



# ResearchGate (RG) for Author Name Disambiguation (AND)

Mehmet Ali Abdulhayoglu &  
Bart Thijs

# Outline

- Introduction and previous endeavors
- How to scrape RG and access the data
- Comparisons with original data & Results
- Discussion and conclusion

# Introduction

- Importance on (individual) research evaluation
- Interior data indexed in DBs such as
  - Author names
  - Journal names
  - Publication titles
- Thomson Reuters WoS also indexes
  - E-mail address (13.8%)
  - Researcher IDs (8.6%)

# Introduction

- A few successful attempts leveraging external data  
(Kanani et al., 2007; Yang et al., 2008, Kang et al., 2009; Pereira et al., 2009)


- Google's CSE API



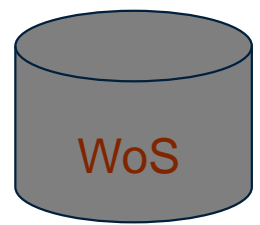
Very limited number of queries to issue

Each web site means different web design

# ResearchGate (RG)

- Social network for scholars (publication titles, affiliations)
- Some bibliometric studies based on RG (Thelwall & Kousha (2015); Ortega (2015))...but not AND related
- RG as a tool to confirm clustering results
- Connected components (CC) based document network
  - Author names, E-mail addresses, Affiliation, Co-authors
  - Weak CC partitioning  is greedy risk for false positives
- RG & CC joint results are only retained

# Crawling RG



N-GRAM SIM > 0,90

IF YES

COMPARE AUTHOR NAMES

TITLE

TOP RESULT

http://www.researchgate.net/publicliterature.PublicLiterature.search.html?type=keyword&search-keyword= + *A triangular model for publication and citation statistics of individual authors* + "&search-abstract=&search-journal=&search=S search"

Publication search

Advanced search is a quick and easy way to find and discover research, read abstracts in your field, or find a journal where you can publish your own work.

Keyword Search

Similar Abstracts

A triangular model for publication and citation statistics of individual authors

Search

28 publication matches found:

Article: **A triangular model for publication and citation statistics of individual authors**  
Wolfgang Glänzel - Sarah Heeffer - Bart Thijs  
[Show abstract]  
No preview · Article · Feb 2016 · Scientometrics

Article

**A triangular model for publication and citation statistics of individual authors**

Wolfgang Glänzel Sarah Heeffer Bart Thijs

Scienometrics (Impact Factor: 2.18). 02/2016; DOI: 10.1007/s11192-016-1870-0

ABSTRACT

One of the most important requirements of building applicable models and meaningful indicators for the use of scientometrics at the micro and meso level is the correct identification

One of the co-author profile link: [https://www.researchgate.net/profile/Wolfgang\\_Glaenzel](https://www.researchgate.net/profile/Wolfgang_Glaenzel)

Wolfgang Glänzel KU Leuven, Leuven  
Quantitative Social Research, Probability Theory  
PhD  
41.57

OVERVIEW CONTRIBUTIONS INFO STATS

<b>260</b> PUBLICATIONS	10.45k Reads	7.851 Citations	408.77 Impact Points	<a href="#">View stats</a>
----------------------------	-----------------	--------------------	-------------------------	----------------------------

University of Leuven  
Department of Managerial Economics, Strategy and ...  
Leuven, Belgium

10,940 publications 31,983 authors by WoS

# Clustering

<i>WoS Name</i>	<i>WoS Last Name</i>	<i>RG Name</i>	<i>Metaphone</i>	<i>Regular Expression</i>	<i>3-grams Score</i>
w glanzel	glanzel	wolfgang glänzel	wolfgang glanzel	glanzel	<b>0.51</b>
w glanzel	glanzel	sarah heeffe	sarah heeffe	-	0.13
w glanzel	glanzel	bart thijs	bart thijs	-	0.00
s heeffe	heeffe	wolfgang glänzel	wolfgang glanzel	-	0.10
<b>s heeffe</b>	<b>heeffe</b>	<b>sarah heeffe</b>	<b>sarah heeffe</b>	<b>heeffe</b>	<b>0.63</b>
s heeffe	heeffe	bart thijs	bart thijs	-	0.00
bart thijs	thijs	wolfgang glänzel	wolfgang glanzel	-	0.13
bart thijs	thijs	sarah heeffe	sarah heeffe	-	0.24
<b>bart thijs</b>	<b>thijs</b>	<b>bart thijs</b>	<b>bart thijs</b>	<b>thijs</b>	<b>1.00</b>



[https://www.researchgate.net/profile/Wolfgang\\_Glaenzel](https://www.researchgate.net/profile/Wolfgang_Glaenzel)



Assign a distinct index number

# Clustering

- However, for an author there might appear different RG author profile pages

<b>WoS Name</b>	W Glanzel
<b>URL</b>	<a href="https://www.researchgate.net/profile/Wolfgang_Glaenzel">https://www.researchgate.net/profile/Wolfgang_Glaenzel</a>
<b>Affiliation</b>	University of Leuven; Institute for Research Organisation, Hungarian Academy of Sciences, Budapest, Hungary
 <b>SAME INDEX</b>	
<b>WoS Name</b>	W Glanzel
<b>URL</b>	<a href="http://www.researchgate.net/researcher/5559900990_W_Glanzel">http://www.researchgate.net/researcher/5559900990_W_Glanzel</a>
<b>Affiliation</b>	University of Leuven; Institute for Research Organisation, Hungarian Academy of Sciences, Budapest, Hungary



