

The evolution of the system of cities literature: fragmentation or integration?

Antoine Peris¹, Evert Meijers¹, and Maarten van Ham^{1,2}

¹OTB Research Institute for Housing, Urban and Mobility Studies,
Delft University of Technology, Delft, The Netherlands

²School of Geography and Geosciences, University of St Andrews,
St Andrews, UK

June 9, 2017

Abstract

The study of relations between cities has long been a major focus in urban research. For decades, this field has grown integrating contributions from many disciplines. But today, the field seems rather fragmented. This study aims to analyse the body of literature that has developed over the last 25 years to identify schools of thought within this literature and to see whether they relate to each other or draw their inspiration from other scientific fields. A delineation strategy is first created to collect the set of papers addressing systems of cities. Then, a semantic map is extracted from the title and abstract of papers. It reveals the methodological and empirical aspects of the different schools within the literature as well as their main objectives and ‘imported’ concepts. By coupling the semantic map with the citation networks of the publications, this study confirms the fragmentation of the field and shows three main foci: studies that develop concepts and theory, those that are empirically oriented, and those that are planning oriented. Some of them appear as very homogenous in term of vocabulary and citation patterns, some others act as bridges between the different subfields.

Keywords: system of cities, urban network, bibliometrics, text mining.

1 Introduction

Cities do not function in isolation, but are organized in systems of cities characterized by strong interdependencies that can take place at the scale of a large region, a nation, a continent or even at the global scale (Pumain, 2011). A large literature has developed around this idea of interrelated cities since the end of the 19th century. From the observation of the regularities in the size

distribution of cities in a country (Auerbach, 1913; Gibrat, 1931; Zipf, 1949) to the formulation of the central place theory (Christaller, 1933), scholars from many disciplines have long been interested in the interdependencies of cities. These early contributions were the basis of a surge of work on intercity relationships in the 1960s and 1970s, addressing many aspects of a system of cities such as the size, location and specialisation of cities as well as the uneven circulation of people, goods and information among them (Berry, 1964; Bourne and Simmons, 1978; Pred, 1977). The definition of a system of cities by Allan Pred (1977, p.13) is still valid today: “a national or regional set of cities that are interdependent in such a way that any significant change in the economic activities, occupational structure, total income or population of one member city will directly or indirectly bring about some modification in the economic activities, occupational structure, total income or population of one or more other set members”. Nowadays, this definition can also be extended to global urban systems because of long-distance interrelationships between cities at the top of national urban hierarchies becoming more common.

In the last 25 years the literature on systems of cities has developed further and expanded, but the current landscape of research seems rather fragmented. Increasingly the term paradigm change is used (Batty, 2013; Capello, 2000; Knox, 1995; Meijers, 2007) which shows the wish of some researchers to position themselves at the opposite of other research traditions in the field. Some of them refer to contributions coming from geography or regional science, some others are inspired by disciplines such as sociology, economics or the interdisciplinary movement of complexity theories. There are different positions about the scale at which the most important urban processes take place. For some researcher, in the context of globalisation, the global scale has become the most determining (Taylor and Derudder, 2015). For others, the erosion of national borders in this context put the regional scale at the centre of economic processes (Kloosterman and Musterd, 2001; Parr, 2014). At the opposite, another research stream stress the importance of the national scale which determine many structures and parameters that experience strong path dependencies (Pumain, 1997). Important differences also exist from an analytical point of view. While some studies are more focused on stakeholders (Alderson and Beckfield, 2006; Sassen, 1991), some other rather to look at the emergent properties of system of cities from the basic interactions between urban agents such as in simulation framework (Sanders et al., 1997) or the methodological individualism of economics (Fujita et al., 1999). Differences in ontological and epistemological perspectives translate into wildly varying objectives of research, ranging from identifying universal laws of urbanisation (Bettencourt et al., 2007) to much more policy oriented studies (Meijers and Romein, 2003). So it seems that within the system of cities literature separate approaches or subfields have emerged.

