- ❑ What is Webometrics?
- ❑ What Webometric techniques are there?
- ❑ What might Webometric techniques be useful for?
- ❑ Potential problems and pitfalls

webometrics.info - Isidro F. Aguillo, Internetlab



Webometrics Ranking of World Countries

Webometrics: Definition and Scope 1/2

- Definition

- “The study of quantitative aspects of the construction and use of Internet resources, structures and technologies on the Web drawing on bibliometric and informetric approaches” (Björneborn, 2004; PhD thesis www.db.dk/lb/)



- A bit like citation analysis, but with web data?

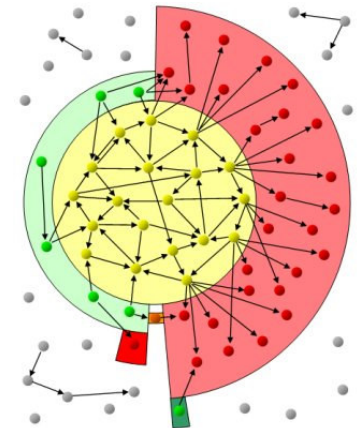
- Scope (1)

- Web page content analysis
- Web link structure analysis
- Web technology analysis (e.g. search engine performance)
- (Björneborn & Ingwersen, 2004: “Towards a framework for Webometrics” www.db.dk/lb/)



Webometrics: Definition and Scope 2/2

- Scope (2)
 - Webometrics research typically centres around
 - Gathering, interpreting or visualising *counts/the structure* of links or text (e.g. citations, author names) in sets of web pages, or
 - The performance of commercial search engines



The structure of the UK academic web (Björneborn, 2004)

Webometrics: Techniques 1/2

Commercial search engine searches can give lots of interesting free data via result count statistics

- Basic search engine result counts
 - E.g. Google “Informetrics” to count how many pages mention Informetrics or to list these pages (up to 1,000)
 - Advanced search engine queries such as link searches
 - E.g. MSN search using the link: or linkdomain: commands
 - E.G. linkdomain:hio.no -site:hio.no – counts/lists the number of pages outside HIO that link to HIO – that MSN knows about (up to 1000)
-

Webometrics: Techniques 2/2

- Commercial search engine Applications Programming Interfaces (APIs)
 - Programs automatically submitting 1,000-10,000 queries per day to Google, MSN, Yahoo!
 - Allows bigger research projects with less human intervention.
 - E.g. [LexiURL Searcher](http://lexiurl.wlv.ac.uk) at lexiurl.wlv.ac.uk
 - Web crawlers – download and extract data from all the pages in one or more web site
 - E.g. SocSciBot socscibot.wlv.ac.uk
-

Webometrics: Applications

- **Web citation analysis** – like ISI citation analysis but using citations to journal articles from web pages
 - Google scholar/Google phrase search for article title
 - Google link or text search for online article URL
 - Types of research questions (e.g. from Kayvan Kousha, Liwen Vaughan)
 - Do Google/Google Scholar citations correlate with ISI citations?
 - Are there Web/ISI disciplinary differences?



Webometrics: Applications

- **Link analysis/web structure analysis**

- How do links structure the web?
- What factors affect link creation? (e.g. geographic)



- **Information analysis**

- Longitudinal –changes over time in web content
- Accuracy/spread of information