# Re-crawling RG

<i>Publications</i>	<i>Clustering ID for an author</i>	<i>Found RG Index for the related cluster ID</i>	<i>RG URL to re-crawl and search for the related publications</i>	<i>Results after re-crawling</i>
Pub1	23	9	<a href="https://www.researchgate.net/profile/Wolfgang_Glaenzel">https://www.researchgate.net/profile/Wolfgang_Glaenzel</a>	9
Pub2	23	-		9
Pub3	23	9		9
Pub4	23	9		9
Pub5	23	9		9
Pub6	23	9		9
Pub7	23	-		9
Pub8	23	-		-

# Re-crawling RG

Wolfgang Glänzel KU Leuven, Leuven  
Quantitative Social Research, Probability Theory  
PhD  
41.57

OVERVIEW CONTRIBUTIONS INFO STATS

**260**  
PUBLICATIONS

10.45k Reads 7,851 Citations 408.77 Impact Points View stats

University of Leuven  
Department of Managerial Economics, Strategy and ...  
Leuven, Belgium

↓

[https://www.researchgate.net/profile/Wolfgang\\_Glaenzel/publications](https://www.researchgate.net/profile/Wolfgang_Glaenzel/publications)  
[https://www.researchgate.net/profile/Wolfgang\\_Glaenzel/publications/2](https://www.researchgate.net/profile/Wolfgang_Glaenzel/publications/2)  
[https://www.researchgate.net/profile/Wolfgang\\_Glaenzel/publications/3](https://www.researchgate.net/profile/Wolfgang_Glaenzel/publications/3)  
...

↓

OVERVIEW CONTRIBUTIONS INFO STATS

PUBLICATIONS

Cluster ID	Wos Name	Wos Title
23	W Glanzel	Title_1
23	W Glanzel	Title_2
23	W Glanzel	Title_3
23	W Glanzel	Title_4
23	W Glanzel	Title_5
23	W Glanzel	Title_6

Article: Greetings from the new Editor-in-Chief  
Wolfgang Glänzel  
(Show abstract)  
No preview - Article - Jan 2014 - Scientometrics  
Request full-text

Article: High-end performance or outlier? Evaluating the tail of scientometric distributions  
Wolfgang Glänzel  
(Show abstract)  
No preview - Article - Oct 2013 - Scientometrics  
Request full-text

Article: The 8th International Conference on Webometrics, Informetrics and Scientometrics & 13th COLLNET Meeting

Similarity > 0.70 ?

Similarity > 0.70 ?

Similarity > 0.70 ?

# Results

	RG Crawling	RG Re-Crawling	Total
Correct Matches	15,520	716	16,236
Citation	42,383	83,336	125,719

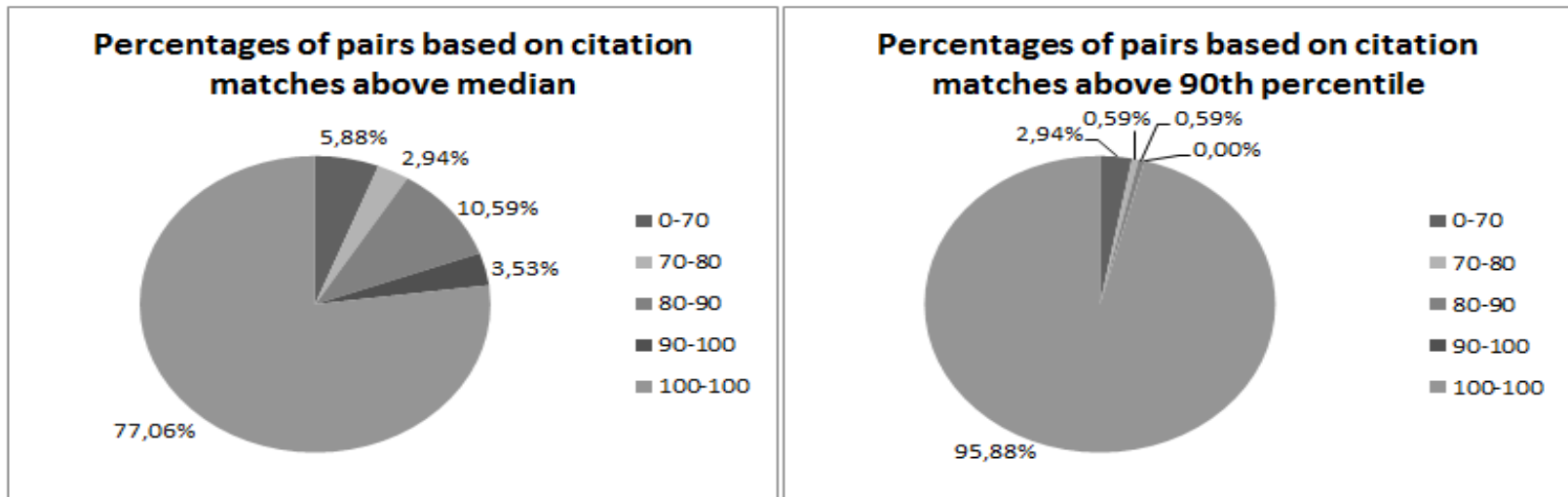
- 16,236 (73,40%) clusters were confirmed
- 125,719 (85,72%) citations of whole dataset
- Re-crawling RG contributed a lot

# Results

- Too many singletons so we chose only those authors having at least 10 publications (183 authors).
- All publications and their citations were found manually for comparison
- F score of 0.955 based on only CC results
- F score of 0.947 based on common CC & RG results
- CC failed on distinguishing names from Far East

# Results

- Bornmann (2013): Median value of citations is used for some evaluation schemes.
- Bornmann (2013): Top 10% cited publications for an author can be counted as significant publications.



# Conclusions

- Highly cited papers are found in RG
- RG can be confirmation tool to clustering tasks, especially when large datasets and reliability are in question
- RG can be a partly solution to very common Chinese names