

Exploration of an interdisciplinary scientific landscape

J. Raimbault^{1,2,3}

*juste.raimbault@polytechnique.edu

¹Center for Advanced Spatial Analysis, University College London

²UPS CNRS 3611 Complex Systems Institute Paris

³UMR CNRS 8504 Géographie-cités

MLIS 2020
Machine Learning II
October 27th, 2020

Importance of interdisciplinary approaches for knowledge itself but also to solve complex issues such as global change and sustainability

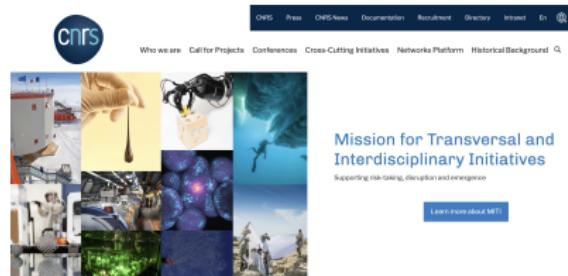
→ different types of integration between disciplines in theory
[Chavalarias et al., 2009] and in practice [Dupuy and Benguigui, 2015]



The screenshot shows the European Research Council (ERC) homepage. It features the ERC logo (a stylized 'erc' in orange dots) on the left. The main title 'European Research Council' is in orange, with the subtitle 'Supporting top researchers from anywhere in the world' in smaller white text below it. Below the title are navigation links: 'FUNDING ▶', 'PROJECTS & FIGURES ▶', 'NEWS & EVENTS ▶', 'MANAGING YOUR PROJECT ▶', and 'ABOUT ERC ▶'. At the bottom, there's a breadcrumb trail: 'Home ▶ News ▶ Supporting Interdisciplinarity, a Challenging Obligation'.

[Supporting Interdisciplinarity, a Challenging Obligation](#)

SUPPORTING INTERDISCIPLINARITY, A CHALLENGING OBLIGATION



The screenshot shows the CNRS (Centre National de la Recherche Scientifique) website. At the top right, there are links for 'CNRS', 'Press', 'CNRS-News', 'Documentation', 'Recruitment', 'Directory', 'Imprint', 'En', and a magnifying glass icon for search. Below the header is a grid of nine small images representing various scientific fields: a satellite, a close-up of a plant, a robotic arm, a brain scan, a cell, a diver, a forest, a landscape, and two people in a field. To the right of the images, the text 'Mission for Transversal and Interdisciplinary Initiatives' is displayed, along with the subtitle 'Supporting risk-taking, disruption and emergence' and a blue button labeled 'Learn more about MTI'.

Quantitative Studies of Science beyond bibliometrics
[Cronin and Sugimoto, 2014]:

- Maps of science [Börner et al., 2012][Leydesdorff and Rafols, 2009]
- Modeling science dynamics [Börner et al., 2011]
- Modeling social processes in science [Edmonds et al., 2011]
- Citation [Shibata et al., 2008], co-authorship, semantic networks [Gaumont et al., 2017]
- Quantitative epistemology [Chavalarias and Cointet, 2013]

- Difference between multi- and interdisciplinary, and empirical, theoretical or methodological interdisciplinarity
[Huutoniemi et al., 2010]
- Specialization indices (Rao-Stirling) [Larivière and Gingras, 2010]
- Diversity indices (Leinster-Cobbold) [Mugabushaka et al., 2016]
- Network-based indices [Leydesdorff, 2007]
[Rafols and Meyer, 2009]
- Semantic aspects of interdisciplinarity [Nichols, 2014]
[Bouveyron et al., 2016]

Difficulty to quantify interdisciplinarity (i) through multiple complementary dimensions; (ii) when data is not straightforward to gather.

For a case study journal in Geography (Cybergeo, European Journal of Geography <https://journals.openedition.org/cybergeo/>)

- construct a database from heterogenous sources
- quantify interdisciplinarity with citation and semantic networks

