

# LIBRARY CATALOG ANALYSIS IS A USEFUL TOOL IN STUDIES OF SOCIAL SCIENCES AND HUMANITIES

**Daniel Torres-Salinas**

EC3. Univ de Navarra and Univ Granada

**Henk F. Moed**

CWTS. Leiden University



10th INTERNATIONAL CONFERENCE ON  
SCIENCE AND TECHNOLOGY INDICATORS  
VIENNA. SEPTEMBER. 2008

# 0. INTRODUCTION



In our research we focus on the evaluation of **books** in the **Social Sciences and Humanities** introducing an exploratory methodology:

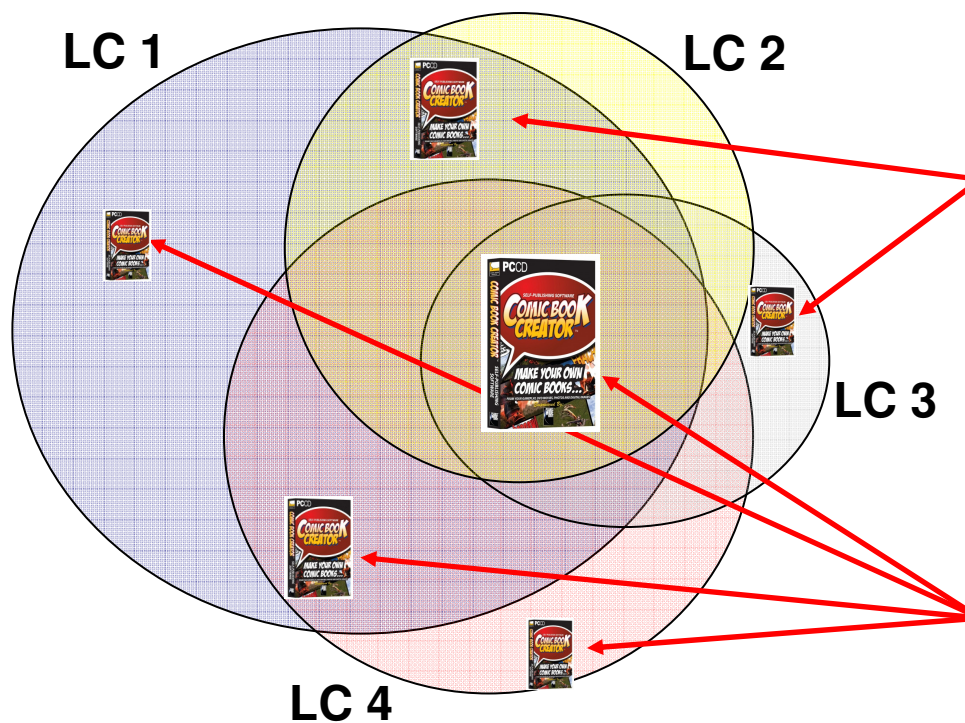
### **LIBRARY CATALOG ANALYSIS (LCA)**

**LCA** has two main **objectives**:



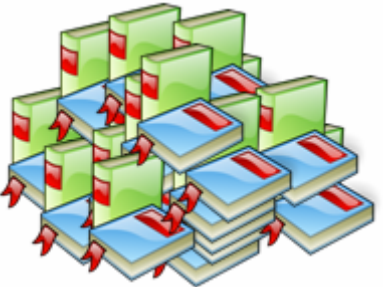

- Create a large database with books from worldwide academic libraries that allow us to describe a research field: **ECONOMICS**
- Use this database for evaluative purposes.
  - National Performance: **SPAIN**
  - **PUBLISHERS** performance
  - Institucional Performance: **University of Navarra**

# 0. INTRODUCTION

A representation of a research field by Online library catalogs.



A tool for the analysis of research performance.

OUTPUT	UNIT
	<b>Social Scientist</b> 
	<b>Department</b> 

## CITATION ANALYSIS vs LIBRARY CATALOG ANALYSIS

### MAIN CONCEPTS



## CITATION ANALYSIS vs LIBRARY CATALOG ANALYSIS

### INDICATORS

Create a set of Papers Published by unit

Measure the prestige of a journal

Numbers of citations received by a paper

Number received citations per paper (CIT/P)

Compare overall CIT/P average with a unit's CIT/P

Geographical spread of authors citing the unit's papers

...



Create a set of Books Published by unit

Measure the prestige of a Publisher

Numbers of catalog inclusions of a book

Numbers of catalog inclusions per book (CAT/P)

Compare overall CAT/P average with a unit's CAT/P

Geographical spread of catalogs containing the unit's book

...

“  
LCA can be defined as the  
application of bibliometric  
techniques to a set of library  
online catalogs”

## RESEARCH QUESTIONS

### **METHODOLOGICAL**

- How can we create a database of books?
- What indicators can we create?

### **PRACTICAL**

- Can we describe a scientific field with LCA?
- Can we use LCA for evaluative purposes?



# 1. METHODOGY



## A GENERAL VIEW OF THE PROCESS

- Selection of academic libraries



- Download books from Library Catalogs
- Compilation of books from UNAV



- Creation of an offline relational database



- Cleaning and normalization



- Query the database

## SELECTION OF ACADEMIC LIBRARIES

- **University's prestige:** Position in the *Essential Science Indicators* for Economics (WoK).
- **Catalog availability:** not all libraries offer public access for downloading bibliographical data (Z39.50)
- **Spanish emphasis:** the institutional evaluation is related to Spain and the *University of Navarra*.

## DOWNLOAD BOOKS FROM LIBRARY CATALOGS

Z39.50 ANSI / NISO Z39.50-1995 (ISO 23950)



Copac Academic & National Library Catalogue

 [Contact](#)

### Copac academic and national library catalogue

The Copac® library catalogue gives *free access* to the merged online catalogues of [major University and National Libraries in the UK and Ireland](#), including the British Library

**Search Copac** [start searching Copac](#)

Copac includes details of materials held in libraries throughout the UK, plus Trinity College Dublin Library in Ireland. Copac includes the catalogues of all the UK National Libraries, a wide range of major University libraries, as well as specialist collections such as the National Art Library (V&A Museum). [More about Copac...](#)

#### Information pages

- ▶ [About](#)
- ▶ [Libraries](#)
- ▶ [FAQ](#)
- ▶ [Support](#)
- ▶ [Interfaces](#)
- ▶ [News](#)

## DOWNLOAD BOOKS FROM LIBRARY CATALOGS

Z39.50 ANSI / NISO Z39.50-1995 (ISO 23950)

WHAT DO YOU NEED?

**QUERY STRATEGY:**

**Subject Headings (KEYWORDS)**

**ECONOM\***

- Economy
- Economical
- Economist

• ....