This paper aims to contribute to the system of cities literature by analysing the evolution of the body of literature that has developed in the last 25 years. The objective is to identify clusters within this literature and the degree to which they relate to each other or draw their inspiration from other scientific fields. Accessing interdisciplinarity in this research field is all the more important be-

cause there are frequent calls for interdisciplinarity in urban system research (Pflieger, Rozenblat, 2010), and because there is clear evidence that innovation in geography is fostered by collaborations among disciplines (Ducruet, Beauguitte, 2013). This study of the evolution of the urban systems literature does not take the form of a classical literature review paper, rather it adopts a bibliometrical approach to analyse a set of 1,258 papers on intercity relationships from 1994 onward. We combine two analyses, namely text mining techniques on the content of papers to assess their disciplinary background, and an analysis of the pattern of citations to study the network of relations between separate clusters. We follow Chavalarias and Cointet (2013) who define scientific fields as sets of keywords delineating a research area. Investigating these two different networks has been inspired by the recent development of the hyper-network approach in bibliometrics that have been applied to papers of a journal (Raimbault, 2016) or to the classification of patents (Bergeaud et al., 2017).

The following section of the paper presents the bibliometric analysis, in particular also the underlying data collection and reviews its results. In the subsequent section, we present the content-analysis based on the vocabulary of the papers, discuss its evolution and analyse the interplay between research approach and citation behaviour to see whether papers with the same vocabulary are more likely to cite each other or if they refer to contribution from other subfields. The last section concludes and discusses the implications of our findings.

2 Materials and method

2.1 Bibliometric analysis in social sciences

Citation networks not only reveal intellectual connections, but also the social organization of science (Leydesdorff, 1998). This dimension of science stands central when studying the formation of concepts but can be enriched by the analysis of the text related to the production of knowledge (Callon et al., 1983). Words have a central place in science because scientists are first of all readers and writers (Latour and Woolgar, 2013). Consequently, to study the evolution of a notion or a concept, we can in addition to exploring citation patterns also study the semantic network extracted from a set of papers. The semantic network can be obtained through text mining techniques applied to titles and abstracts of the studied papers. A major step of this research was to analyse the vocabulary within scientific publications on systems of cities, and then study its evolution to explore how this scientific field has developed. We follow Chavalarias and Cointet (2013) who worked on the evolution of scientific fields using only the vocabulary of publications. Rather than using predefined categories, this approach allows a bottom-up reconstruction of science. We used co-word analysis to identify subfields and mine their key vocabulary. This kind of analysis is used in bibliometrics to identify proximity. The basic hypothesis is that two words co-occurring often within individual papers will have a great probability to be

strongly related (Chavalarias and Cointet, 2013). After this semantic analysis, the citation network is used to see whether people using different approaches are embedded within very homogeneous clusters or if there are exchanges between different schools of thought.

The Scopus database appears to index most social science journals and was therefore chosen as point of departure for defining the body of literature to analyse. This implies that we focus on papers published after 1994, because the Scopus database does not systematically provide information preceding this date. The data have been processed with the programmes VOSviewer and R, and visualized with VOSviewer and Gephi.

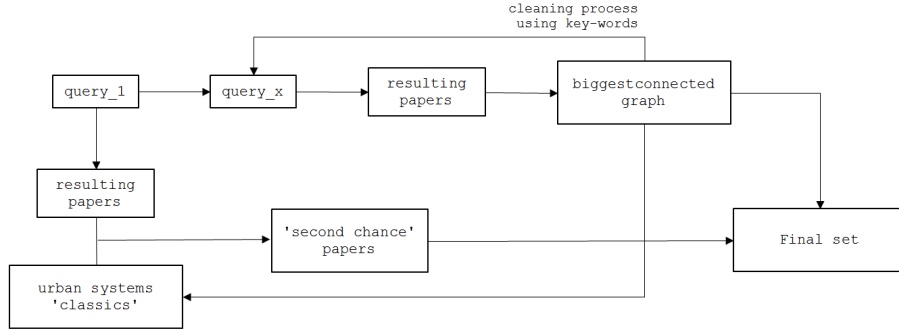
Despite the new insights on the field that this approach can bring, it has a number of limitations. First of all, we cannot claim that we are dealing with an all-encompassing, exhaustive set of papers on systems of cities for several reasons other than the less strong representation of pre-1994 papers. As this field is predominantly enriched by social scientists, many contributions, notably some of them published in books are missing. In social science, 40% of the citations relate to books rather than scientific papers (Hicks, 1999). Even if Scopus sometimes includes book chapters, we have to accept that some contributions are missing. Moreover, there is a big chance that papers published in English are overrepresented in our corpus because we did the query in English. We decided to focus on only one language to avoid the issue of translation of scientific concepts and because most of the non-English literature is indexed in Scopus with an English title and abstract (which we focus on). But obviously such a translation of abstracts into English does not always occur, which means that our corpus might be biased. Nonetheless, we do think that analysing a large corpus of publications can bring new insights on the field because it covers an important part of the scientific production in a given period. The fact that it has been collected predominantly by using a key-word strategy rather than by only climbing up or going down the chain of citations allows to avoid the teleological dimension of classical literature reviews. This approach, of course, does not replace the fundamental work of extensive readings but allows to frame the literature in a novel way.