- **Search engine stability**

- How do search engine results differ over time and between search engines

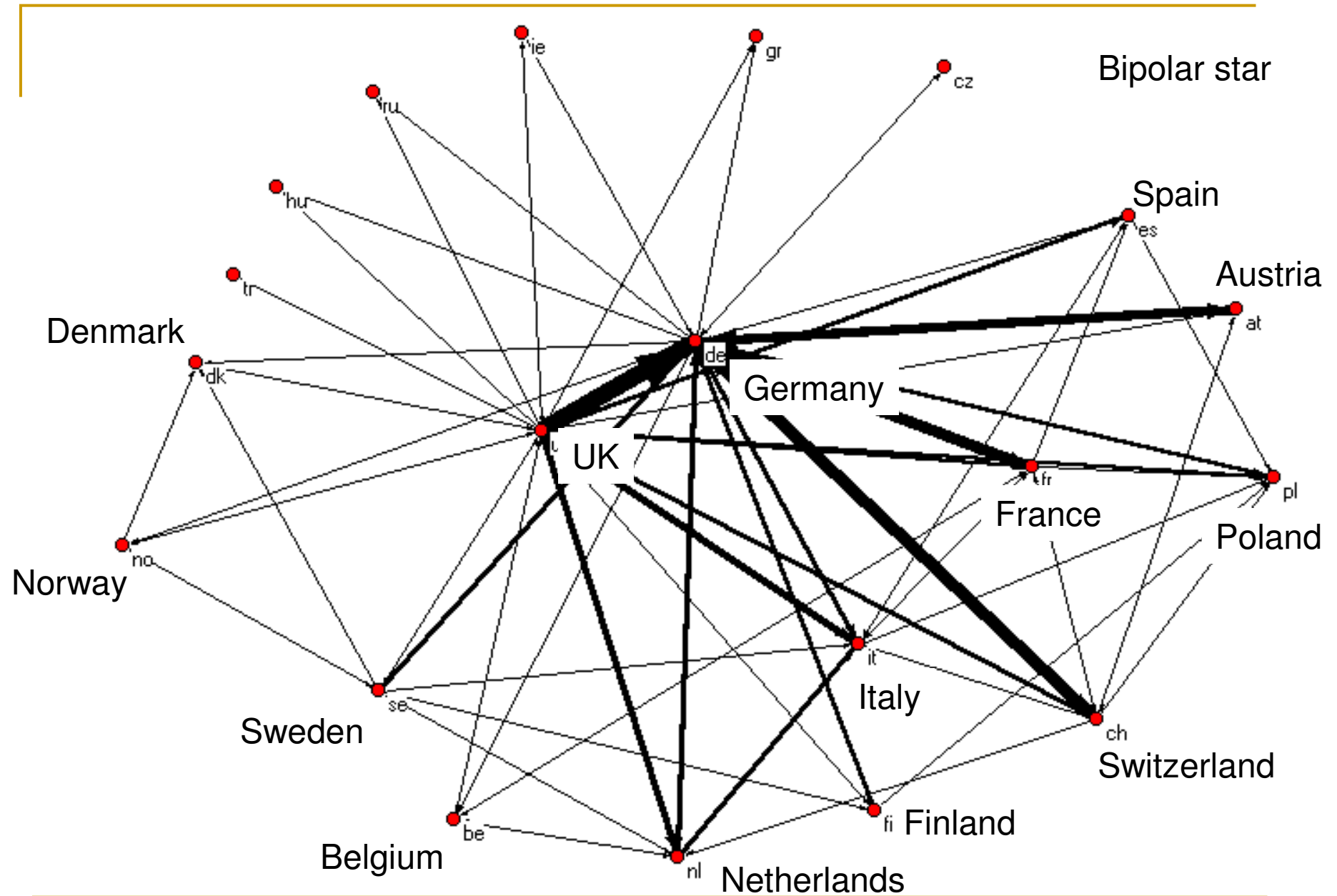


Webometrics: Problems and pitfalls

- Search engine problems
 - Partial and uneven coverage of the web
 - Sometimes unreliable result counts
 - Replication of information (e.g. mirror sites)
 - Theoretical problems
 - Web pages created for many different purposes
 - What do counts pages mean?
 - Much Webometrics is methodological –developing and assessing the validity of techniques
-