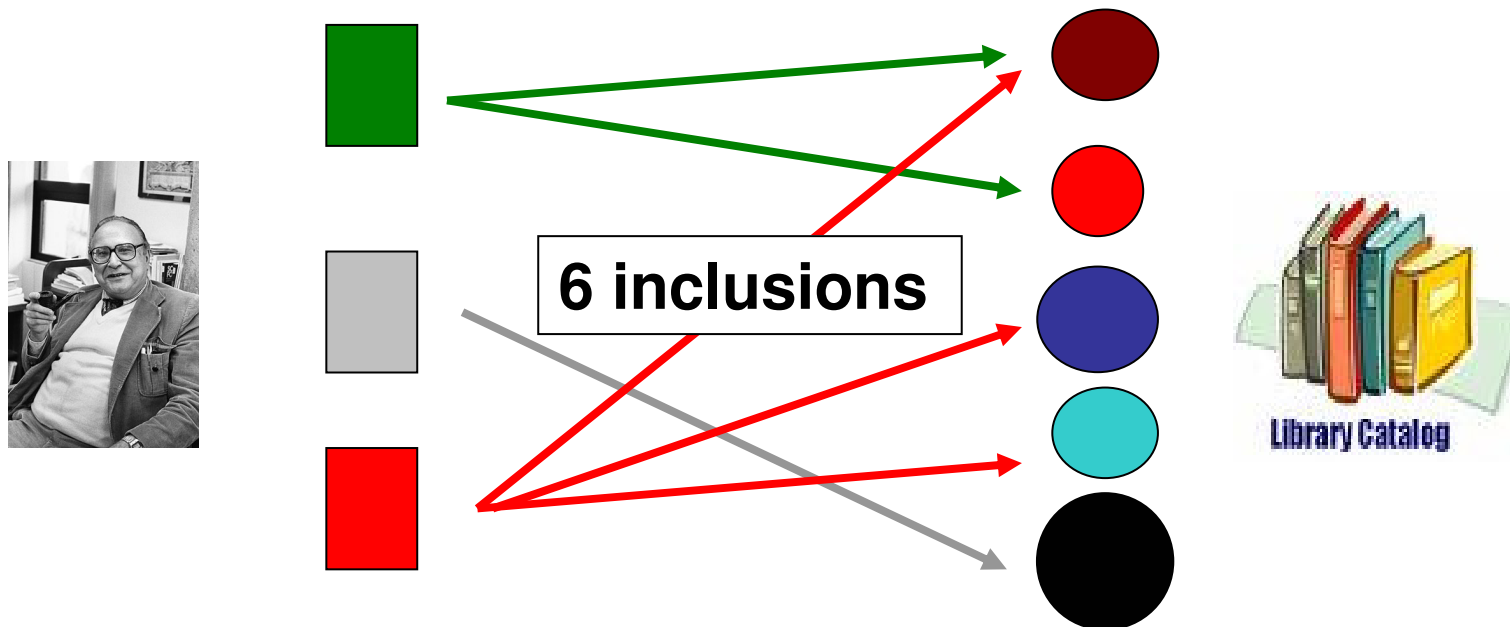
## INDICATORS

TWO MAIN INDICATORS: *INCLUSION* AND *DIFFUSION*

BASIC CONCEPTS FOR THE CONSTRUCTION OF THE INDICATORS

A Unit Publishing a set of **3** book titles

We use **5** library catalogs



The maximum number of possible inclusions for the set is 15:  $3 \text{ titles} \times 5 \text{ catalogs}$

**INDICATORS****NUMBER OF TITLES**

**The total number of unique book titles for a particular aggregate**

**CATALOGS INCLUSIONS (CI)**

**Total number of inclusions**

**CATALOG INCLUSION RATE (CIR)**

**The total number of catalog inclusions divided by the number of titles in the set**

**RELATIVE CATALOG RATE (RCIR)**

**It is defined as the ratio of CIR of the aggregate to be assessed and CIR of the aggregate that serves as a benchmark in the assessment**

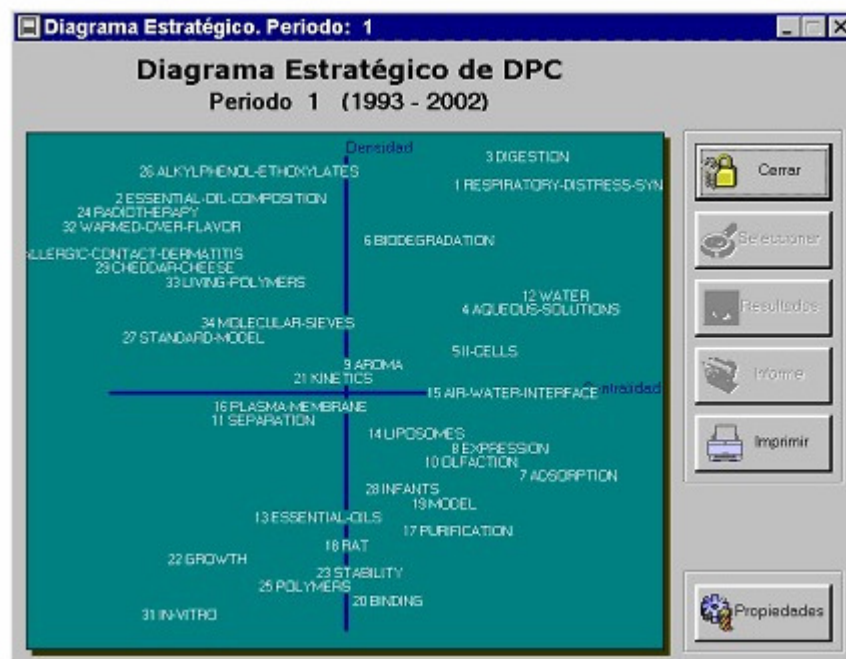
**DIFFUSSION RATE (DR)**

**The percentage of catalog inclusions of book titles produced by a given aggregate relative to the total number of possible catalog inclusions**

