Network analysis in quantitative epistemology towards open science and reflexivity

J. Raimbault[1,2,3]

*j.raimbault@ucl.ac.uk

[1]Center for Advanced Spatial Analysis, University College London
[2]UPS CNRS 3611 Complex Systems Institute Paris
[3]UMR CNRS 8504 Géographie-cités

CASA Networks Seminar
November 26th, 2021

Introduction

Knowledge production processes studied from diverse viewpoints:

- Sociology of science [Callon and Latour, 2013]
- Philosophy of science [Hacking, 1999]
- Innovation economics [Grupp et al., 1998]
- Cognitive science [De Mey, 2012]
- Scientometrics [Mingers and Leydesdorff, 2015]

→ Recent applications of network science [Fortunato et al., 2018]

→ Towards a *quantitative epistemology* [Chavalarias and Cointet, 2013]

This presentation: *Synthesis of different network analysis applied to quantitative epistemology, within the context of open science and reflexivity*

PLOS | ONE

RESEARCH ARTICLE

Classifying patents based on their semantic content

Antonin Bergeaud[1☯], Yoann Potiron[2☯], Juste Raimbault[3,4☯*]

1 Paris School of Economics - EHESS and Bank of France, Paris, France, **2** Faculty of Business and Commerce, Keio University, Tokyo, Japan, **3** UMR CNRS 8504 Géographie-cités, Université Paris VII, Paris, France, **4** UMR-T 9403 IFSTTAR LVMT, Ecole Nationale des Ponts et Chaussées, Champs-sur-Marne, France

☯ These authors contributed equally to this work.
* juste.raimbault@polytechnique.edu

Check for updates

Abstract

In this paper, we extend some usual techniques of classification resulting from a large-scale data-mining and network approach. This new technology, which in particular is designed to be suitable to big data, is used to construct an open consolidated database from raw data on

Why study patents?

$\rightarrow$ Although imperfect, patents are the most commonly used measure of innovation in economics
$\rightarrow$ Particular case of the ecology and evolution of knowledge; diffusion of knowledge

Examples:

- [Griliches, 1998]: patent as an economic indicator
- [Youn et al., 2015] interaction between technological fields; combinatorial nature of inventions
- [Bruck et al., 2016] citation network analysis to detect emerging research fronts
- [Gerken and Moehrle, 2012] [Tseng et al., 2007] semantic analysis (remains limited to specific fields and time windows)

Proposed approach
Complement existing patent office classifications using patent semantic content

Why?: more endogenous, flexible and informative (?). And comparison with *technological* classification to detect outliers.

Takeaway results

→ *An endogenous semantic classification is constructed for the full USPTO abstracts and titles, 1976-2012*

→ *Information carried in the semantic classification is complementary to the technological classification*

US006833896B2

(12) **United States Patent**
Lee

(10) **Patent No.:** **US 6,833,896 B2**
(45) **Date of Patent:** **Dec. 21, 2004**

(54) **METHOD OF FORMING AN ARRAY SUBSTRATE FOR AN IN-PLANE SWITCHING LIQUID CRYSTAL DISPLAY DEVICE HAVING AN ALIGNMENT FILM FORMED DIRECTLY ON A THIN FILM TRANSISTOR**

(75) Inventor: **Yun-Bok Lee**, Seoul (KR)

(73) Assignee: **LG.Philips LCD Co., Ltd.**, Seoul (KR)

6,407,791	B1 *	6/2002	Suzuki et al.	....	349/129
6,642,972	B2 *	11/2003	Yoo et al.	....	349/40
2001/0040648	A1 *	11/2001	Ono et al.	....	349/43
2002/0093600	A1 *	7/2002	Chol	....	349/43

FOREIGN PATENT DOCUMENTS

JP	09-005764	1/1997
JP	09-073101	3/1997
JP	09-101538	4/1997
JP	09-105908	4/1997

Technological classes: 349 (Liquid Crystal Cells, Elements and Systems), 257 (electrodes)

Prior art keywords (Google Patents): layer, method, gate, electrode, substrate

Most central keywords (our method): substrat, semiconductor, insul, gate, transistor

Descriptive statistics

Summary statistics for the USPTO database between 1976 and 2013

- 4,666,365 utility patents from 1976 to 2013
- 70,000 in 1976 to 278,000 in 2013
- 38,756,292 citation links (84% of within-class citations)
- 270,877 patents with no citations 5 years next to publication

Does this mean that innovation is 4 time larger en 2013 than in 1976...
Not necessary. Although definitely more specialized.

Database construction

Construction of a Database from US Patent and Trademark Office redbook 1976-2012 (full patent description), which provide raw data but on separate files and different formats

Data Collection Procedure

- Automatic download of raw data file
- Parsing depending on format : `dat` or `xml` (varying schema)
- Uniformisation and storing in MongoDB

$\rightarrow$ 4,666,365 patents with text data (abstract); dated by application date (current state of knowledge, differs from the granted date which implies review processes and an exogenous intervention)

Extracting relevant n-grams

Text-mining in python with `nltk` [Bird, 2006], method adapted from [Chavalarias and Cointet, 2013]. Advantage over LDA: scalability and flexibility

- Parsing and tokenizing / pos-tagging (word functions) / stemming with `nltk`
- Selection of potential *n-grams* (with $1 \leq n \leq 3$) with the rule $\bigcap\{NN \cup VBG \cup JJ\}$
- Database insertion for instantaneous use (several days $\rightarrow$ 1min)
- Estimation of *n-grams* relevance, following co-occurrences statistical distribution (*termhood* score as chi-2 score)

Estimation of n-gram relevance

- Termhood-based selection of a fixed number of keywords K_W=100, 000 (*difficulty to select terms in general*)
- Estimation on temporal moving windows $\left[t - T_w; t\right]$ (fixed to T_W=4y after sensitivity analysis)
- Filtering of network edge (parameter θ_w), with an additional exogenous control by technological class keyword concentration to filter nodes (parameter θ_c)
- Low sensibility to the length of the time window for the network structure

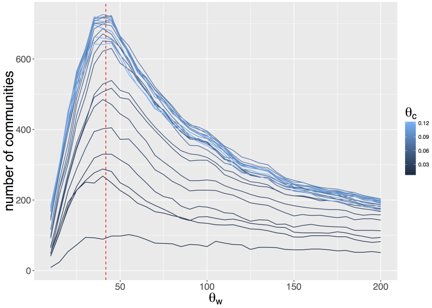
Technical aspects

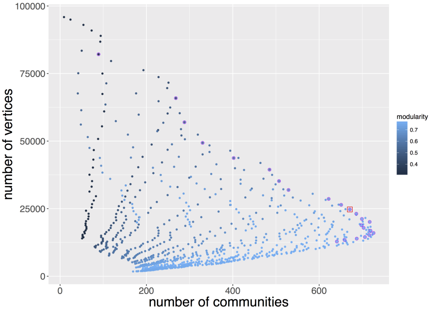
$\rightarrow$ Temporal complexity: $\mathcal{O}\left(N_P\right)$ for keyword extraction and co-occurrences (constant l_{max}^2); parallelization on a 60 cores server for a "reasonable" computation time.