Scientometrics (2019) 119:617–641
<https://doi.org/10.1007/s11192-019-03090-3>



Exploration of an interdisciplinary scientific landscape

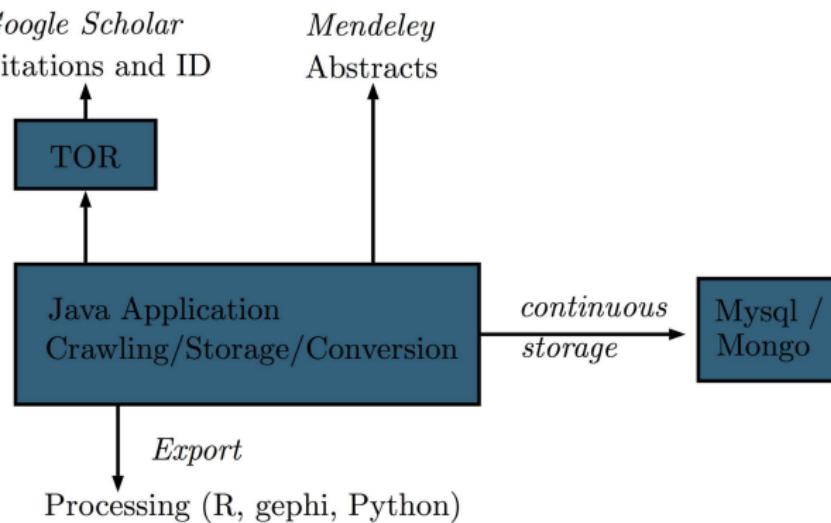
Juste Raimbault^{1,2} 

Received: 3 December 2017 / Published online: 25 March 2019
© Akadémiai Kiadó, Budapest, Hungary 2019

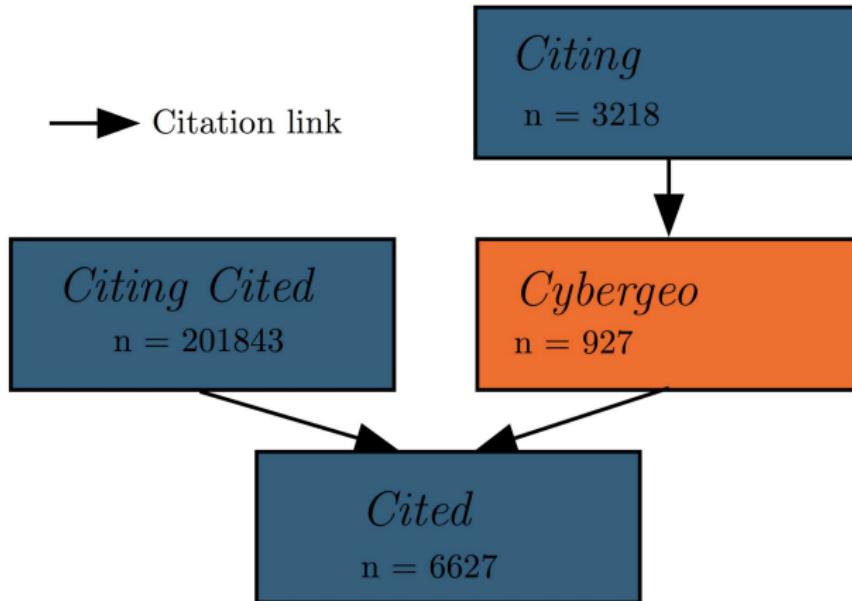
Data collection

Data collected from heterogenous sources: journal production database, google scholar for citation links, Mendeley for abstracts.

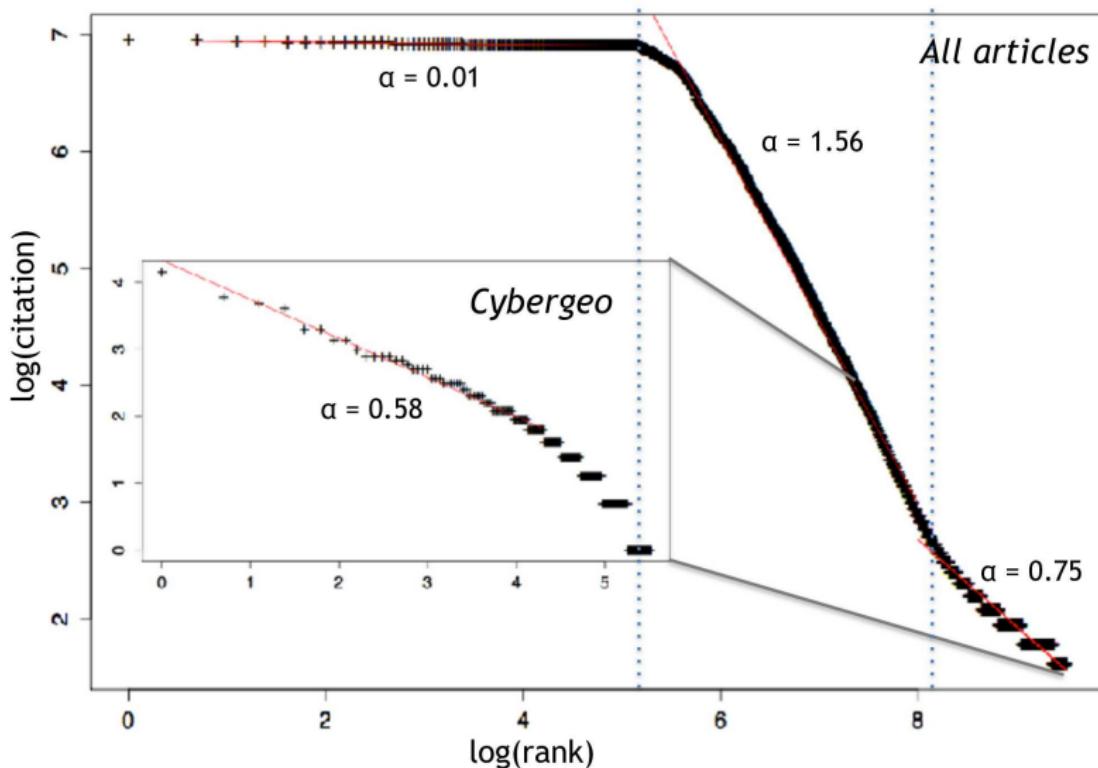
Refactored open source java library for data collection:
<https://github.com/JusteRaimbault/BiblioData>