## OTHER INDICATORS: CO-WORD ANALYSIS



CALLON AND COURTIAL





## 2. RESULTS

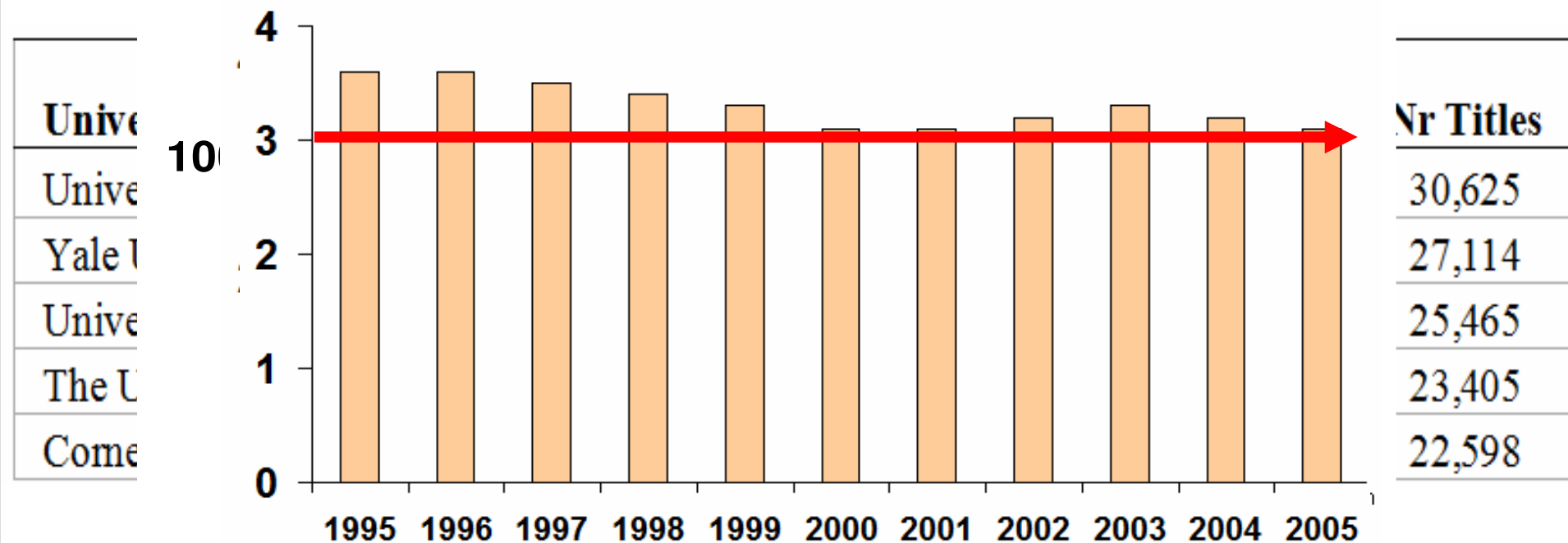


## 2. RESULTS

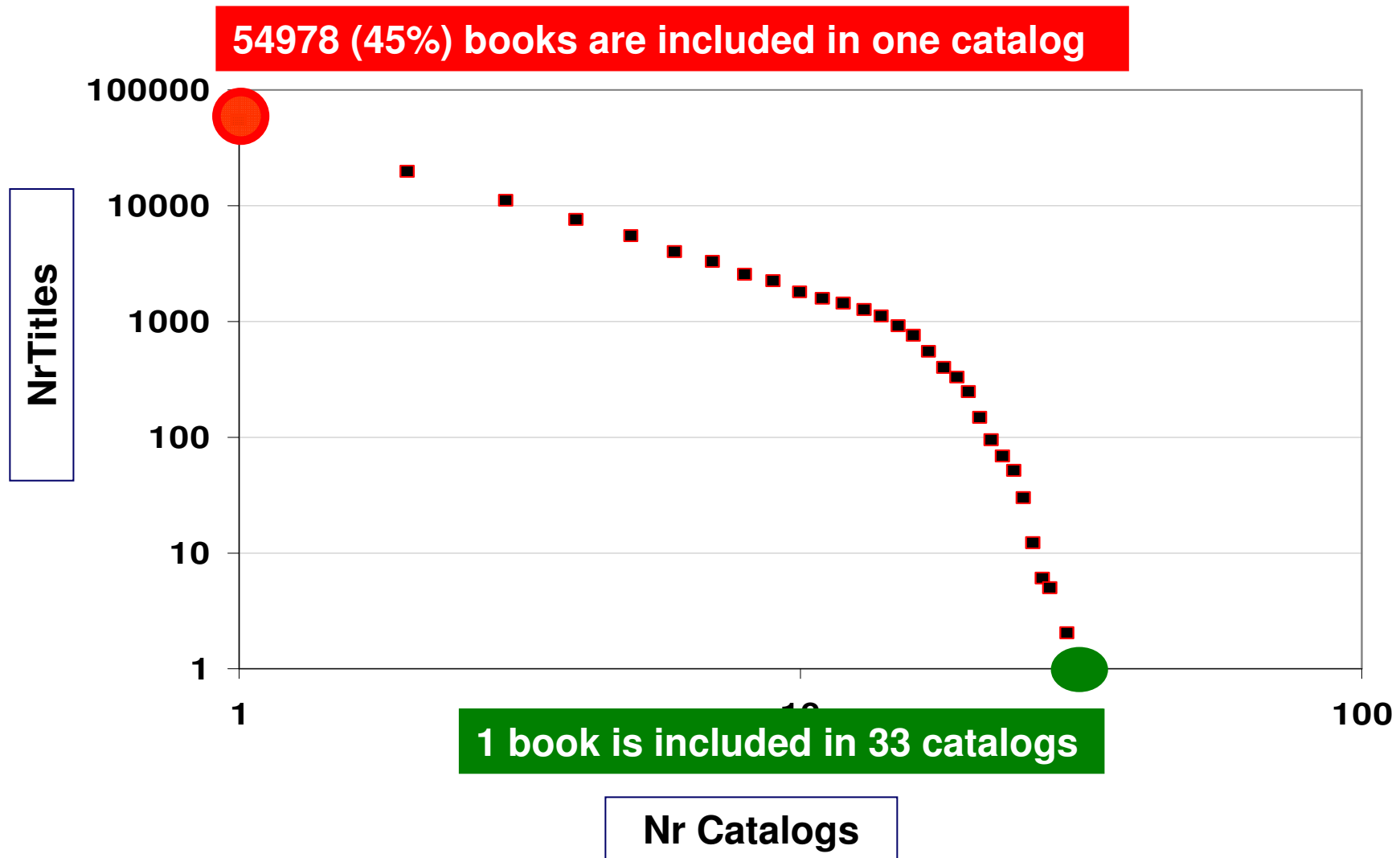
### GENERAL DESCRIPTION OF THE DATABASE

#### Overall Statistics

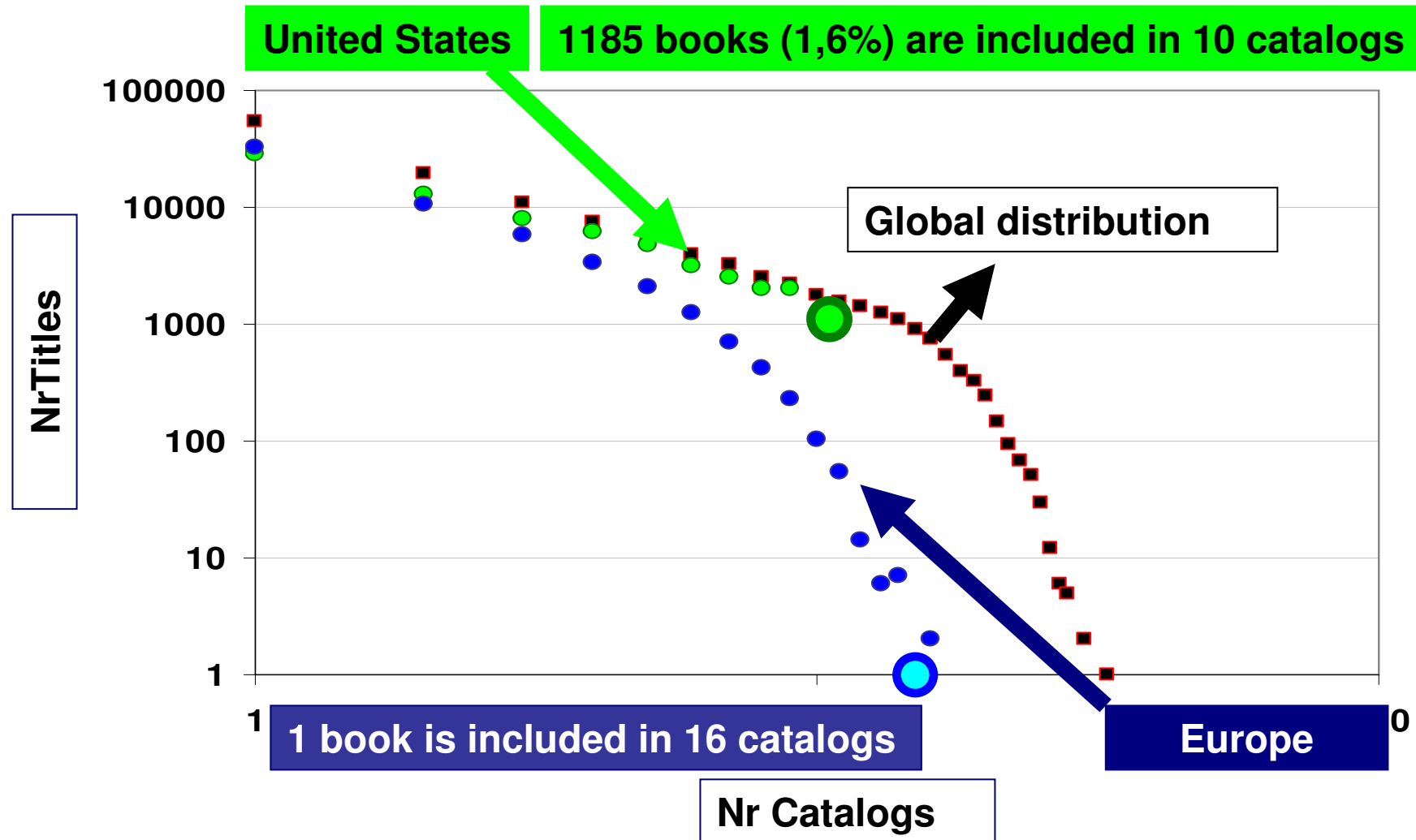
Number of countries	7
Number of host Universities	42
Number of unique book titles	121,147
Number of catalog inclusions	417,033
Overall Catalog Inclusion Rate	3.4



