

# Quantifying Interdisciplinarity of a Generalist Journal

Juste Raimbault

Received: date / Accepted: date

**Abstract** Insert your abstract here. Include keywords, PACS and mathematical subject classification numbers as needed.

**Keywords** Bibliometrics · Semantic Analysis

## 1 Introduction

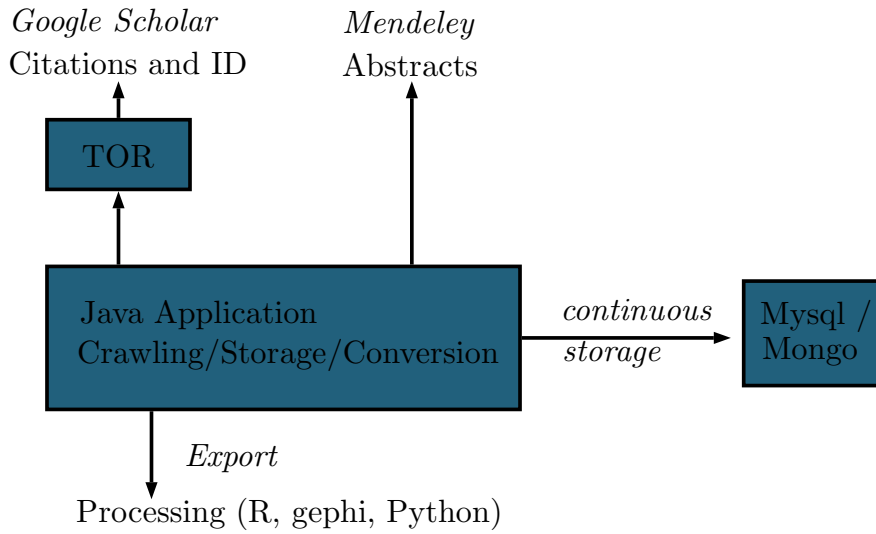
Semantic analysis does not contain all the information on disciplinary compartmentation nor on patterns of propagation of scientific knowledge as the ones contained in citation networks for example. Furthermore, data collection in the previous algorithm is subject to convergence towards self-consistent themes because of the proper structure of the method. It may give more insight about scientific social patterns of ontological choices in modeling to study communities in broader networks, that would more correspond to disciplines (or sub-disciplines depending on granularity level).

Previous works in quantitative epistemology using various types of networks have shown interesting potentialities. For the citation network, a good predicting power for citation patterns is for example obtained in [?]. Co-authorship networks can also be used for predictive models [?]. A multilayer network approach was recently proposed in [?], using bipartites networks of papers and scholars, in order to produce measures of interdisciplinarity. Disciplines can be stratified into layers to reveal communities between them and therein collaboration patterns [?]. Keyword networks are used in other fields such as economics of technology [?,?].

We describe here a study implementing these ideas for the particular case of a scientific journal for which bibliographical data is difficult to obtain, that is **cybergeo**, an electronic journal in theoretical and quantitative geography,

---

J. Raimbault  
UMR CNRS 8504 Géographie-cités  
E-mail: juste.raimbault@polytechnique.edu



**Fig. 1** Heterogeneous Bibliographical Data Collection. Architecture of the application for content (semantic data), metadata and citation data collection.

that is concerned with open science issues such as peer-review ethics transparency [?]. Our approach combine semantic communities analysis (as done in [?] but with keyword extraction ; [?] analyses semantic networks of political debates) with citation network to extract e.g. interdisciplinarity measures.

## 2 Material and Methods

### 2.1 Data Collection

#### 2.1.1 Implementation

The general architecture for data collection is presented in Fig. ?? . Citation data is collected from **Google Scholar**, that is the only source for incoming citations [?] in our case as the journal is not referenced in other databases. We are aware of the possible biases using this single source [?]<sup>1</sup>, but these critics are more directed towards search results than citation counts.

Text processing is done the same way as in previous section, expect that a particular treatment is done to language detection using *stop-words* and a specific tagger **TreeTagger** is used for other languages than english [?].