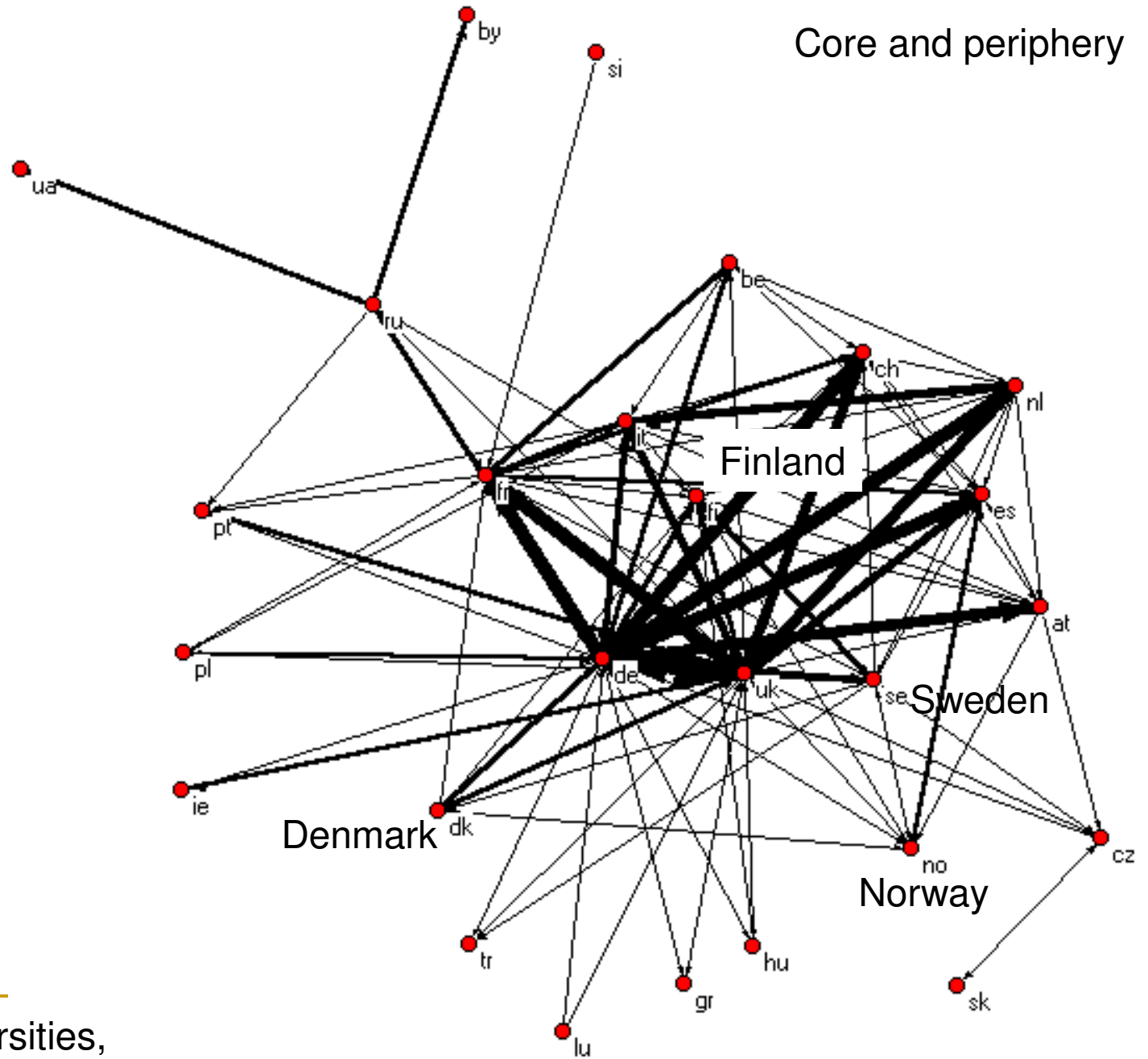
EU and European Integration

- Some findings about the extent of web interlinking within Europe
 - Data: links within Europe to and from European universities
 - Used a list of universities in 42 European countries and the MSN API to estimate the number of links to and from each university to and from each other country
-