$\rightarrow$ Memory complexity: co-occurrence matrices in $\mathcal{O}\left(K_W^2\right)$; necessitates around 600Go RAM when parallelized.

$\rightarrow$ Database management: MongoDB (nosql suited to big data).

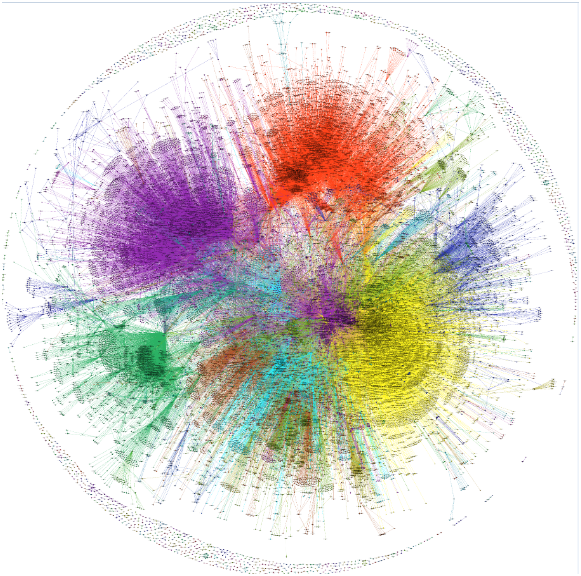
Optimizing network structure





Pareto optimization on cutoff parameters with the objectives of modularity, size and number of communities

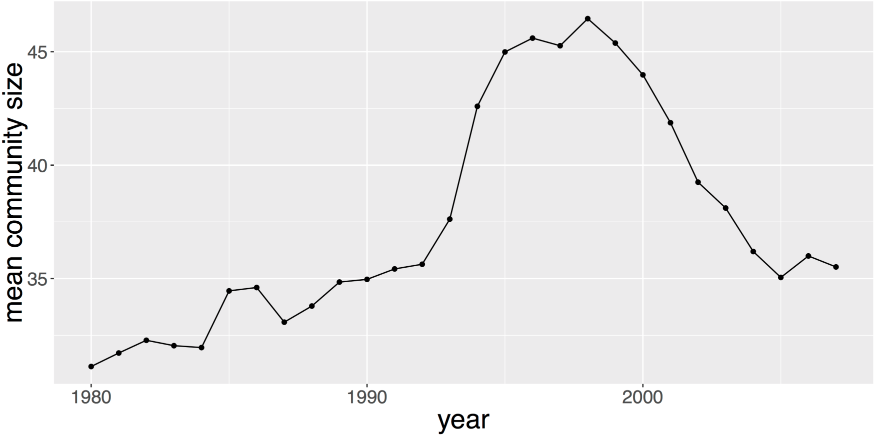
Semantic network visualization (2000-2004)



Classes Examples (2000-2004)

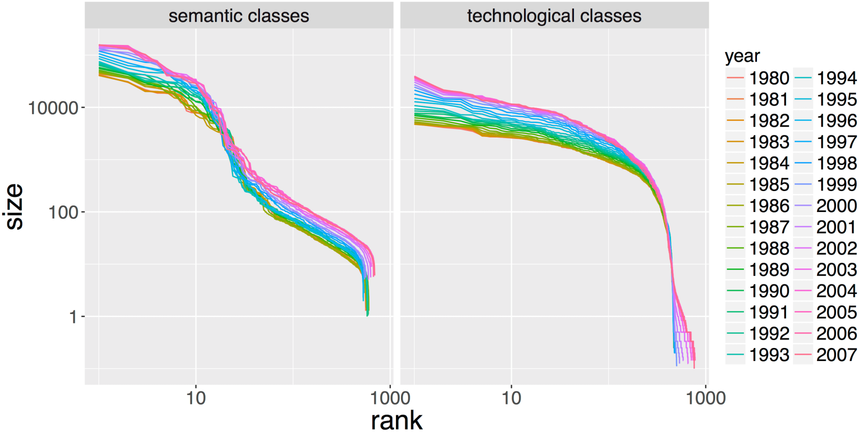
- Memory devices : `semiconductor memori devic; memori cell plural; memori cell transistor; layer ferroelectr`
- Chemical analysis : `time-of-flight mass spectromet; chromatograph column; ion trap mass`
- Particular steel : `martensit; austenit stainless steel`
- Laser : `emit laser beam; vertic caviti surfac; vcsel`
- Sewing : `circular knit machin; stitch; sew machin; embroideri`
- Lithography : `lithograph mask; project beam radiat; heat-sensit; planograph print plate`
- Tobacco : `cigarett filter; cigarett pack; tobacco; tobacco rod`

Size of classes



A peak in average class size around 1998

Hierarchical class structure



Fat-tail distribution of class size for both classification, closer to a power-law for semantic classes

UCL

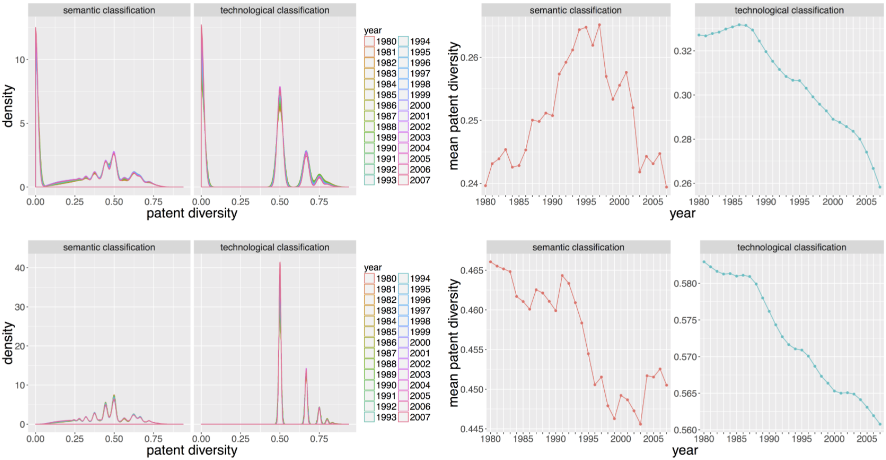
Diversity

$$D_i^z = 1 - \sum_{j1}^{N^z} p_{ij}^2, \text{ with } z \in \{tec, sem\}$$

Originality and Generality

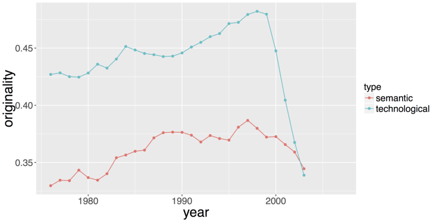
$$O_i^z = 1 - \sum_{j=1}^{N^z} \left(\frac{\sum_{i' \in I_i} p_{i'j}}{\sum_{k1}^{N^z} \sum_{i' \in I_i} p_{i'k}} \right)^2 \text{ and } G_i^z = 1 - \sum_{j=1}^{N^z} \left(\frac{\sum_{i' \in \tilde{I}_i} p_{i'j}}{\sum_{k1}^{N^z} \sum_{i' \in \tilde{I}_i} p_{i'k}} \right)^2$$