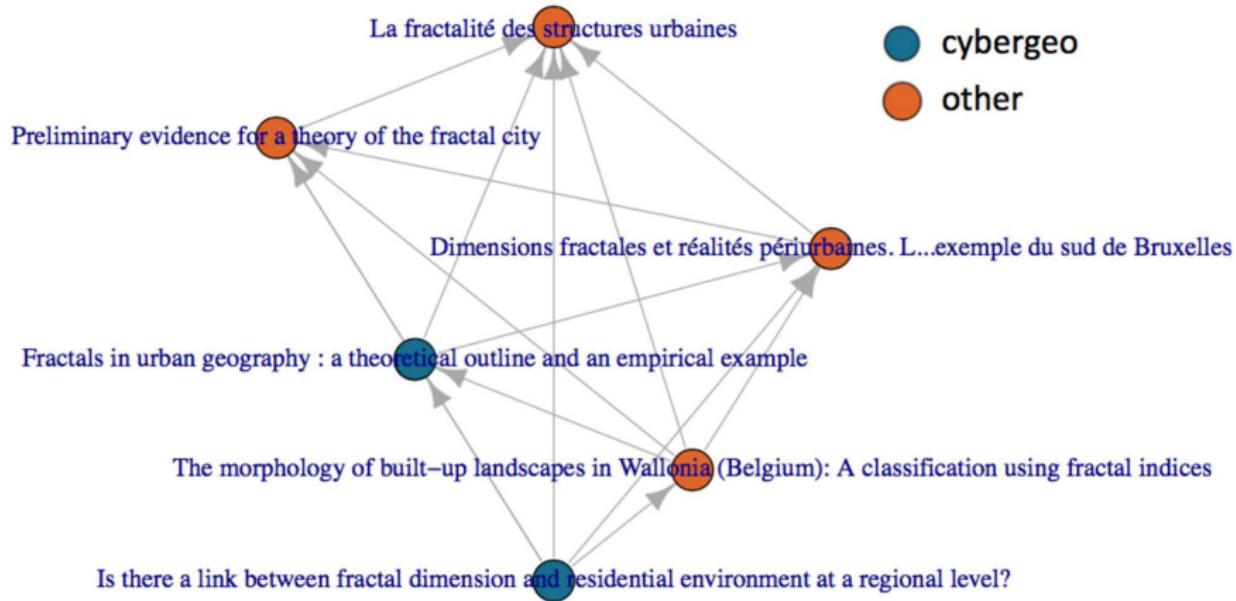
Citation network with abstract coverage: $\simeq 2.1 \cdot 10^5$ papers