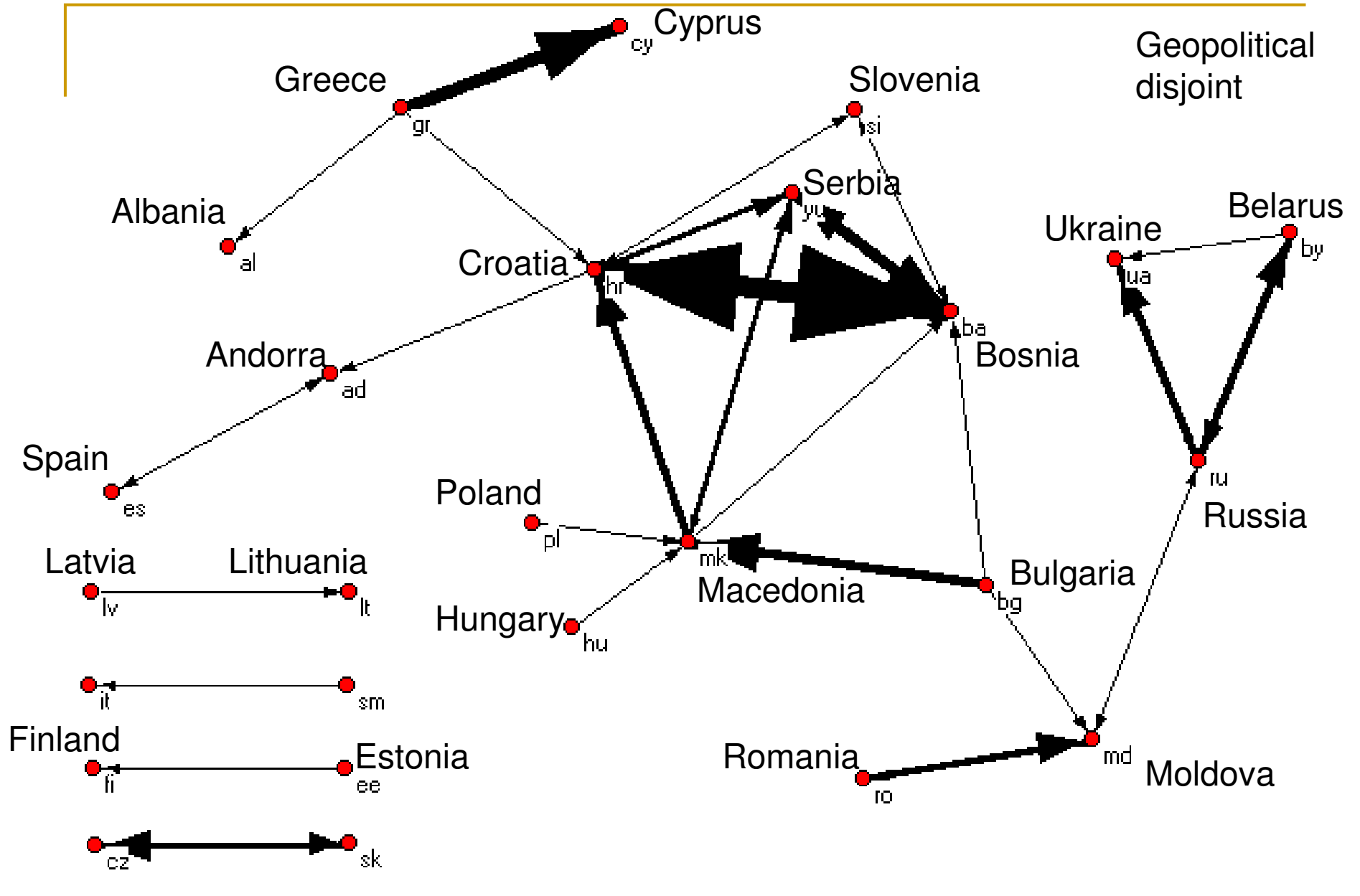


Links to universities,
Only top links and countries shown

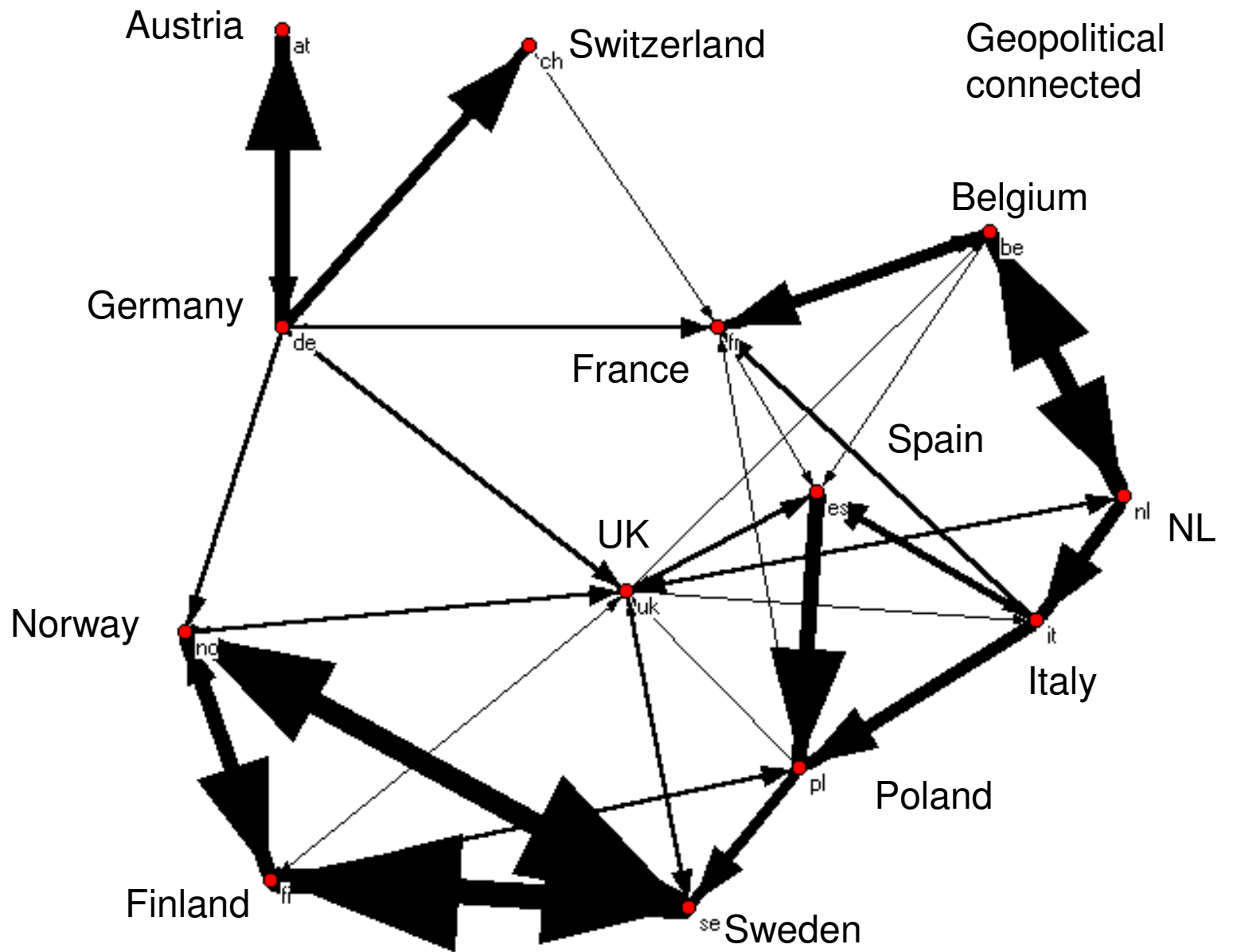
Core and periphery



Links *from* universities,
Only top links and countries shown



Normalised linking, only top links and linked countries shown



Normalised linking, smallest countries removed

NetReAct:

Webometrics/Scientometrics Case Study

- The role of networking in life science researcher mobility
- Investigation of a sample of European life science research groups
 - Interviews
 - Questionnaires
 - Bibliometrics
 - Webometrics

NetReAct 

The role of networking in research activities

Franz Barjak, Wolfgang Glanzel, Simon Robinson, Xuemei Li



empirica



NetReAct: Webometrics Objectives

- Objectives
 - Identify patterns of international communication via links
 - Assess the effectiveness of web site use
 - Identify national difference hypotheses for testing with the bibliometric data
-