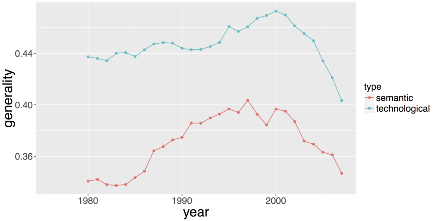
Evolution of patent diversities



General increase in average invention specialization seen both for semantic and technological; semantic regime shift in 1996

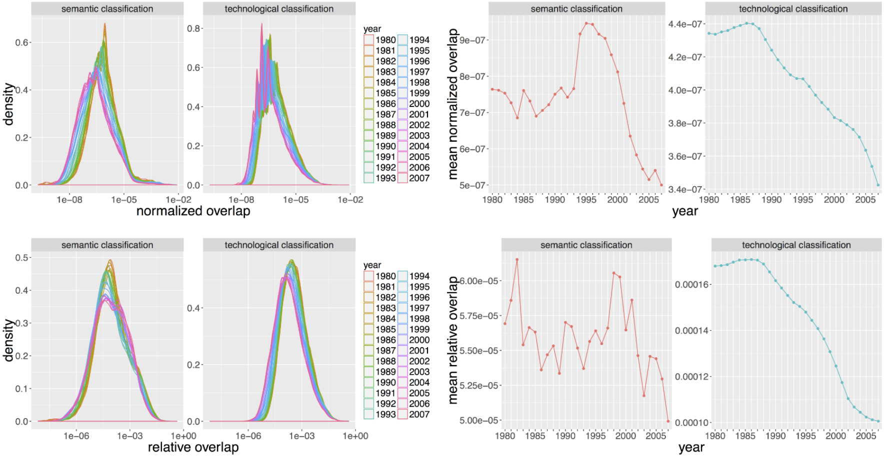
Originality and generality measures





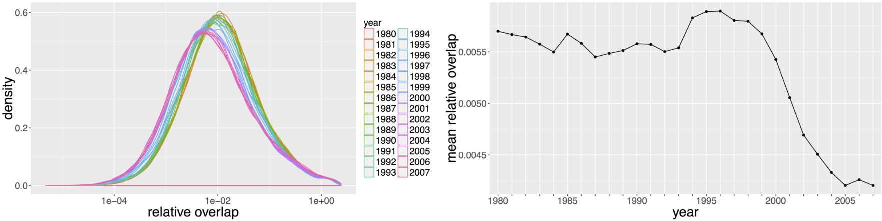
Systematically lower originality and generality (citation-based measures) for the semantic classification (consistent with the higher modularity shown thereafter).

Interaction between classes: intra-classification



Increased technological specialization; qualitative regime shift confirmed in 1996 for the semantic classification

Inter-classification overlaps



Constant then decreasing average overlap between technological and semantic classes; confirms the change in nature of inventions around 1996.

→ an impact of new information technologies (regarding knowledge production: from expert and contextualized to automatized search in references e.g.) ? Linked to structural changes in economy (increase in firm concentration since 90s) ?

Definition of modularities

Directed simple modularity [Nicosia et al., 2009]

$$Q_d^z = \frac{1}{N_P} \sum_{1 \leq i,j \leq N_P} \left[A_{ij} - \frac{k_i^{in} k_j^{out}}{N_P} \right] \delta\left(c_i, c_j\right),$$

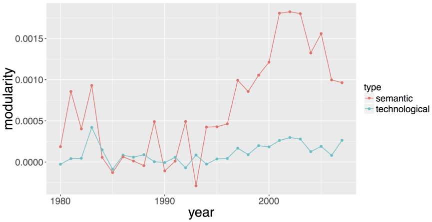
Multi-class modularity:

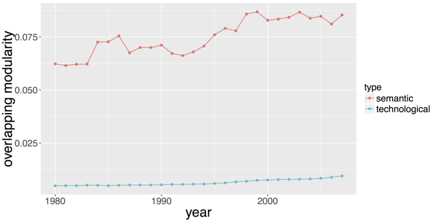
$$Q_{ov}^z = \frac{1}{N_P} \sum_{c=1}^{N^z} \sum_{1 \leq i,j \leq N_P} \left[F\left(p_{ic}, p_{jc}\right) A_{ij} - \frac{\beta_{i,c}^{out} k_i^{out} \beta_{j,c}^{in} k_j^{in}}{N_P} \right],$$

where

$$\beta_{i,c}^{out} = \frac{1}{N_P} \sum_j F\left(p_{ic}, p_{jc}\right) \text{ and } \beta_{j,c}^{in} = \frac{1}{N_P} \sum_i F\left(p_{ic}, p_{jc}\right).$$

Modularity of classifications





Semantic classification is more significant regarding the structure of the citation network, both for single-class and multi-class modularities (confirmed statistically using a simple Stochastic Block Model).

- We complete the analysis by developing a statistical model aiming at quantifying performance of both technological and semantic classification systems
- Intuitively, we look at within class citations proportion (for both technological and semantic approaches)
- Question: is the semantic classification better at predicting future within class citations?
  $\rightarrow$ short answer is yes.

Possible developments

Extensions

- Look at the correlation between patent quality indicators and centrality indices.
- Use patent-firm matching to study a *firm-level* semantic network.
- Extending the analysis to other patent offices (EPO/JPO).

Economic

- Linking semantic measures to values of firms
- Can semantic proximity helps understand M&A?
- Firms trajectories in relation to their life-cycle

Technology

- Measure of complementarity between technology?
- Possible application to detection of emerging research fronts
- An interactive exploration of semantic content?