Citation network properties



Example of the largest citation clique



Citation communities



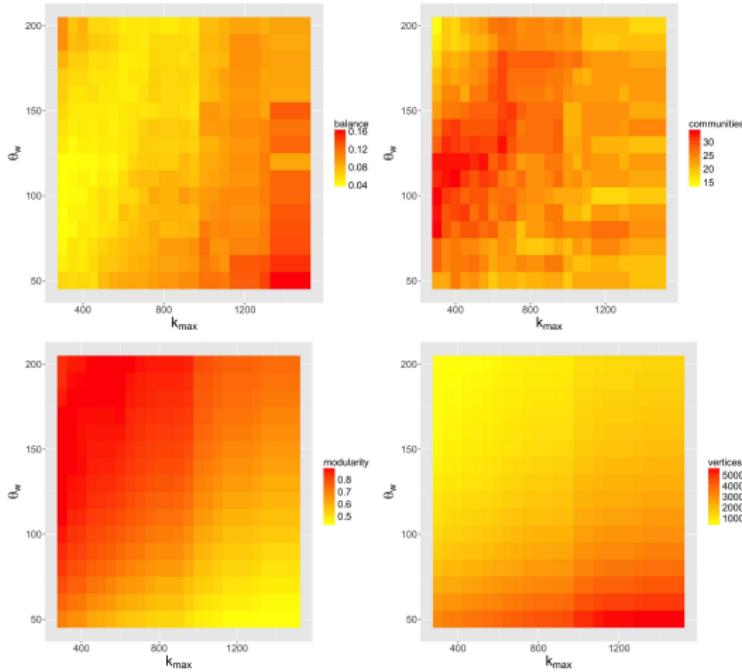
*Louvain algorithm community detection to construct endogenous citation communities
(modularity of 0.71)*

Construction of the semantic network as a co-occurrence network between keywords extracted following [Bergeaud et al., 2017]:

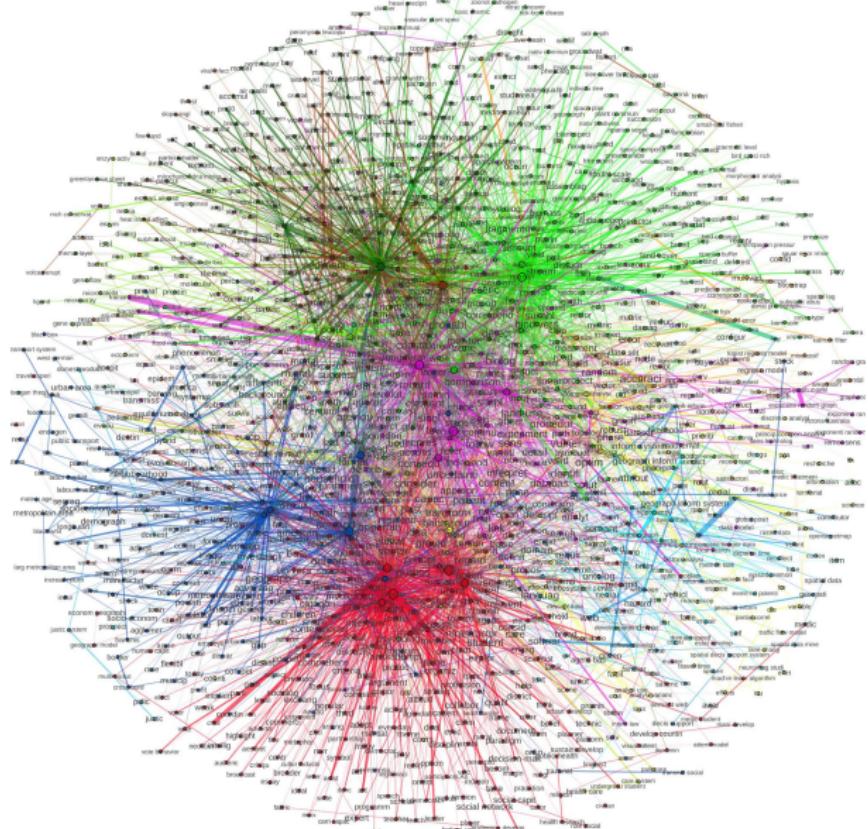
- 1 Language detection using stop-words [Baldwin and Lui, 2010]
- 2 Part-Of-Speech tagging (`nltk` or `TreeTagger` [Schmid, 1994]); stemming
- 3 Construction of potential n-grams: nouns and adjectives up to size 4
- 4 Estimation of n-gram relevance following the deviation to the expected statistical distribution of co-occurrences (chi-squared test)

Sensitivity analysis of the semantic network

Additional filtering procedure to remove relevant but common keywords:
filter on edge weight and maximal degree; values chosen based on multiples objectives including modularity and network size.