2.2 A delineation procedure to study complex scientific subfields

The collection of the corpus of relevant publications is a very important and sensitive step because it has a strong influence on the outcomes of the analysis. There is no consensus on how to delineate a scientific field and collect related publications. Three main strategies tend to be used: the key-words strategy (Van Meeteren et al., 2015) where the set is obtained by collecting all the papers mentioning some chosen key-words; the journal-level strategy (Leydesdorff and Zhou, 2007; Liu, 2005) that supposes that specific scientific areas are covered by a limited number of journals; and, the citation-based strategy (Waltman, van Eck, 2012) which supposes that scientific fields can be conceived as clusters of individual publications citing each other. According to Zitt

(2015) “mixed strategies with learning processes, adaptive queries and multistep protocols, with possible combination of supervised and automatic stages” are welcomed in bibliometric studies and information retrieval. This is especially the case with interdisciplinary fields which are not necessarily institutionalised. For example, previous analysis of the urban studies literature have been based on mixed journal-level/key-words strategies (Kamalski and Kirby, 2012; Wang et al., 2012).

Figure 1: The algorithm of the delineation procedure



Social sciences and humanities often deal with complex notions that can have more than one signified and signifier. Expressions do not refer necessarily to a single concept and a concept can be addressed with several expression. This is the case for the notion that we are studying. The most common expressions used in the studies of a set of interdependent cities are “system of cities”, “urban system”, “city-system”, “urban network” and “city network”. Some of these expressions can be associated with underlying theories. For example, in the 1990s researchers started to criticize the strict hierarchical relations between cities conveyed by the central place model and proposed a ‘network’ model for which the relationship among cities can be conceived as horizontal and non-hierarchical (Camagni and Salone, 1993). Such associations are not systematic and most of the time, these terms are used interchangeably. Among these expressions, ‘urban system’ is the most ambiguous. In the late 1970s, Allan Pred (1977, p. 219) was already underlining the “inconsistent connotation” of the ‘urban system’ term that was used in connection to both an individual city and to a set of cities. This explains why during our first attempts of corpus delineation, a vast majority of papers were about intra-urban infrastructure networks (water, electricity, roads) and about urban metabolism (a research stream focusing on the material flow analysis of a city). For this reason, we adopted a multistep process with adaptive queries refined after each iteration and mixed with the analysis of citation patterns. The different steps are schematized in Figure 1 and are described below.

We first mined the references of the existing literature reviews (Ducruet and Lugo, 2013; Van Meeteren, 2016; Pumain, 2006) to select relevant keywords on

the subject. This first step allowed to identify a large spectrum of the field because they cover different aspects of the literature. We then added other terms based on our knowledge of the subject.