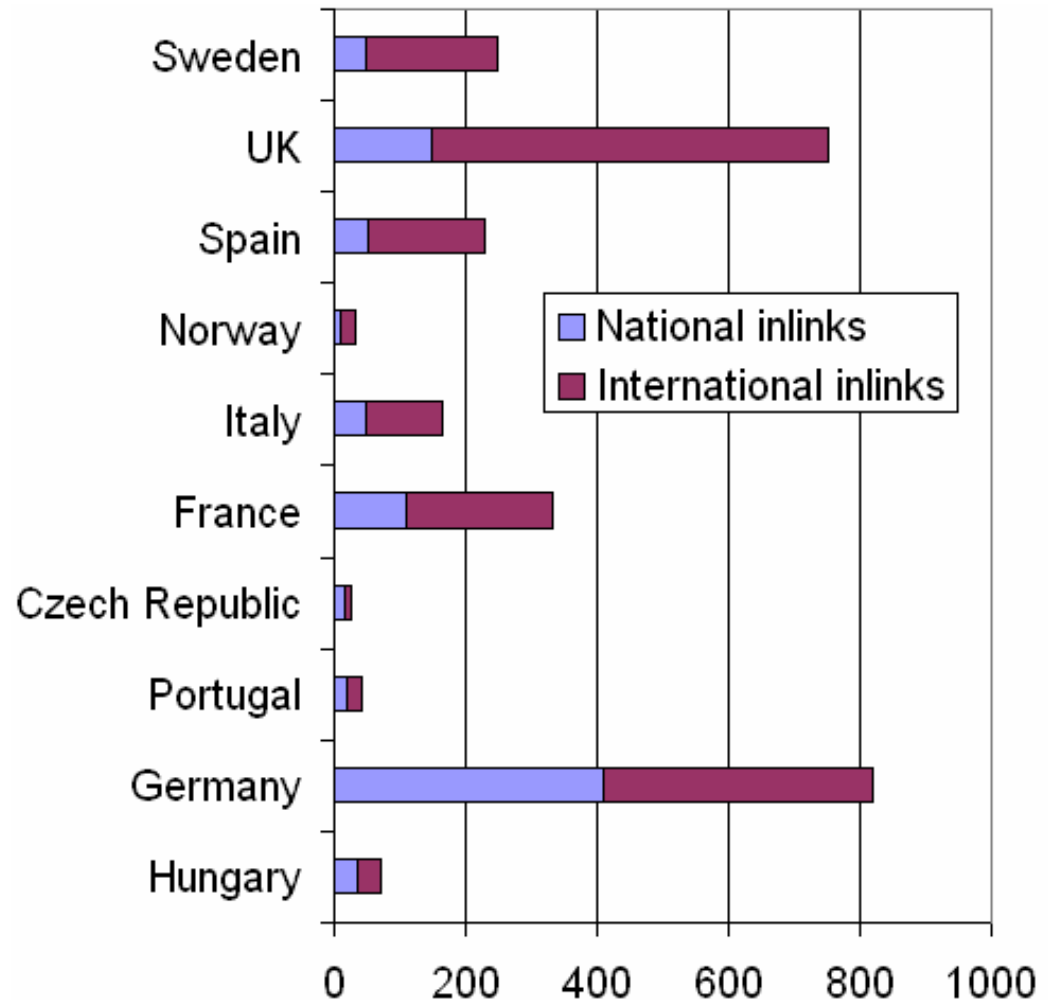
NetReAct: Webometrics Data

- Links *to* research group web sites (via LexiURL Searcher: Yahoo!/Microsoft/Google APIs)
 - Using the advanced search link: or linkdomain: commands
 - E.G. [linkdomain:hio.no -site:hio.no](#)
 - Links *from* research group home pages (via LexiURL)
 - The program LexiURL downloads the web pages and extracts the URLs
-