Semantic network visualization

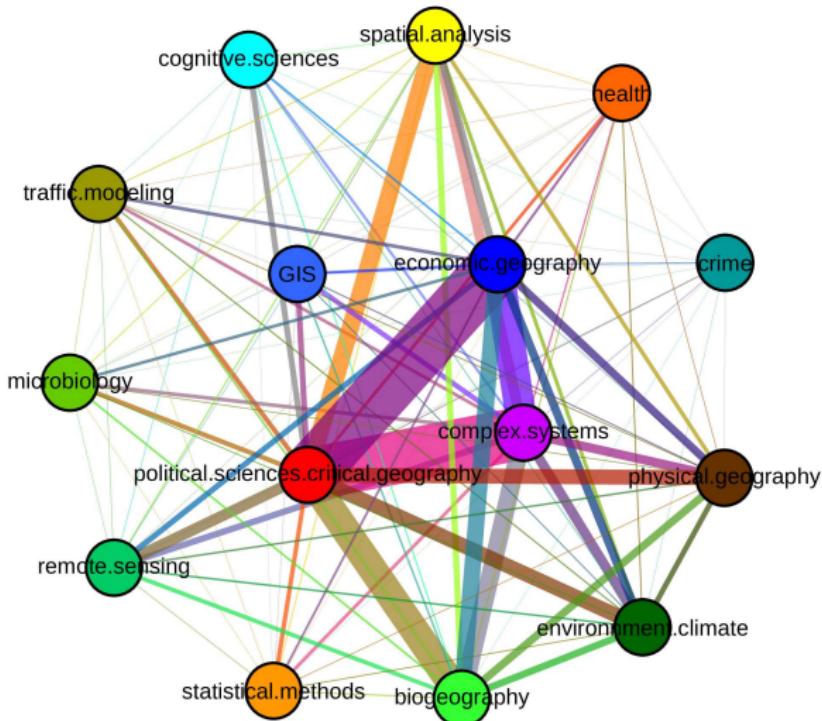


Semantic communities

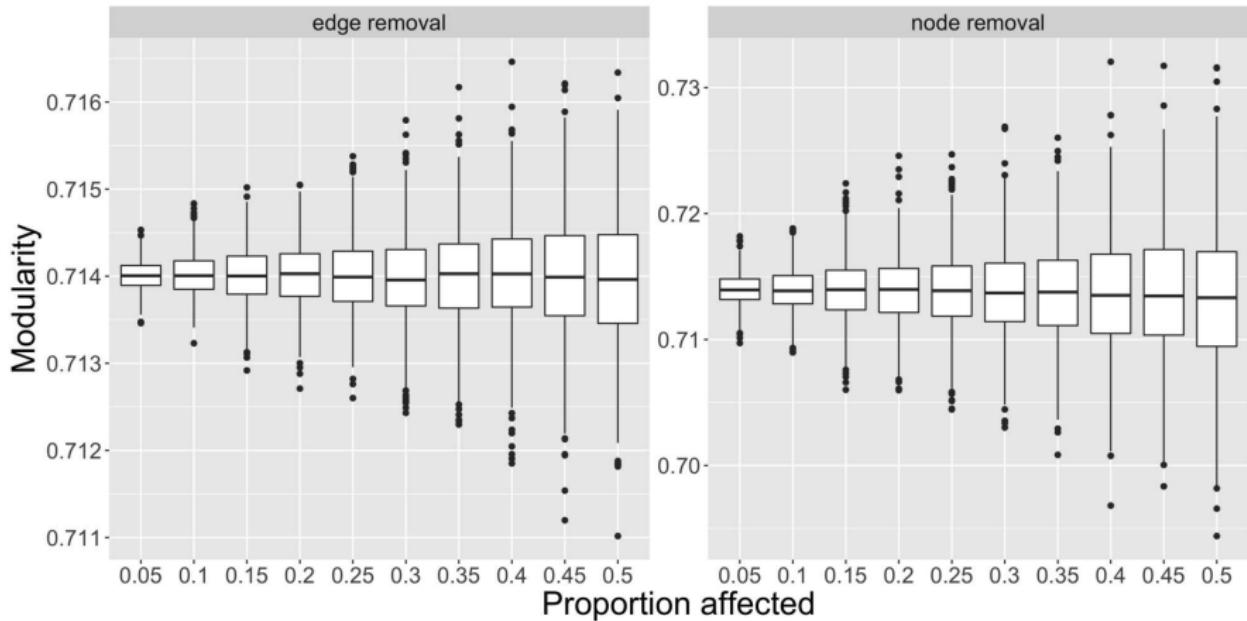
Table 1 Semantic communities reconstructed from community detection in the semantic network