(1) *TITLE-ABS-KEY("city system*" OR "city-system*" OR "urban system*" OR "system* of cit*" OR "city network*" OR "urban network*" OR "network* of cit*" OR "settlement system*" OR "system* of settlements" OR "central place system*" OR "megaregion*" OR "polycentric urban region*" AND PUBYEAR >1994*

The first query resulted in a set of 6,079 documents. We then used the VOS viewer application (van Eck and Waltman, 2014) to visualize them. This application allowed us to transform bibliometric data into citation networks where publications are nodes and citations are edges. It returned a 'Saturn' type of configuration with a big connected graph of publications in the centre and a ring of unconnected publications surrounding. We took a random sample of 100 publications from the outer-ring to read their title and abstract. The vast majority of them were not about our subject so we decided to keep only the biggest connected graph. We then took a random sample of 100 publications in this graph and read their title and abstract. The first counting resulted in 36 relevant publications and 64 irrelevant ones. Following Milanez et al., (2016) we decided to exclude some publications with an iterative cleaning procedure that excluded papers containing certain keywords that are not relevant in our field, by adding them in the query with the AND NOT operator. After each iteration when new words were added to the query, we checked random samples of 100 papers that were part of the biggest connected graph to ensure that they were rightly kept and a similar sample of the papers that were now excluded to see whether this was fair. . When, after several iterations, we established that 95% of the papers in these samples were correctly included or dismissed, we considered our set as good enough to analyse. We then submitted the query to experts of the field for validation. The final query is the following:

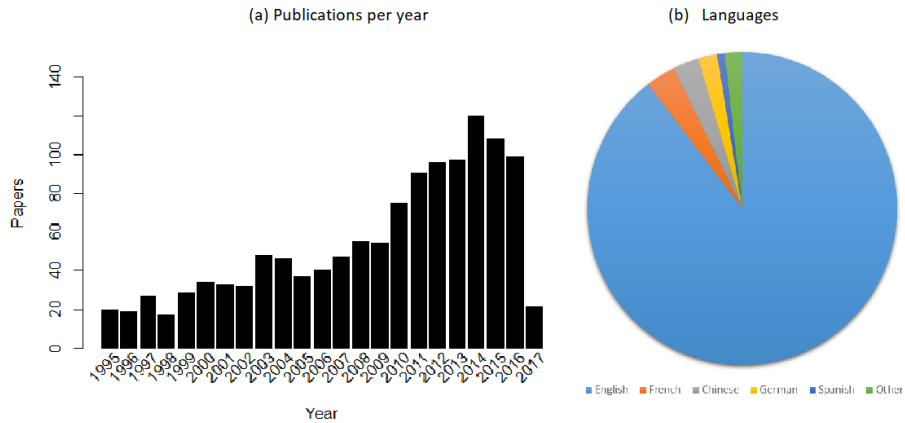
(x) *TITLE-ABS-KEY("city system*" OR "city-system*" OR "urban system*" OR "system* of cit*" OR "city network*" OR "urban network*" OR "network* of cit*" OR "settlement system*" OR "system* of settlements" OR "central place system*" OR "megaregion*" OR "polycentric urban region*" AND NOT "dispute settlement system*" AND NOT "traffic control" AND NOT "urban metabolism*" AND NOT "urban ecosystem*" AND NOT "parking*" AND NOT "smart cit*" AND NOT "urban traffic" AND NOT "space syntax" AND NOT "flood*" AND NOT "land use change*" AND NOT "urban ecology" AND NOT "hazard*" AND NOT "emergy" AND NOT "sewage" AND NOT "nitrogen" AND NOT "sensors" AND NOT "mobile landscapes" AND NOT "radial major roads" AND NOT "carbon metabolic network" AND NOT "route perception" AND NOT "waste") AND PUBYEAR >1994*

After extracting the biggest connected graph, it resulted in a set of 1,163 papers.

During the manual check of the excluded publications, we observed that some of them should have been included based on their title and abstract. Most of them were non-English language publications or publications citing the classical literature but not the current one. We then designed a ‘security net’ for these two cases. We based it on the ‘referencing structure function’ (Zitt and Bassecoulard, 2006), which is the “fraction of the literature which can be retrieved under two interplaying constraints: a minimum threshold on citation scores for the cited repertoire Y , and a minimum closeness of the article with this repertoire, measured by the number X of references in common with this repertoire.” We used $Y = 10$ for English literature and $Y = 8$ for non-English literature, considering that a non-English paper cited 8 times in the corpus is as least as important as an English paper cited 10 times. For the number of references in common, we set $X > 1$. The publications cited more than 10 times by the biggest connected graph can be considered as urban systems classics. We manually sorted them to keep only texts about systems of cities and not more general scholarly work theorising urbanization (Appendix 1).