Similar method applied to a scientific corpus

Study around the interdisciplinary geography journal Cybergeo

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

Check for updates

Exploration of an interdisciplinary scientific landscape

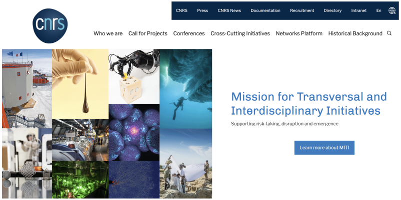
Juste Raimbault[1,2]

Received: 3 December 2017 / Published online: 25 March 2019

Interdisciplinarity

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

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



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]

Defining and measuring interdisciplinarity

- 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]

Proposed approach

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/`)

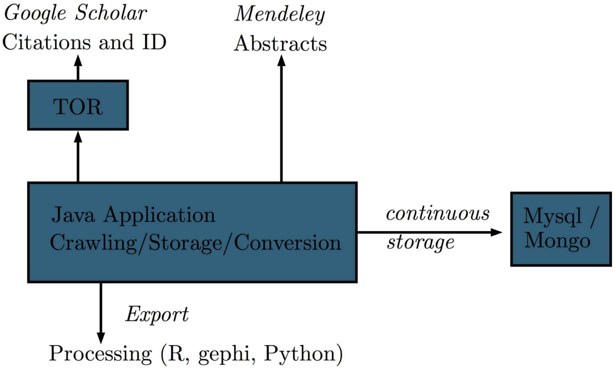
$\rightarrow$ construct a database from heterogenous sources

$\rightarrow$ quantify interdisciplinarity with citation and semantic networks

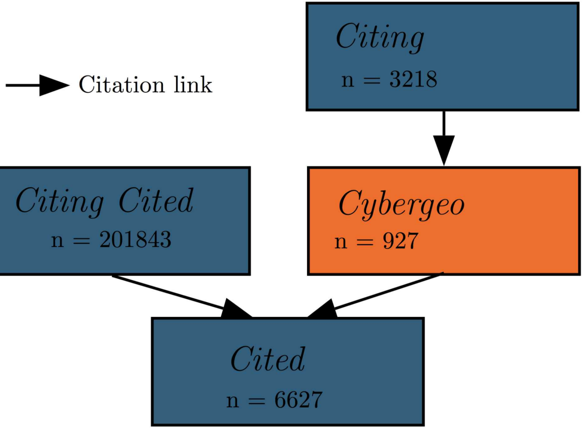
Data collection

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

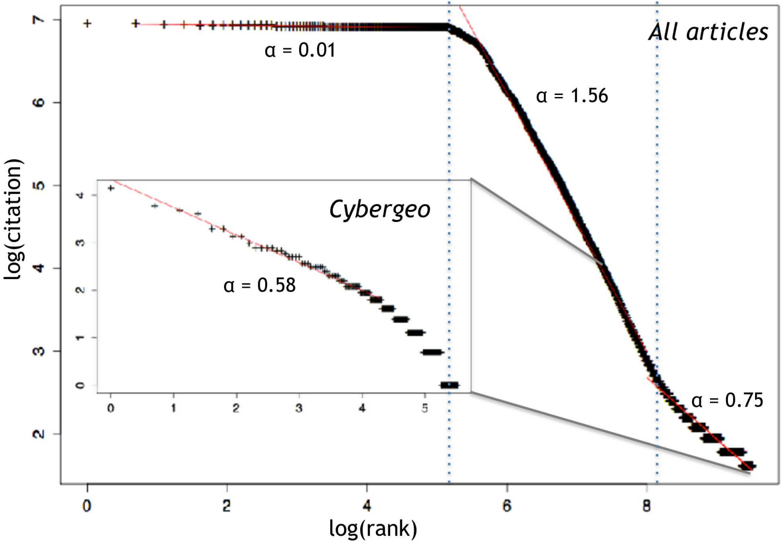
Refactored open source java library for data collection:
`https://github.com/JusteRaimbault/BiblioData`



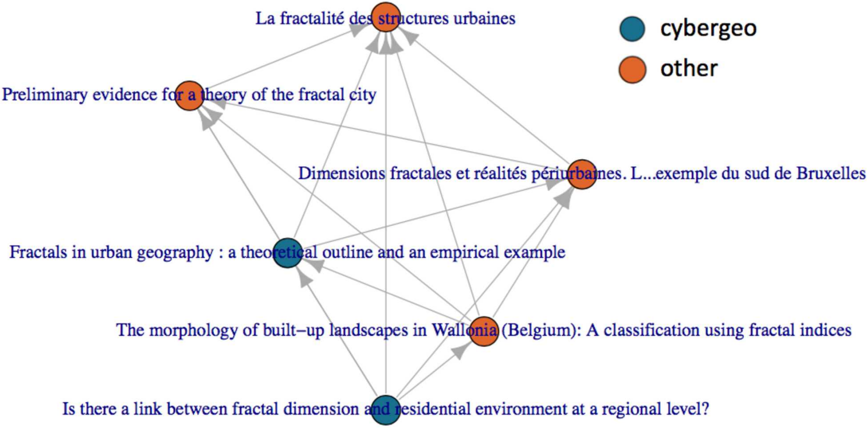
Citation network with abstract coverage: $\simeq 2.1 \cdot 10^5$ papers



Citation network properties



Example of the largest citation clique



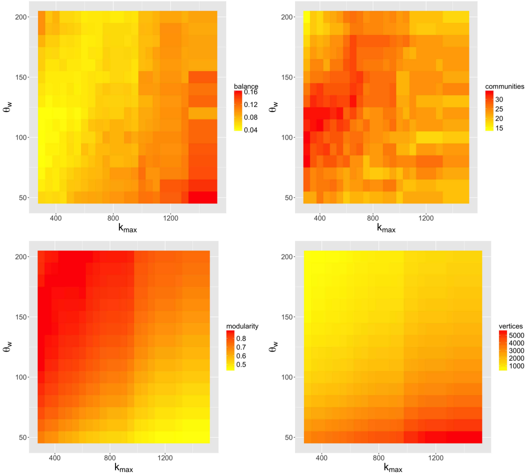
Louvain algorithm community detection to construct endogenous citation communities (modularity of 0.71)

Semantic network

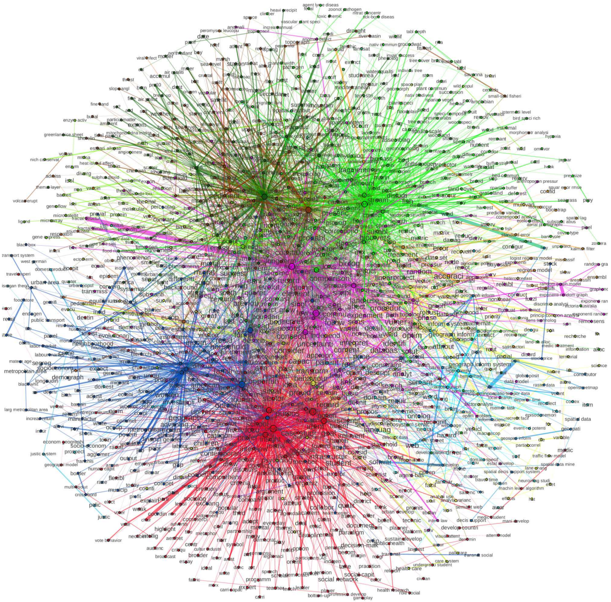
Construction of the semantic network as a co-occurence network between keywords extracted following [Bergeaud et al., 2017]:

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-occurences (chi-squared test)