Name	Size	Keywords
Political sciences/critical geography	535	decision-mak, polit ideolog, democraci, stakehold, neoliber
Biogeography	394	plant densiti, wood, wetland, riparian veget
Economic geography	343	popul growth, transact cost, socio-econom, household incom
Environment/climate	309	ice sheet, stratospher, air pollut, climat model
Complex systems	283	scale-fre, multifract, agent-bas model, self-organ
Physical geography	203	sedimentari, digit elev model, geolog, river delta
Spatial analysis	175	spatial analysi, princip compon analysi, heteroscedast, factor analysi
Microbiology	118	chromosom, phylogeneti, borrelia
Statistical methods	88	logist regress, classifi, kalman filter, sampl size
Cognitive sciences	81	semant memori, retrospect, neuroimag
GIS	75	geograph inform scienc, softwar design, volunt geograph inform, spatial decis support
Traffic modeling	63	simul model, lane chang, traffic flow, crowd behavior
Health	52	epidem, vaccin strategi, acut respiratori syndrom, hospit
Remote sensing	48	land-cov, landsat imag, lulc
Crime	17	crimin justic system, social disorgan, crime

Interactions between communities

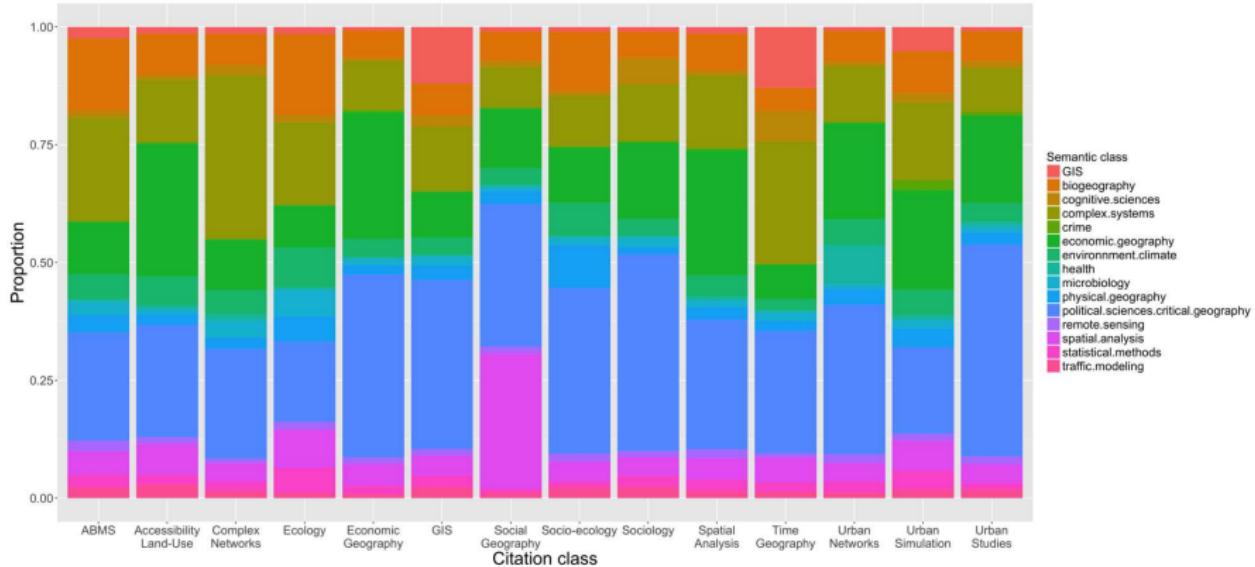


Sensitivity to corpus definition

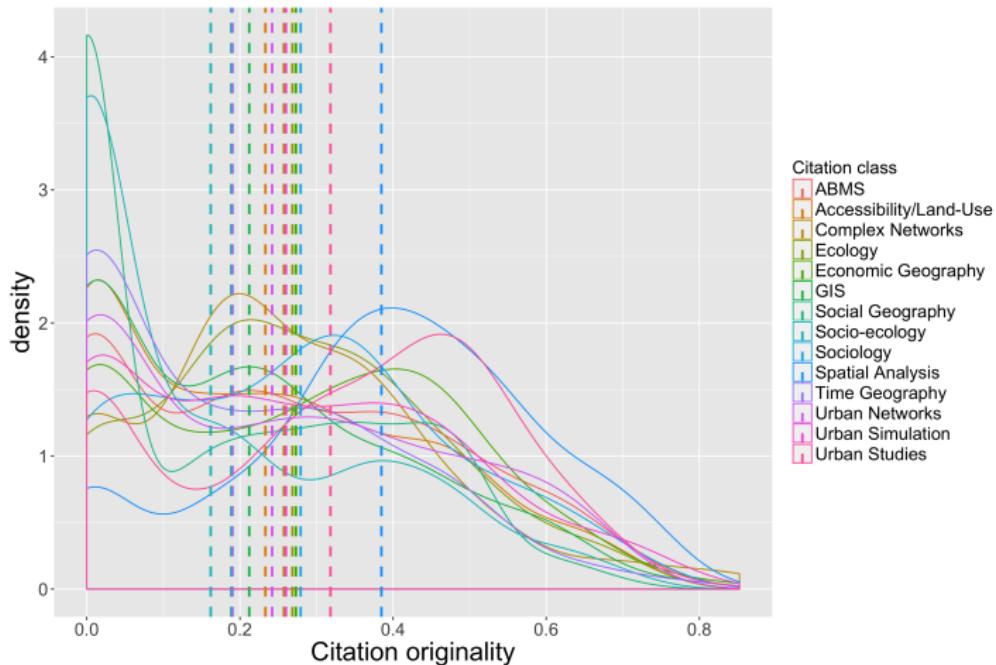


Impact of node and edge removal on the optimal modularity of the citation network

Semantic composition of citation communities

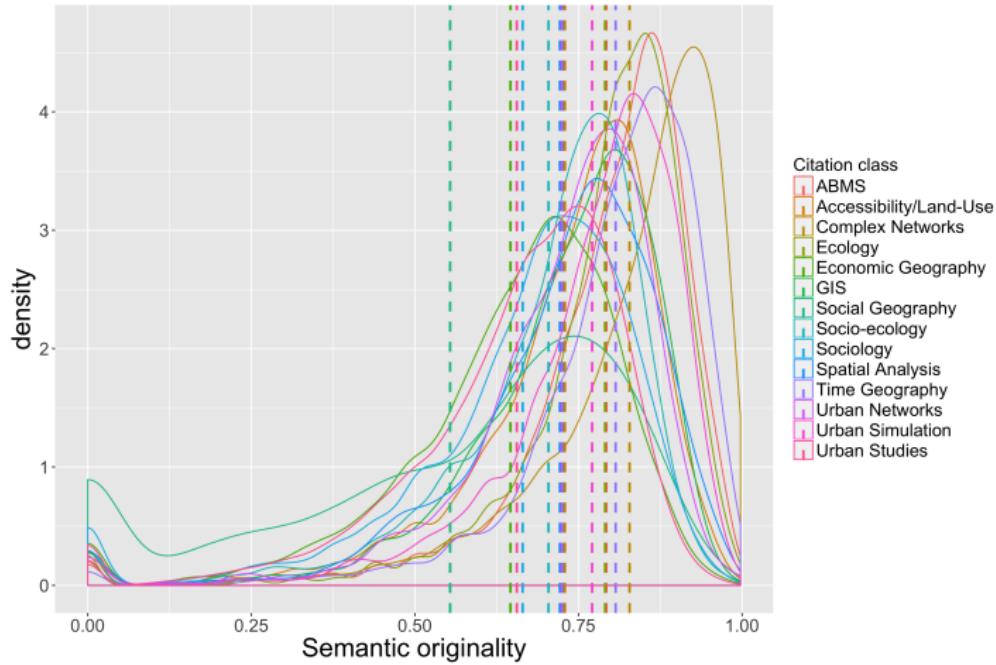


Measuring interdisciplinarity



Distribution of originalities (Herfindhal index) for how references are cited by different citation classes