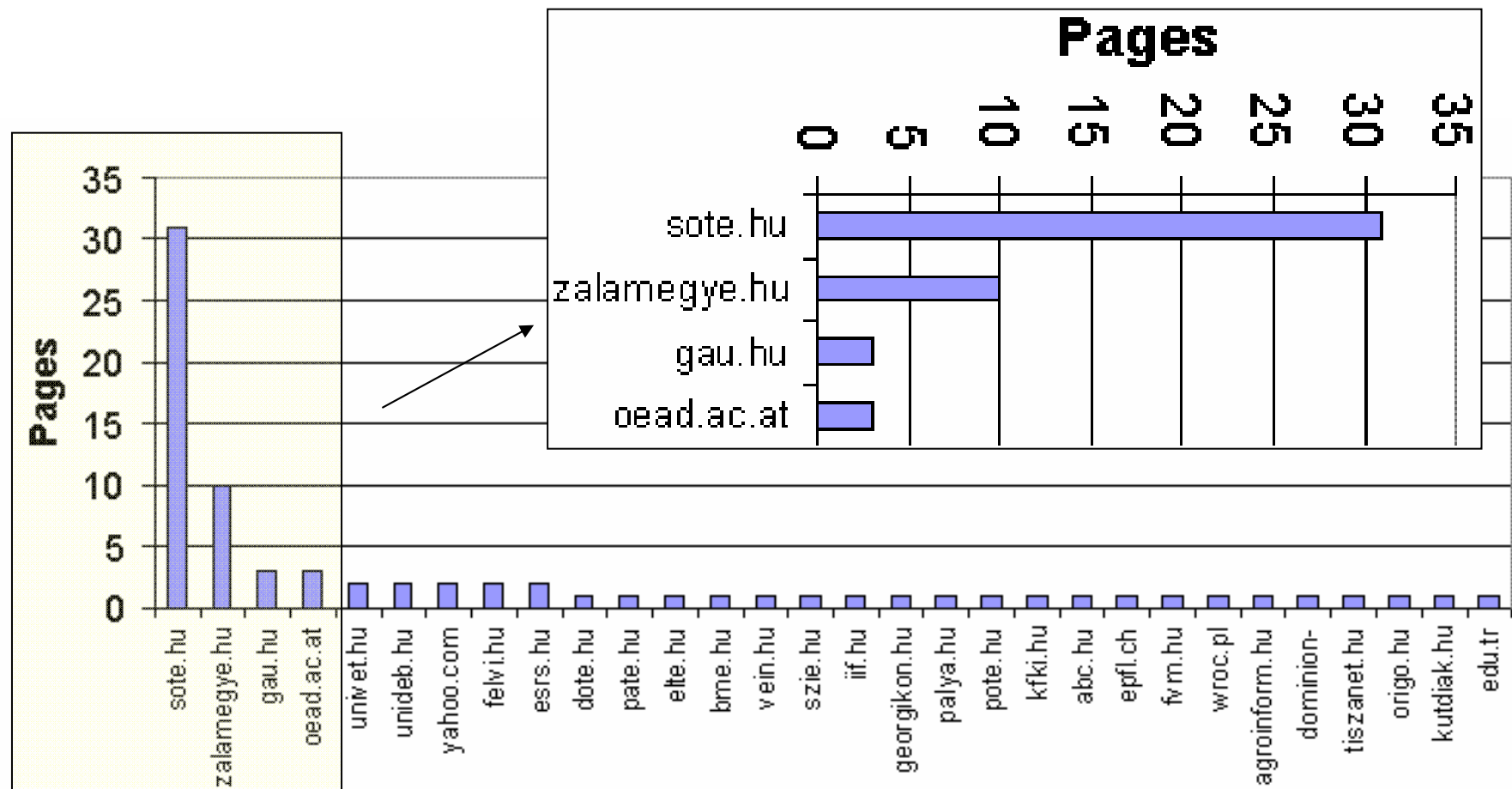
NetReAct: Webometrics Results 1/4

National and international inlinks to Life Science Research Groups

More international?
(doesn't take size into account)