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

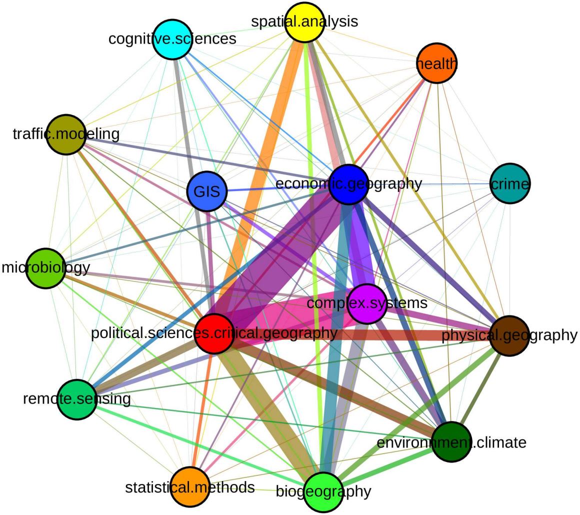


Semantic communities

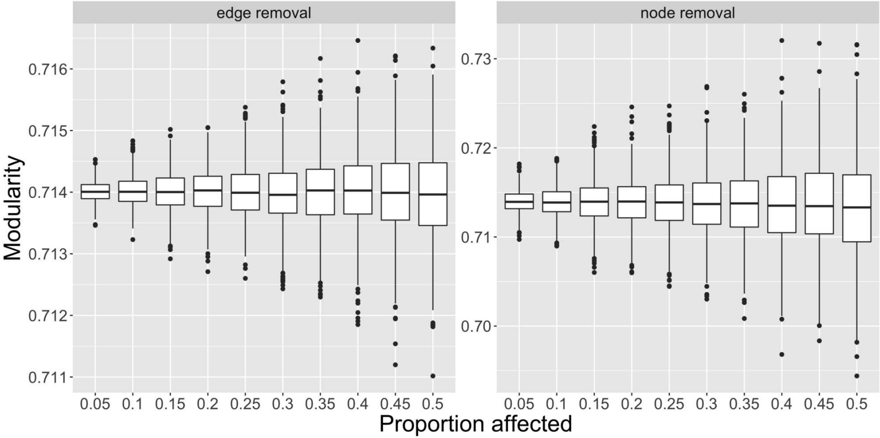
Table 1 Semantic communities reconstructed from community detection in the semantic network

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
Environnment/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, phylogenet, 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

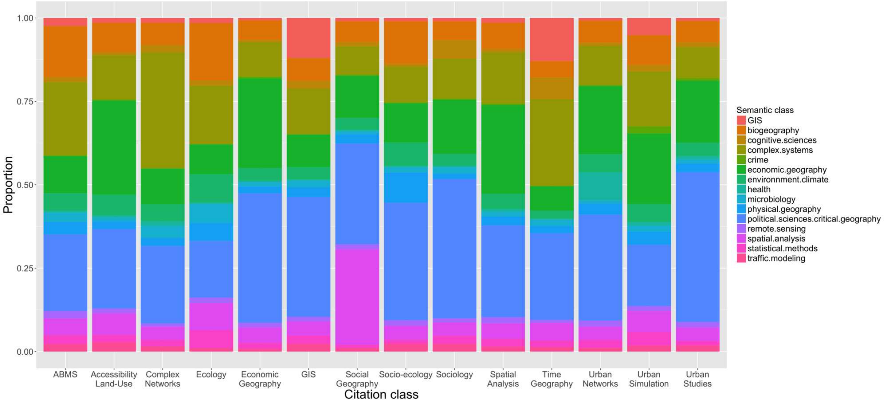


Sensitivity to corpus definition

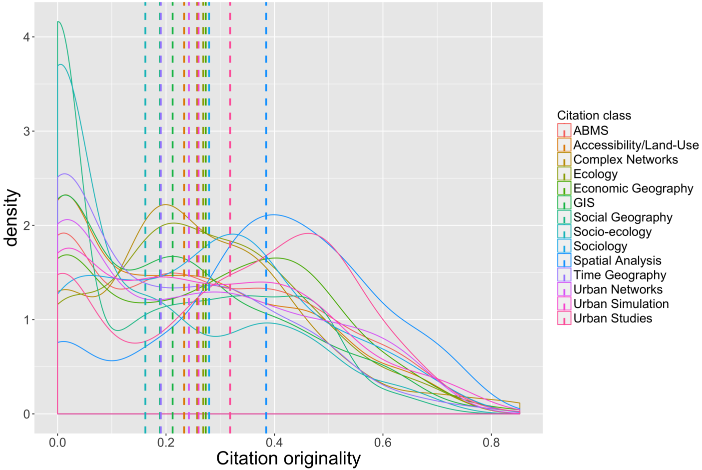


Impact of node and edge removal on the optimal modularity of the citation network

Semantic composition of citation communities

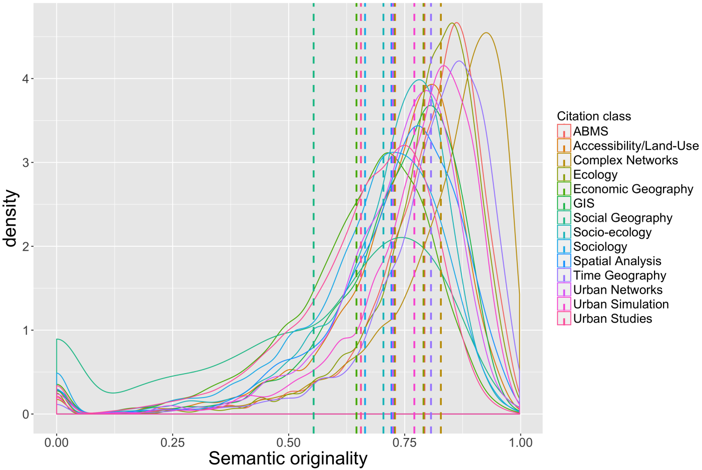


Measuring interdisciplinarity



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

Measuring interdisciplinarity



Distribution of semantic originalities

Discussion

Developments

$\rightarrow$ Journal dynamics and benchmarking, reflexivity for authors; fostering Open Science [Raimbault et al., 2019]

$\rightarrow$ Performance of the semantic classification for citation link prediction

$\rightarrow$ Correspondence of terms between disciplines

Applications

$\rightarrow$ Quantification of Domains of Knowledge [Raimbault, 2017]

$\rightarrow$ Complementary dimensions in the structure of science

$\rightarrow$ Spatial diffusion of knowledge [Raimbault, 2020b]

Multiple methods applied to the Cybergeo corpus: keywords, LDA, citation and semantic network, geographical dimension

Article

Urban Analytics and City Science

EPB: Urban Analytics and City Science
0(0) 1–16

DOI: 10.1177/2399808319870816
journals.sagepub.com/home/epb
SAGE

Empowering open science with reflexive and spatialised indicators

Juste Raimbault
UPS CNRS 3611 ISC-PIF, France; CASA, UCL, United Kingdom; UMR CNRS 8504 Géographie-cités, France