## DISTRIBUTION OF BOOK TITLES AMONGS CATALOGS



## DISTRIBUTION OF BOOK TITLES AMONGS CATALOGS



## 2. RESULTS

### Publication languages of book titles

All 42 libraries			% Titles in the country where libraries are located						
Language	Nr Titles	% Titles	Australia	Canada	Ireland	Spain	Sweden	UK	USA
English	75,211	61	93	86.5	92.2	38.1	86.5	89	67.8
Spanish	13,348	10.8	0.9	1.6	2.8	52.4	2.4	1.7	7.2
French	6,209	5.0	0.9	3.8	2.8	4.0	0.7	2.6	3.5
Chinese	5,858	4.8	0.7	3.4	0.0	0.0	0.0	0.1	4.5
German	4579	3.7	0.3	1.0	0.5	0.9	0.4	1.5	3.9
Russian	3104	2.5	0.1	0.4	0.0	0.0	0.1	1.4	2.7
Italian	2344	1.9	0.1	0.3	0.3	1.1	0.0	0.9	1.7
Japanese	2191	1.8	0.8	1.0	0.1	0.0	0.1	0.0	1.7
Indonesian	1167	1.0	1.2	0.8	0.0	0.0	0.0	0.0	1.3
Portuguese	1142	0.9	0.0	0.0	0.0	0.3	0.1	0.2	1.0

**English and Spanish account for 72% of the book titles**

**In 5 out of the seven countries the English is the main language**

**United States libraries has miscellaneous collections**

**Spain has a distinct profile with a strong presence of books in english**

## AUTHORS: DISTRIBUTION AND AVERAGE

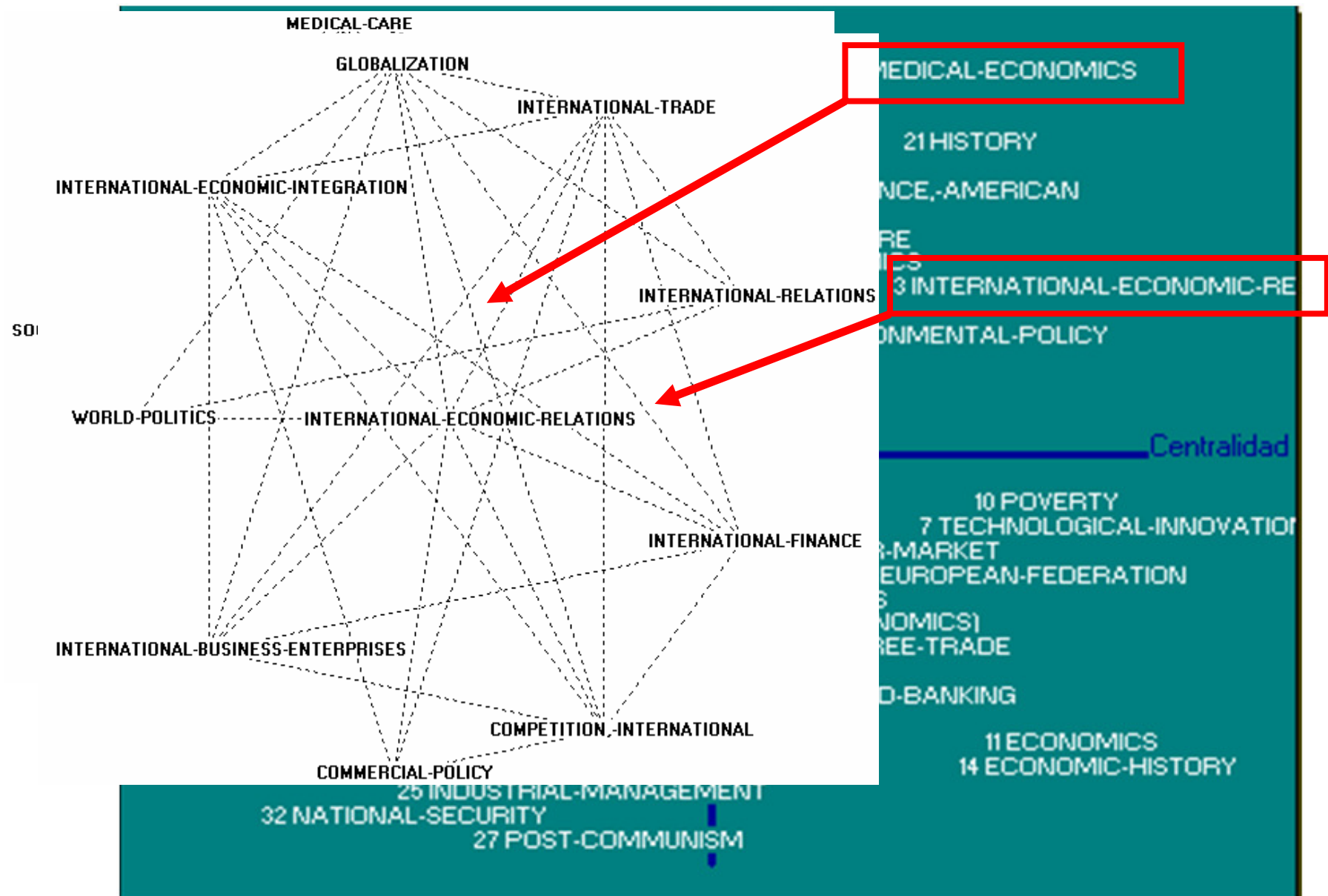
**Avg  
authors  
1.63**

<b>Nr Authors</b>	<b>Nr Titles</b>	<b>% Titles</b>
1	74,952	62
2	24,853	21
3	15,424	13
4	4,208	3.5
5	1,013	0.84
6	507	0.42
7	88	0.07
8	62	0.05
9	16	0.01
10	18	0.01
11	6	0.00
12	7	0.01
13	1	0.00
14	1	0.00