<sup>1</sup> or see <http://iscpif.frblog201602the-strange-arithmetic-of-google-scholars>

### 3 Results

#### 3.1 Citation Network Properties

#### 3.2 Semantic Communities Construction

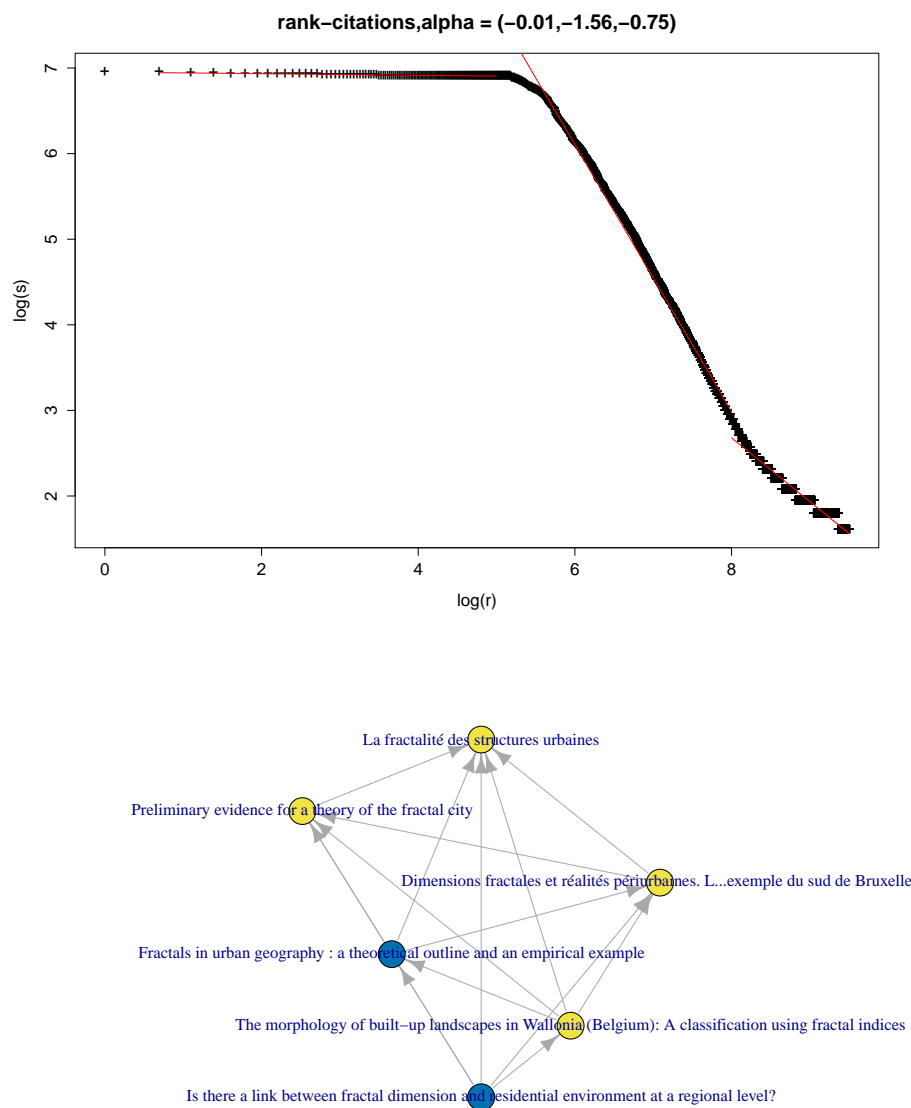
#### 3.3 Measures of Interdisciplinarity

We show in figures 2 and 3 preliminary results on citation and semantic network. We are able by the reconstruction of the citation network at depth  $\pm 1$  from the original 1000 references of the journal to retrieve around  $45 \cdot 10^6$  references, on which  $2.1 \cdot 10^6$  are retrieved with abstract text allowing semantic analysis. We retrieve by community detection in the semantic network typical geographical disciplines, such as :

- Hydrology : water, basin, river, capac
- Traffic : traffic, road, vehicl
- Biogeography : habitat, soil, veget, ecosystem
- Political Science : polit, cultur, societi, debat
- Economy : market, economi, privat, competit, industri
- Transportation : transport, travel
- Teledetection : cluster, imag, classif, satellit
- Education : educ, age, student, school
- Health : diseas, infect
- GIS : gi, geograph inform system
- Social geography : neighborhood, resid

Distribution of keywords within reconstructed disciplines provides an article-level interdisciplinarity, and we can construct various measures at the journal level. Combination of citation and semantic layers in the hyper-network provide second order interdisciplinarity measures. The construction of null models for comparison and the collection of currently missing data (journals for other papers) are currently ongoing so these results are not presented here.

**Acknowledgements** The author would like to thank the editorial board of Cybergeo, and more particularly Denise Pumain, for having offered the opportunity to work on that subject and provided the production database of the journal.



**Fig. 2** Properties of the citation network. Top : Rank-size plot of in-degrees ; three superposing successive regimes must correspond to different literature types or practices across disciplines. Bottom : example of a maximal clique in the citation network, paper of **cybergeogeo** being in blue.