Pierre-Olivier Chasset
UMR CNRS 8504 Géographie-cités, France; LISER, Luxembourg

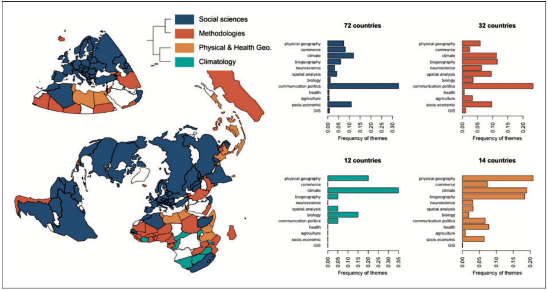
Clémentine Cottineau
UMR CNRS 8504 Géographie-cités, France; UMR CNRS 8097 Centre Maurice Halbwachs, France

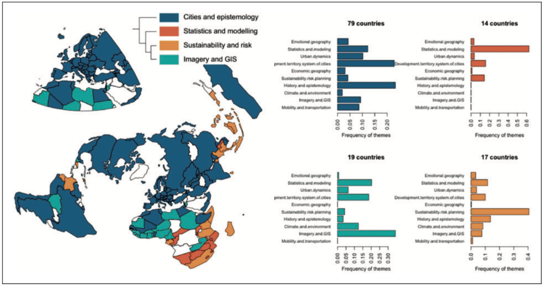
Hadrien Commenges, Denise Pumain and Christine Kosmopoulos
UMR CNRS 8504 Géographie-cités, France

Arnaud Banos
UMR IDEES 6266, France

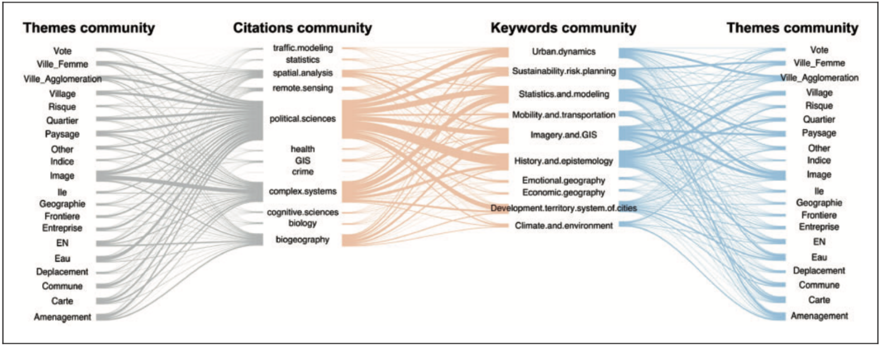
Geographical content of papers

Themes and their geographical distributions





Complementarity of methods



Interactive web application to foster reflexivity

CybergeoNetworks

analytics.huma-num.fr/geographie-cites/cybergeonetworks/

CybergeoNetworks | The Project | Overview | Citation network | Full-text Semantic network | Keyword network | Geo-semantic Networks

Cybergeo turns 20: it's time to look back for reflection and to anticipate future evolution! cybergeo.revues.org/

The editorial policy

First entirely electronic journal for social sciences in the world, peer reviewed, European, open (free of charge for authors and readers), with a focus on geography and widely open to the diversity of research agendas and methodologies in all countries. Cybergeo is a success story with now more than one million papers downloaded every year.

An app to look back

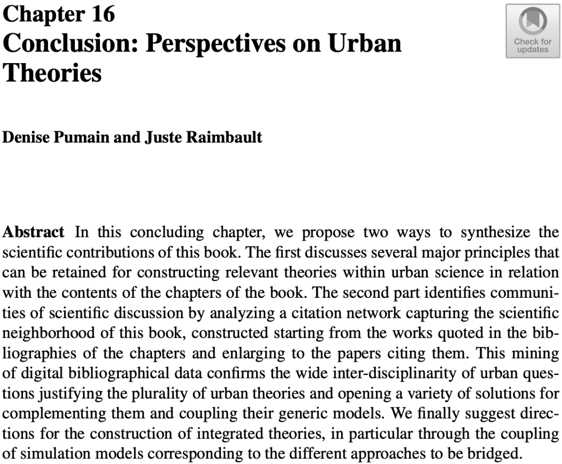
This app builds on 20 years of publication in Cybergeo. You can play with data, drawing geographical networks of authoring, studying and citing through countries, analyzing semantic networks per key words and articles' content, you can review twenty years of epistemological and thematic trends in a variety of fields of scientific interest. The networks tell who studies what, where and how. Data are regularly updated.

About the app

All data, materials and source codes are freely available on this repository: github.com/AnonymousAuthor3/cybergeo20

Multiple urban perspectives

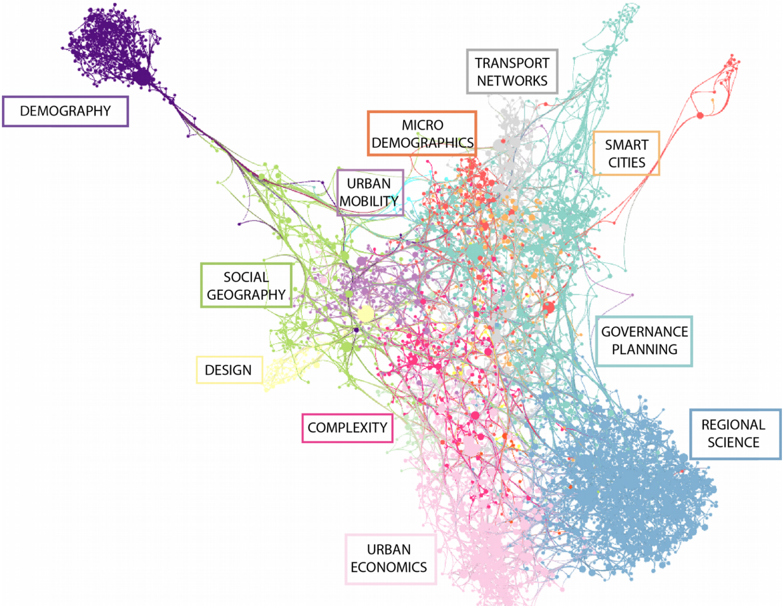
Citation network analysis of the scientific neighborhood of perspectives developed in the Geodivercity ERC book