**96% of the titles  
1,2,3 authors**

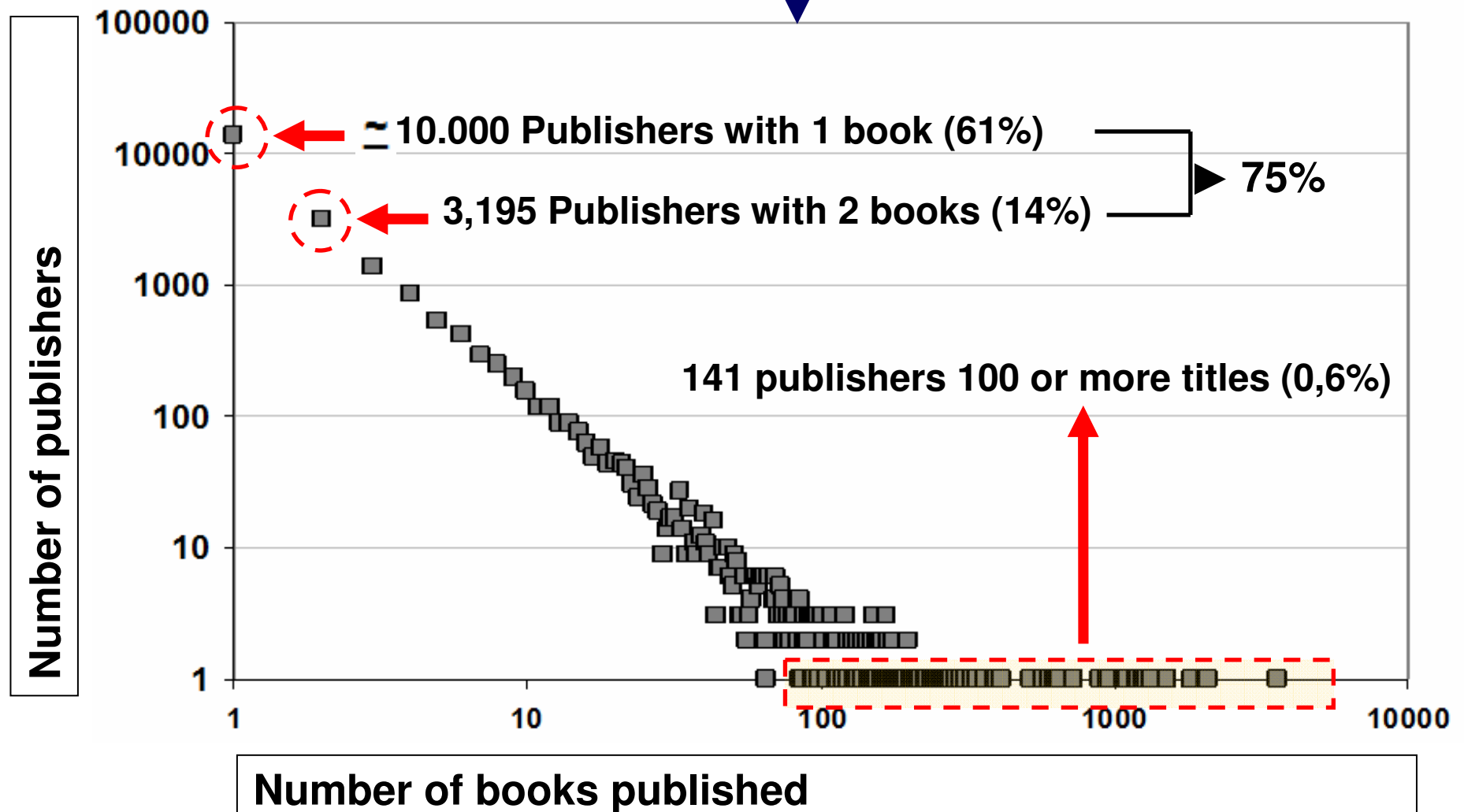
**The others have  
less than 1% of  
titles**