Measuring interdisciplinarity



Distribution of semantic originalities

Developments

- Journal dynamics and benchmarking, reflexivity for authors; fostering Open Science [Raimbault et al., 2019]
- Performance of the semantic classification for citation link prediction
- Correspondence of terms between disciplines

Applications

- Quantification of Domains of Knowledge [Raimbault, 2017]
- Complementary dimensions in the structure of science
- Spatial diffusion of knowledge [Raimbault, 2020]

- A multi-dimensional approach to understand patterns of interdisciplinarity
- Open tools and methodology to foster Open Science

Paper at <https://doi.org/10.1007/s11192-019-03090-3>
(open version at <https://arxiv.org/abs/1712.00805>)

Open repositories for the paper and library

<https://github.com/JusteRaimbault/HyperNetwork>
<https://github.com/JusteRaimbault/BiblioData>

Data at <http://dx.doi.org/10.7910/DVN/VU2XKT>

-  Baldwin, T. and Lui, M. (2010).
Language identification: The long and the short of the matter.
In *Human Language Technologies: The 2010 Annual Conference of the North American Chapter of the Association for Computational Linguistics*, pages 229–237. Association for Computational Linguistics.
-  Bergeaud, A., Potiron, Y., and Raimbault, J. (2017).
Classifying patents based on their semantic content.
PloS one, 12(4):e0176310.
-  Börner, K., Glänzel, W., Scharnhorst, A., and Van den Besselaar, P. (2011).
Modeling science: studying the structure and dynamics of science.
Scientometrics, 89(1):347–348.