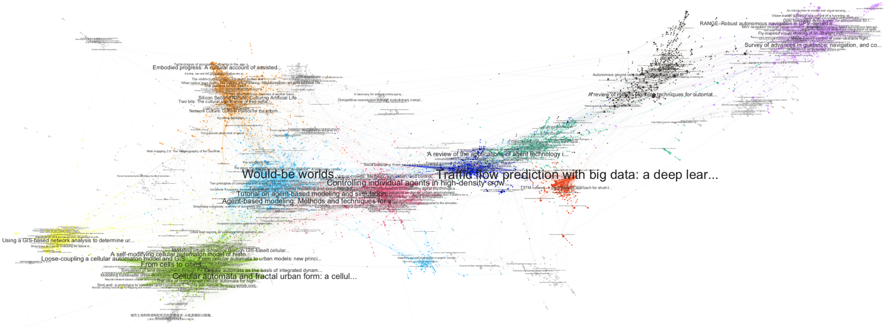

Chapter 16
Conclusion: Perspectives on Urban Theories

Denise Pumain and Juste Raimbault

Abstract In this concluding chapter, we propose two ways to synthesize the scientific contributions of this book. The first discusses several major principles that can be retained for constructing relevant theories within urban science in relation with the contents of the chapters of the book. The second part identifies communities of scientific discussion by analyzing a citation network capturing the scientific neighborhood of this book, constructed starting from the works quoted in the bibliographies of the chapters and enlarging to the papers citing them. This mining of digital bibliographical data confirms the wide inter-disciplinarity of urban questions justifying the plurality of urban theories and opening a variety of solutions for complementing them and coupling their generic models. We finally suggest directions for the construction of integrated theories, in particular through the coupling of simulation models corresponding to the different approaches to be bridged.

Citation network





Citation network of ALife studies of urban systems

Transfer of concepts: Urban morphogenesis, bio-inspired design, urban ecology, autopoiesis [Batty and Marshall, 2009]

Raimbault, J. (2020). Cities as they could be: Artificial life and urban systems. arXiv preprint arXiv:2002.12926.

Knowledge Frameworks

Knowledge Framework : *A systemic framework containing an epistemological component dealing with the nature of knowledge or knowledge production.*

$\rightarrow$ Knowledge management: [Durantin et al., 2016] coupling engineering with design paradigms; [Carlile, 2004] knowledge at the boundaries of disciplines.

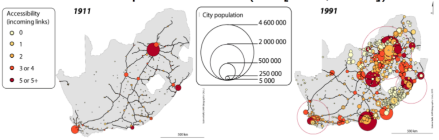
$\rightarrow$ Meta-modeling frameworks: [Cottineau et al., 2015] multi-modeling; [Golden et al., 2012] unified formal description of Complex Systems.

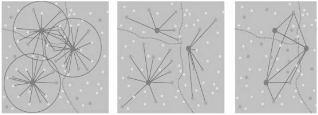
$\rightarrow$ Applied frameworks: [Moulin-Frier et al., 2017] typology of approaches in Artificial Intelligence.

Iterative Construction of Knowledge across Domains

Databases
(ex. [Swerts, 2017])

Empirical studies (ex. [Baffi, 2016])

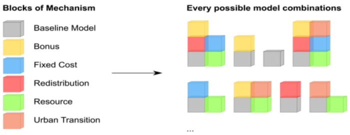




Models (ex. [Pumain, 2008])

OpenMOLE

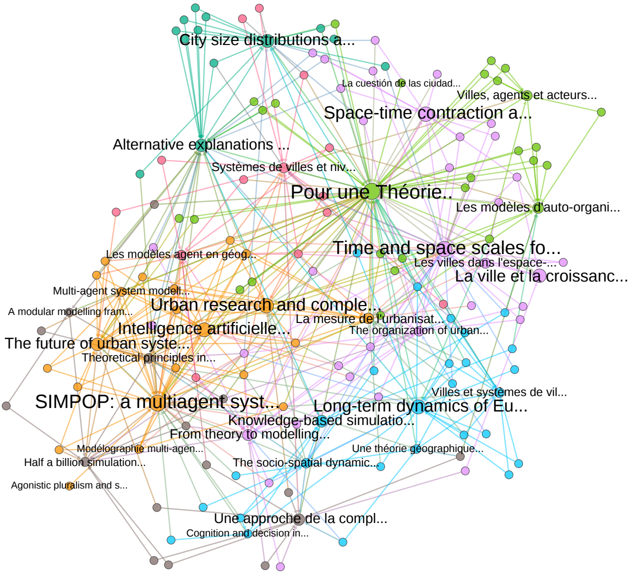
Tools (ex. [Reuillon et al., 2013])



Methods (ex. [Cottineau et al., 2015])

Theory
[Pumain, 1997]
[Pumain, 2010]

Citation Network Analysis



Core citation network of Evolutive Urban Theory

$|V|=155$

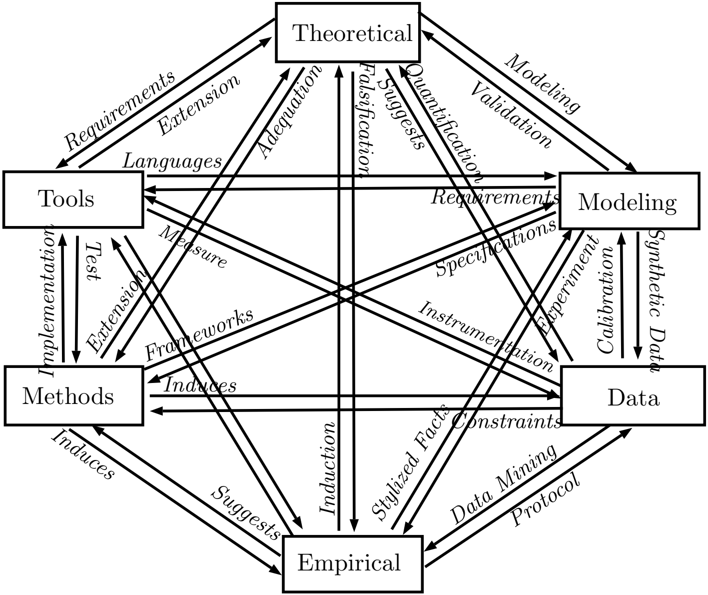
$|E|=449$

7 communities, modularity 0.39

Knowledge Domains

Definition of Knowledge Domains, extending [Livet et al., 2010]

- **Empirical.** Empirical knowledge of real world objects
- **Theoretical.** Conceptual knowledge, implying cognitive constructions
- **Modeling.** The model as the formalized *medium* of the perspective [Giere, 2010]
- **Data.** Raw information that has been collected
- **Methods.** Generic structures of knowledge production
- **Tools.** Implementation of methods and supports of others domains


Theoretical
Tools
Modeling
Methods
Data
Empirical
Requirements
Extension
Adequation
Falsification
Suggests
Quantification
Modeling
Validation
Languages
Requirements
Measure
Specifications
Implementation
Test
Extension
Frameworks
Instrumentation
Experiment
Calibration
Synthetic Data
Induces
Constraints
Induces
Suggests
Induction
Stylized Facts
Data Mining
Protocol

Towards a quantification

$\rightarrow$ Citation and semantic network analysis can *in some cases* provide insights into types of papers (data, methodological, case study, model)

$\rightarrow$ Semi-supervised approach with full-texts to quantify knowledge domains *within* papers

$\rightarrow$ Dynamics of knowledge domains (see e.g. reflexive analysis in [Raimbault, 2018]

Issues:

- Practices (and domains) specific to disciplines - transferability and comparability?
- Limitations of machine learning regarding interpretations
- Data: open access papers, sci-hub?