## THEMATIC MAPPING OF THE DATABASE USING SUBJECT HEADINGS



## Distribution of number of book titles amongst publishers

There are  $\approx 22.000$  publishers in the database





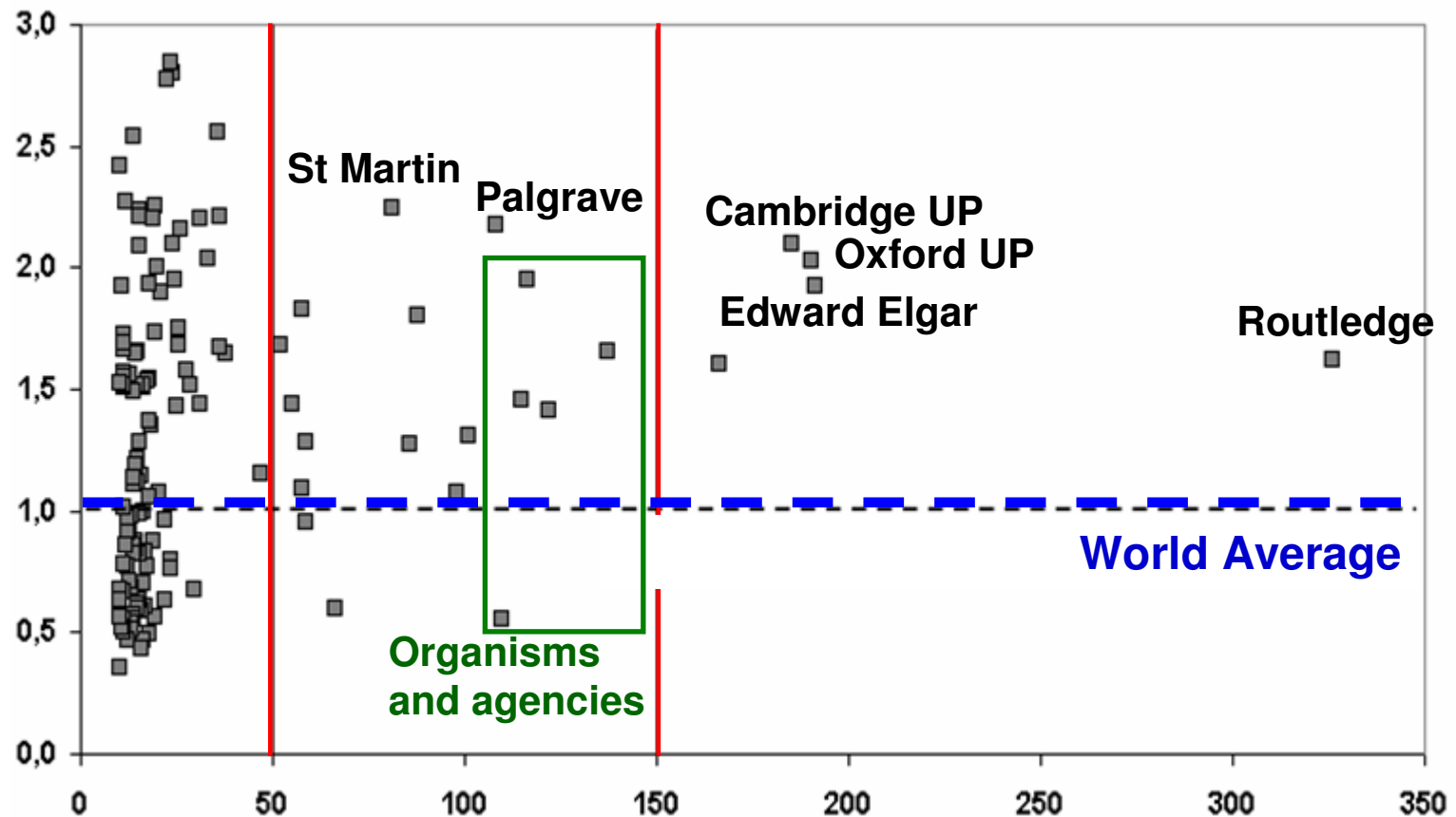
## 2. RESULTADOS

### The 25 most productive publishers in Economics

	Publisher name	Total Nr Titles	Nr Titles per year	Nr Catalog Inclusions	Catalog Inclusion Rate	Relative Catalog Inclusion Rate
→	Routledge	3,589	326	19,985	5.6	1.6
→	Edwad Elgar	2,102	191	13,922	6.6	1.9
→	Oxford University Press	2,092	190	14,576	7.0	2.0
→	Cambridge University Press	2,040	185	14,710	7.2	2.1
	Macmillan Press	1,827	166	10,054	5.5	1.6
	World Bank	<b>Publishers from organizations and agencies</b>				
	Official Publications					
	United States GPO	1,281	116	8,613	6.7	2.0
	OECD	1,260	115	6,308	5.0	1.5
	Her Majesty's Stationery Office	1,209	110	2,315	2	1.9
→	Palgrave	1,188	108	8,905	7.5	2.2
	United Nations	1,109	101	4,993	4.5	1.3
	Springer	1,078	98	3,995	3.7	1.1
	Ashgate	966	88	5,992	6.2	1.8
	Kluwer	944	86	4,133	1	4.4
→	St Martin's Press	894	81	6,904	7.7	2.2
	Harmatan	727	66	1,506	2.1	0.6

## *Number of titles published per year and Relative Catalog Inclusion Rate (RCIR) for major publishers*

Relative Catalog Inclusion Rate



Number of titles per year

## Spanish performance

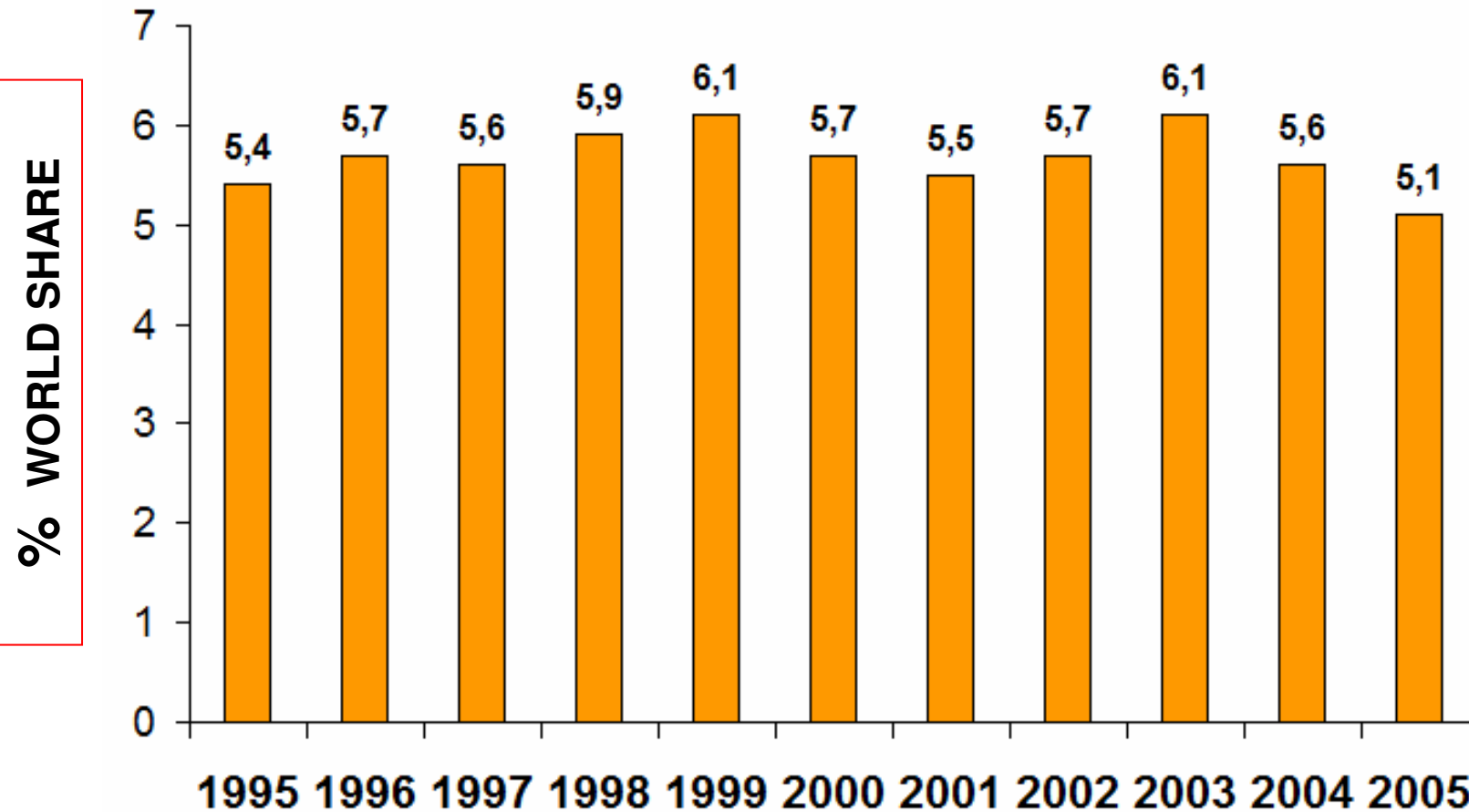
	Nr Titles	% Titles	Nr Catalog Inclusions	Catalog Inclusion Rate	Relative Catalog Inclusion Rate	Diffusion Rate
World	6,704	100	14932	2.2	0.6	0.05
United States	740	11.0	1,494	2.0	0.6	0.02
Spain	6,704	100	13,016	1.9	1.2	0.16
United kingdom	197	2.9	225	1.1	0.5	0.00
North America	763	11.3	1,603	2.1	0.6	0.01
European countries	6,704	100	13,303	1.9	0.9	0.09