-  Börner, K., Klavans, R., Patek, M., Zoss, A. M., Biberstine, J. R., Light, R. P., Larivière, V., and Boyack, K. W. (2012).
Design and update of a classification system: The ucsd map of science.
PloS one, 7(7):e39464.
-  Bouveyron, C., Latouche, P., and Zreik, R. (2016).
The stochastic topic block model for the clustering of vertices in networks with textual edges.
Statistics and Computing, pages 1–21.
-  Chavalarias, D., Bourgine, P., Perrier, E., Amblard, F., Arlabosse, F., Auger, P., Baillon, J.-B., Barreteau, O., Baudot, P., Bouchaud, E., et al. (2009).
French roadmap for complex systems 2008-2009.

-  Chavalarias, D. and Cointet, J.-P. (2013).
Phylomemetic patterns in science evolution—the rise and fall of scientific fields.
PloS one, 8(2):e54847.
-  Cronin, B. and Sugimoto, C. R. (2014).
Beyond bibliometrics: Harnessing multidimensional indicators of scholarly impact.
MIT Press.
-  Dupuy, G. and Benguigui, L. G. (2015).
Sciences urbaines: interdisciplinarités passive, naïve, transitive, offensive.
Métropoles, (16).

-  Edmonds, B., Gilbert, N., Ahrweiler, P., and Scharnhorst, A. (2011). Simulating the social processes of science.
Journal of Artificial Societies and Social Simulation, 14(4):14.
-  Gaumont, N., Panahi, M., and Chavalarias, D. (2017). Methods for the reconstruction of the socio-semantic dynamics of political activist twitter networks.
-  Huutoniemi, K., Klein, J. T., Bruun, H., and Hukkinen, J. (2010). Analyzing interdisciplinarity: Typology and indicators.
Research Policy, 39(1):79–88.
-  Larivière, V. and Gingras, Y. (2010). On the relationship between interdisciplinarity and scientific impact.
Journal of the Association for Information Science and Technology, 61(1):126–131.

-  Leydesdorff, L. (2007). Betweenness centrality as an indicator of the interdisciplinarity of scientific journals. *Journal of the Association for Information Science and Technology*, 58(9):1303–1319.
-  Leydesdorff, L. and Rafols, I. (2009). A global map of science based on the isi subject categories. *Journal of the American Society for Information Science and Technology*, 60(2):348–362.
-  Mugabushaka, A.-M., Kyriakou, A., and Papazoglou, T. (2016). Bibliometric indicators of interdisciplinarity: the potential of the leinster–cobbold diversity indices to study disciplinary diversity. *Scientometrics*, 107(2):593–607.

-  Nichols, L. G. (2014).
A topic model approach to measuring interdisciplinarity at the national science foundation.
Scientometrics, 100(3):741–754.
-  Rafols, I. and Meyer, M. (2009).
Diversity and network coherence as indicators of interdisciplinarity: case studies in bionanoscience.
Scientometrics, 82(2):263–287.
-  Raimbault, J. (2017).
An applied knowledge framework to study complex systems.
In *Complex Systems Design & Management*, pages 31–45.

-  Raimbault, J. (2020).
A model of urban evolution based on innovation diffusion.
Artificial Life Conference Proceedings, (32):500–508.
-  Raimbault, J., Chasset, P.-O., Cottineau, C., Commenges, H., Pumain, D., Kosmopoulos, C., and Banos, A. (2019).
Empowering open science with reflexive and spatialised indicators.
Environment and Planning B: Urban Analytics and City Science, page 2399808319870816.
-  Schmid, H. (1994).
Probabilistic part-of-speech tagging using decision trees.
In *Proceedings of the international conference on new methods in language processing*, volume 12, pages 44–49. Citeseer.

-  Shibata, N., Kajikawa, Y., Takeda, Y., and Matsushima, K. (2008). Detecting emerging research fronts based on topological measures in citation networks of scientific publications. *Technovation*, 28(11):758–775.