Conclusion

$\rightarrow$ Network analysis provides useful theoretical (knowledge production) and practical (literature mapping) insights

$\rightarrow$ Complementarity of methods

$\rightarrow$ Open tools and methodologies to foster Open Science and reflexivity

Some open repositories

Patents: `https://github.com/JusteRaimbault/PatentsMining`

Data collection: `https://github.com/JusteRaimbault/BiblioData`

Interdisc.: `https://github.com/JusteRaimbault/HyperNetwork`

Cybergeo: `https://github.com/Geographie-cites/cybergeo20`

Baffi, S. (2016).
Le chemin de fer et la ville dans les processus de territorialisation en Afrique du Sud : de la séparation à l'intégration territoriale ?
Theses, Université Paris 1 - Panthéon Sorbonne.

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.

Batty, M. and Marshall, S. (2009).
The evolution of cities: Geddes, abercrombie and the new physicalism. centenary paper.
Town Planning Review, 80(6):551–574.

References II

Bergeaud, A., Potiron, Y., and Raimbault, J. (2017).
Classifying patents based on their semantic content.
PloS one, 12(4):e0176310.

Bird, S. (2006).
Nltk: the natural language toolkit.
In *Proceedings of the COLING/ACL on Interactive presentation sessions*, pages 69–72. Association for Computational Linguistics.

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.

Bruck, P., Réthy, I., Szente, J., Tobochnik, J., and Érdi, P. (2016).
Recognition of Emerging Technology Trends. Class-selective study of citations in the U.S. Patent Citation Network.
ArXiv e-prints.

References IV

Callon, M. and Latour, B. (2013).
La science telle qu'elle se fait: anthologie de la sociologie des sciences de langue anglaise.
La découverte.

Carlile, P. R. (2004).
Transferring, translating, and transforming: An integrative framework for managing knowledge across boundaries.
Organization science, 15(5):555–568.

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.

Cottineau, C., Reuillon, R., Chapron, P., Rey-Coyrehourcq, S., and Pumain, D. (2015).
A modular modelling framework for hypotheses testing in the simulation of urbanisation.
Systems, 3(4):348–377.

Cronin, B. and Sugimoto, C. R. (2014).
Beyond bibliometrics: Harnessing multidimensional indicators of scholarly impact.
MIT Press.

References VI

De Mey, M. (2012).
The cognitive paradigm: Cognitive science, a newly explored approach to the study of cognition applied in an analysis of science and scientific knowledge, volume 1.
Springer Science & Business Media.

Dupuy, G. and Benguigui, L. G. (2015).
Sciences urbaines: interdisciplinarités passive, naïve, transitive, offensive.
Métropoles, (16).

Durantin, A., Fanmuy, G., Miet, S., and Pegon, V. (2016).
Disruptive innovation in complex systems.
In *International Conference on Complex Systems Design & Management*, pages 41–56. Springer.

References VII

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.

Fortunato, S., Bergstrom, C. T., Börner, K., Evans, J. A., Helbing, D., Milojević, S., Petersen, A. M., Radicchi, F., Sinatra, R., Uzzi, B., et al. (2018).
Science of science.
Science, 359(6379).

Gaumont, N., Panahi, M., and Chavalarias, D. (2017).
Methods for the reconstruction of the socio-semantic dynamics of political activist twitter networks.

References VIII

Gerken, J. M. and Moehrle, M. G. (2012).
A new instrument for technology monitoring: novelty in patents measured by semantic patent analysis.
Scientometrics, 91(3):645–670.

Giere, R. N. (2010).
Scientific perspectivism.
University of Chicago Press.

Golden, B., Aiguier, M., and Krob, D. (2012).
Modeling of complex systems ii: A minimalist and unified semantics for heterogeneous integrated systems.
Applied Mathematics and Computation, 218(16):8039–8055.

Griliches, Z. (1998).
Patent statistics as economic indicators: a survey.
In *R&D and productivity: the econometric evidence*, pages 287–343.
University of Chicago Press.

Grupp, H. et al. (1998).
Foundations of the economics of innovation.
Books.

Hacking, I. (1999).
The social construction of what?
Harvard university press.

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.

- Livet, P., Müller, J. P., Phan, D., Sanders, L., and Auatabu, T. (2010). Ontology, a mediator for agent-based modeling in social science.

- Mingers, J. and Leydesdorff, L. (2015). A review of theory and practice in scientometrics. *European journal of operational research*, 246(1):1–19.

- Moulin-Frier, C., Puigbò, J.-Y., Arsiwalla, X. D., Sanchez-Fibla, M., and Verschure, P. F. M. J. (2017). Embodied artificial intelligence through distributed adaptive control: An integrated framework. *ArXiv e-prints*.

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.

Nicosia, V., Mangioni, G., Carchiolo, V., and Malgeri, M. (2009).
Extending the definition of modularity to directed graphs with overlapping communities.
Journal of Statistical Mechanics: Theory and Experiment, 2009(03):P03024.

Pumain, D. (1997).
Pour une théorie évolutive des villes.
L'Espace géographique, pages 119–134.

Pumain, D. (2008).
The socio-spatial dynamics of systems of cities and innovation processes: a multi-level model.
In *The Dynamics of Complex Urban Systems*, pages 373–389. Springer.

Pumain, D. (2010).
Une théorie géographique des villes.
Bulletin de la Société géographie de Liège, (55):5–15.

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. (2018).
Caractérisation et modélisation de la co-évolution des réseaux de transport et des territoires.
PhD thesis, Université Paris 7 Denis Diderot.

Raimbault, J. (2020a).
Cities as they could be: Artificial life and urban systems.
arXiv preprint arXiv:2002.12926.

References XV

Raimbault, J. (2020b).
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.

Raimbault, J., Chasset, P.-O., Cottineau, C., Commenges, H., Pumain, D., Kosmopoulos, C., and Banos, A. (2021).
Empowering open science with reflexive and spatialised indicators.
Environment and Planning B: Urban Analytics and City Science, 48(2):298–313.

Reuillon, R., Leclaire, M., and Rey-Coyrehourcq, S. (2013).
Openmole, a workflow engine specifically tailored for the distributed exploration of simulation models.
Future Generation Computer Systems, 29(8):1981–1990.

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.

Tseng, Y.-H., Lin, C.-J., and Lin, Y.-I. (2007).
Text mining techniques for patent analysis.
Information Processing & Management, 43(5):1216–1247.

Youn, H., Strumsky, D., Bettencourt, L. M. A., and Lobo, J. (2015).
Invention as a combinatorial process: evidence from us patents.
Journal of The Royal Society Interface, 12(106).