### SPANISH PERFORMANCE

### SPANISH BOOKS INCLUDED IN DIFFERENT CATALOGS PER COUNTRY

### SPANISH BOOKS INCLUDED IN DIFFERENT CATALOGS PER WORLD REGION

### Spanish performance per year



YEAR OF PUBLICATION

## Institutional performance: The case of University of Navarra

### UNIVERSITY OF NAVARRA PERFORMANCE

(UNNA) have published during 1995-2005 a total of 211

### BOOKS INCLUDED IN DIFFERENT CATALOGS PER COUNTRY

works, 10 per cent of those were monographs or edited

### BOOKS INCLUDED IN DIFFERENT CATALOGS PER WORLD REGION

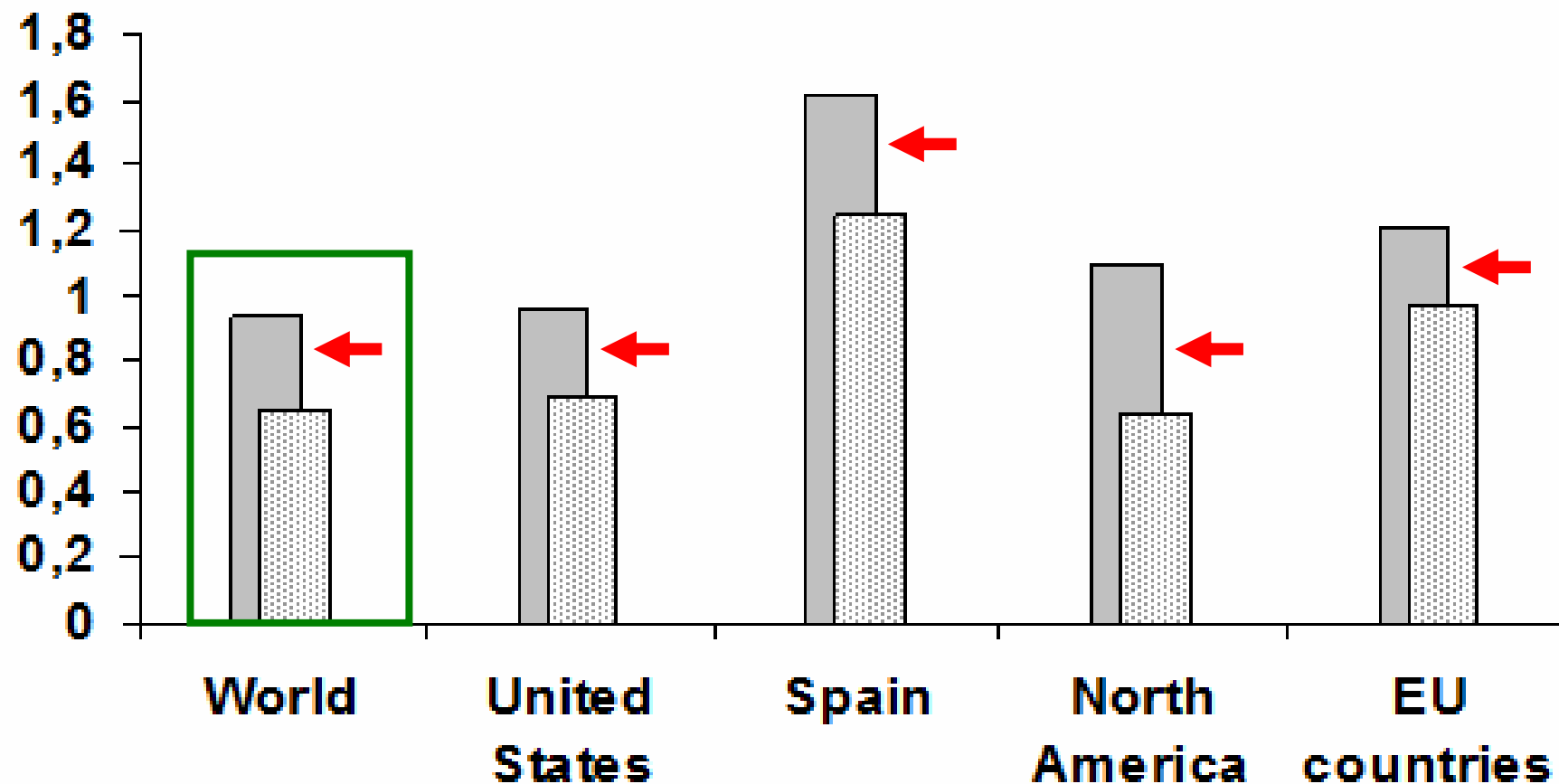
publications carried a Spanish ISBN

	Nr Titles	% Titles	Nr Catalog Inclusions	Catalog Inclusion Rate	Relative Catalog Inclusion Rate	Diffusion Rate
Entire database	44	100	142	3.2	0.9	0.08
United States	11	25	31	2.8	0.9	0.07
Spain	29	66	72	2.5	1.6	0,13
United Kingdom	12	26	20	1.7	0.9	0.04
North America	12	27	43	3.6	1.1	0,07
European countries	40	91	98	2.4	1.2	0.11