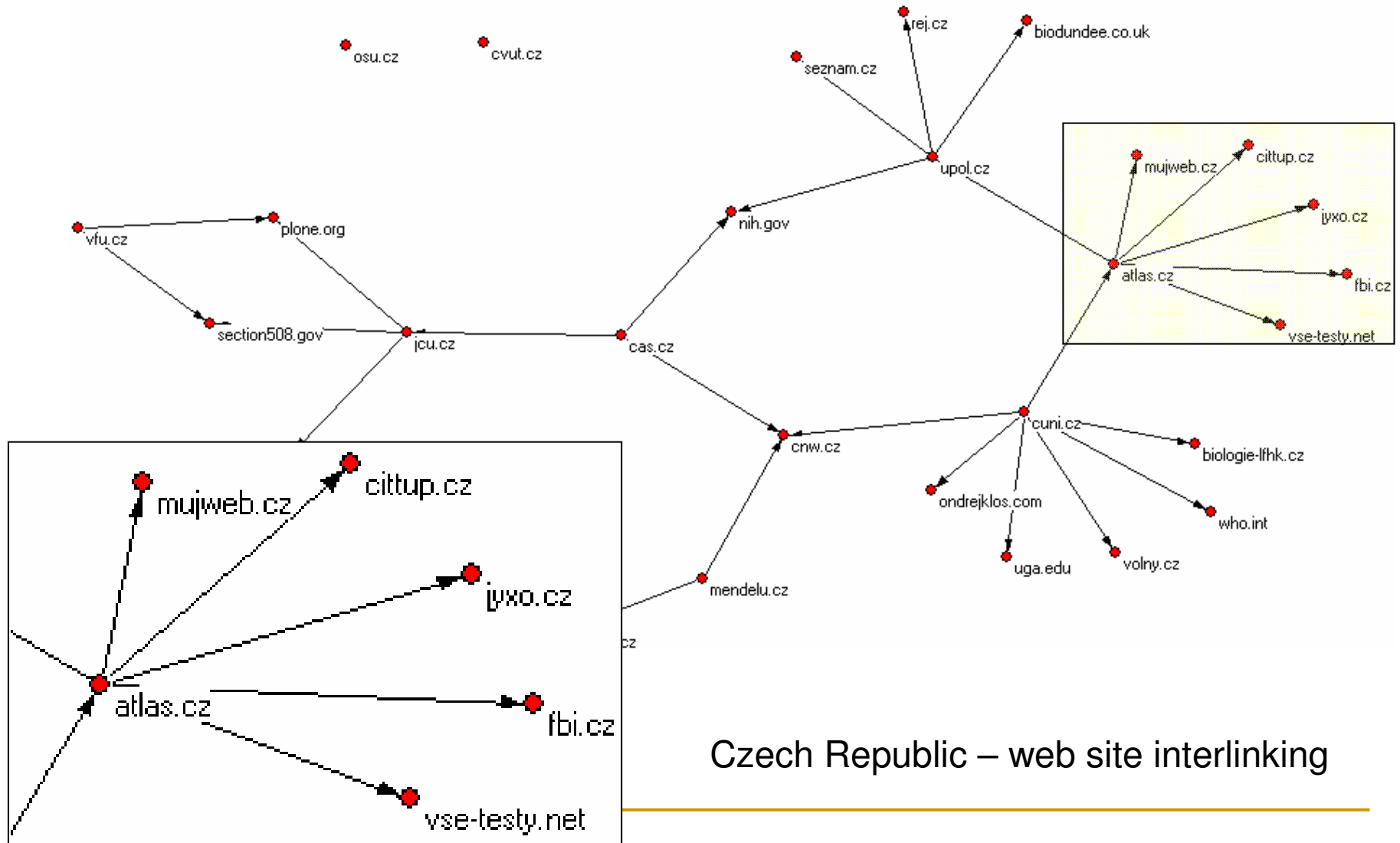


NetReAct: Webometrics Results 2/4



Link source site for pages linking to Hungarian LSRGs

NetReAct: Webometrics Results 3/4



NetReAct: Webometrics Results 4/4

Selected results:-

- General findings:
 - Language *not* a significant factor for linking
 - Smaller countries *not* more international
 - Problem countries:
 - Czech Republic, Norway, Portugal, Hungary almost ignored on the web (few inlinks)
 - Lack of an outward-looking focus (outlinks per LSRG): Portugal and Italy
 - Lack of an international focus: Hungary, Germany and France
 - Unusual features:
 - High government connections in Spain (e.g., mec.es, xunta.es)
 - UK research groups concerned with funding sources
-

The future for Webometrics?

- ❑ In Scientometrics
 - Fast pilot studies for bibliometrics/surveys
 - Measurement of web use/publishing
 - Investigation of wider scientific impact (e.g. industry, government, general population)
- ❑ In library and information science
 - Exploring how information has been constructed (e.g. small worlds, information in Wikipedia, social bookmarking sites)
 - Exploring links as indicators of information use?
- ❑ In social science
 - Exploring interconnections between organisations of interest (e.g. online activist groups)