## Comparing Spain and the University of Navarre



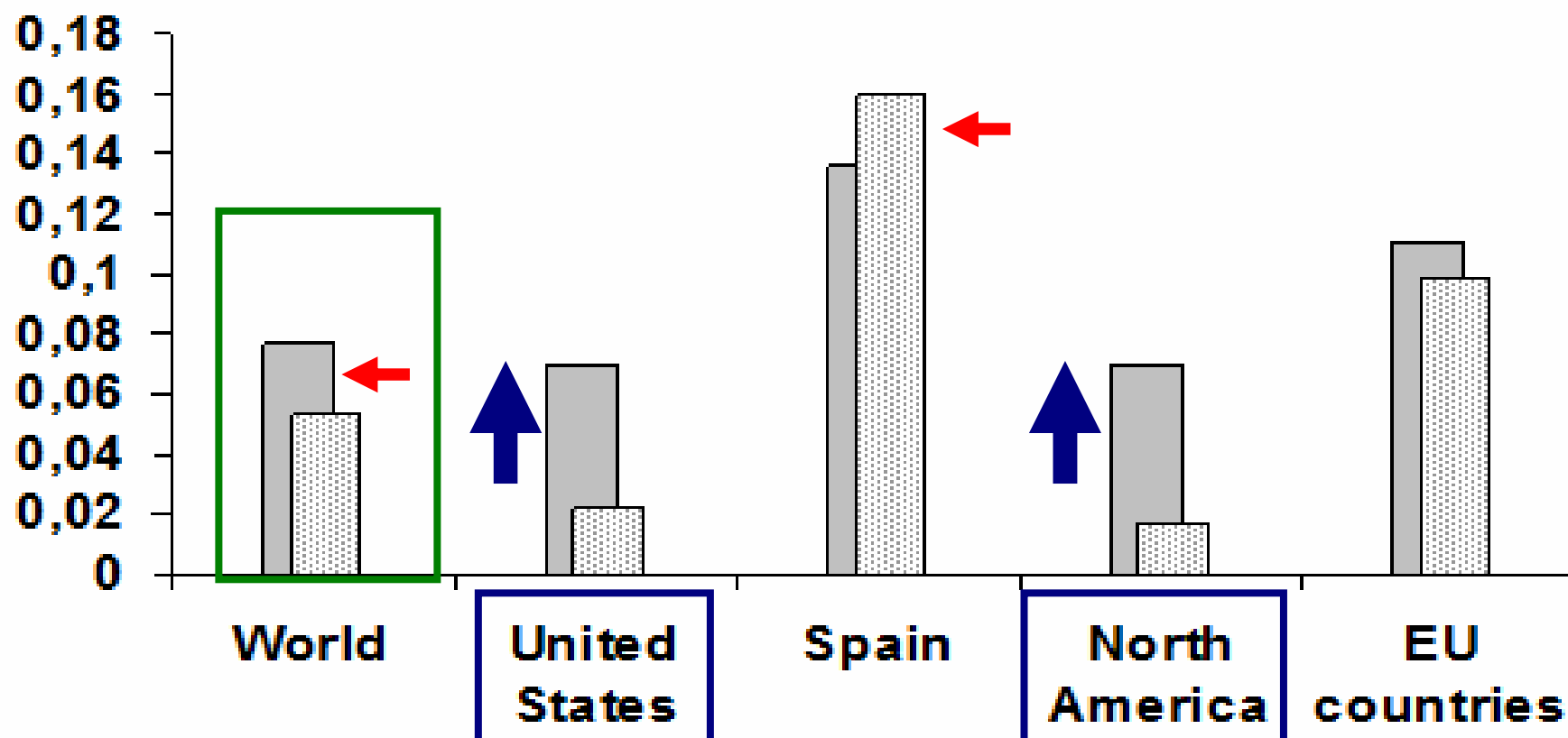
## Relative Inclusion Rate



## Comparing Spain and the University of Navarre

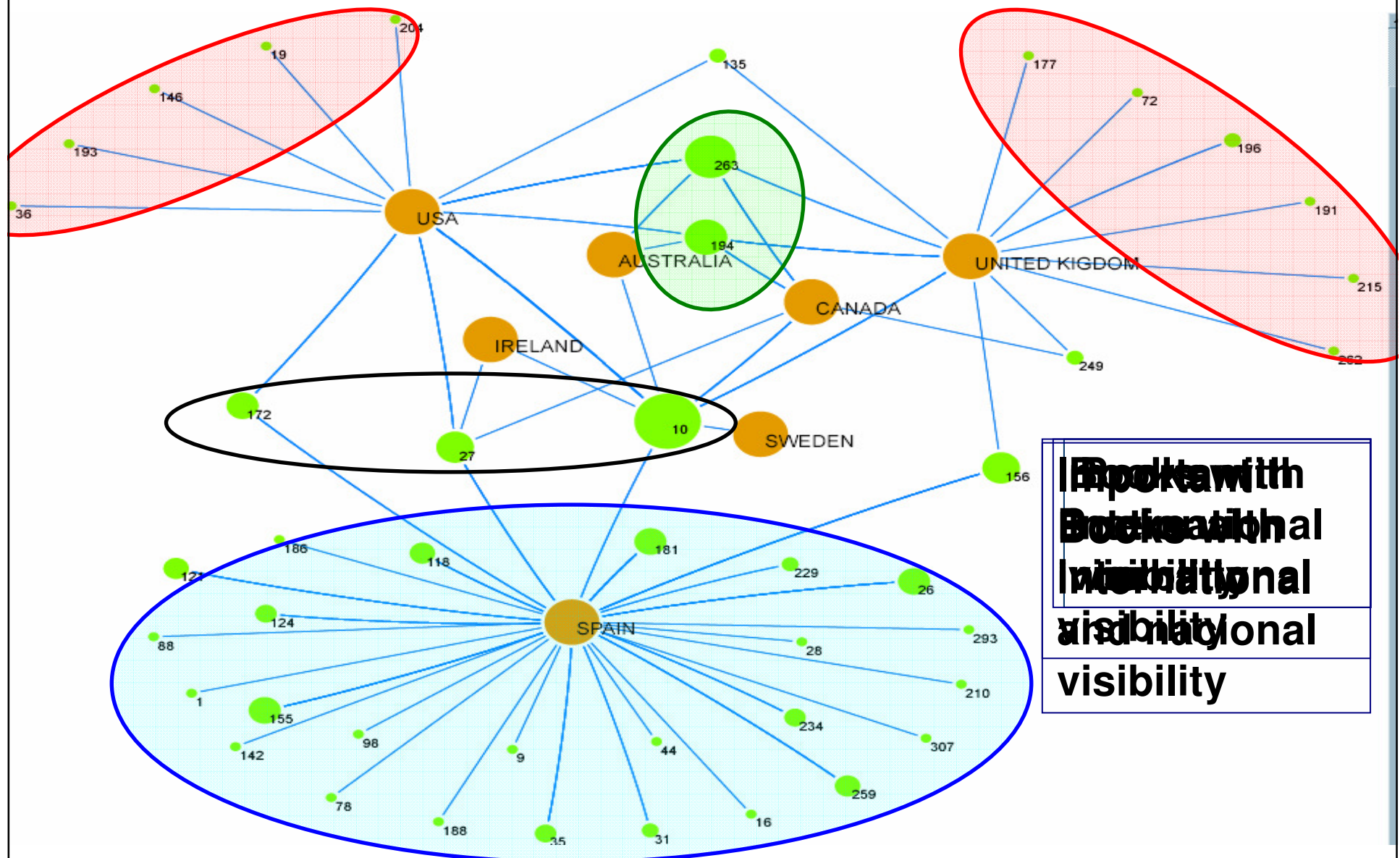


## Diffusion Rate



## 2. RESULTS

### BOOKS FROM THE UNIVERSITY OF NAVARRE: GEOGRAPHICAL SPREAD





# 3. CONCLUSIONS



### 3. CONCLUSIONS

#### TWO MAIN CONCLUSIONS

- 1. The proposed analogy model between citation analysis and library catalog analysis has proven to be valuable**

**1. NATIONAL BIASES.** The set of libraries selected in the current study has a rather strong Anglo-Saxon *bias*

- affect the identification the national publishers
- The coverage of Z39.50 protocol

**2. FIELD DELIMITATION PROBLEMS.** The current, exploratory study used the truncated string *Econom\** .

- This problem is particularly relevant in the assessment of research performance of institutions and affect the identification the national publishers
- Differences in the subject classification systems among libraries

**3. TYPE OF BOOKS.** Library catalogs not apply a categorization of books into types

- Useful types would be: authored versus edited words; books primarily for teaching versus those primarily for research; books for a specialized scientific-scholarly audience versus those for broader audiences within or even outside the scientific-scholarly community

***THANK YOU FOR YOUR  
ATTENTION***

**LIBRARY CATALOG ANALYSIS IS A USEFUL TOOL  
IN STUDIES OF SOCIAL SCIENCES AND  
HUMANITIES**

**Daniel Torres-Salinas** – EC3. Universidad de Navarra  
**Henk F. Moed** – CWTS. Leiden University