Then we extracted from the excluded papers all the texts citing at least two of these ‘urban system classics’ (some being not that old!) to add them to the final set. This operation allowed to reintroduce 95 relevant texts into the final set, leading to a total of 1,258 publications. Figure 2 describes the set of papers and book chapters. The production of publications related to system of cities increased globally over the period, especially from 2010 onwards. The peak in 2014 seems to indicate that the production has slowed down over the last years, but can also be caused by the time lag between the publication of a paper and its indexation in the Scopus database. English language is largely dominant, its importance was probably increased by the use of an English query. French, Chinese and German are also important languages in the literature.

Figure 2: Year repartition and languages of the set of papers and book chapters



3 Results

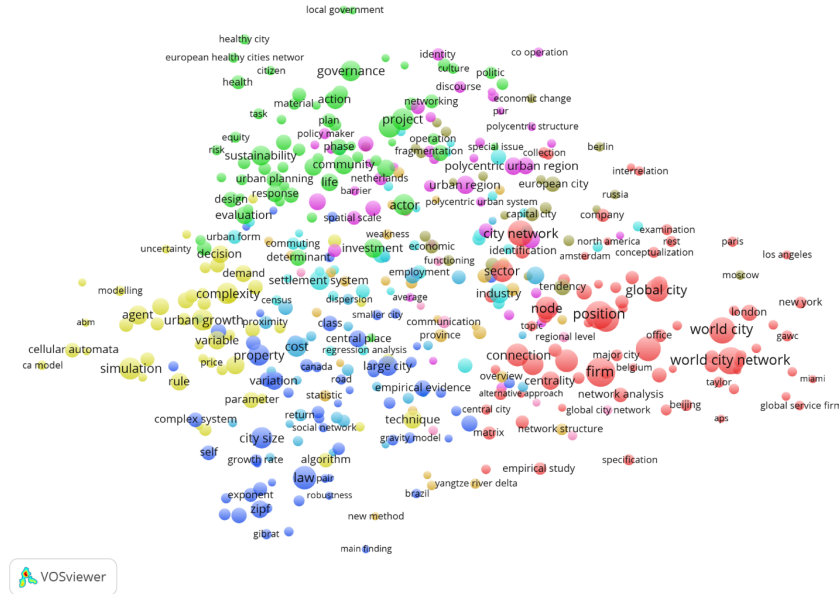
3.1 The vocabulary of urban system literature

We first did a graph of co-occurrences using VOSviewer based on the title and abstract of the 1,258 papers and book chapters in our set. This software allows us to work on noun phrases rather than simple words. It means that the words that are used systematically together will be a single node in the graph (i.e. “cellular automata”, “world city”). A phrase has to occur at least in 8 papers or book chapters to be in the graph (to keep the Figure readable). The software computes a relevance score for each noun phrase based on the frequency of its occurrence. It allows to get rid of standard wording of the scientific literature (i.e. “this paper analyses”, “interesting result”) as well as the stop words, and to keep only the ones with a specific meaning (van Eck, Waltman, 2014). We kept the 60% of the noun phrases with the highest score. For the visualisation (Figure 3), the size of the nodes corresponds to the number of occurrences of the noun phrases in the corpus. The placing of these nodes is based on their co-occurrences, with terms that co-occur often being located closer. VOS viewer also performs a clustering algorithm that allows us to visualize the most associated terms. This leads to the identification of 9 main clusters in the literature on systems of cities. The most representative phrases for each clusters can be found in Table 1, and gives a first description of what these clusters are about. The order of key-phrases in Table 1 is based on the product of the relevance score and the number of occurrences for each phrase as a proxy for its importance within the cluster.

The vocabulary used in studies of intercity relationships reveals many aspects of the different research approaches. We decided to focus especially on five aspects: the urban agents identified, the geographical scope, the methodology, thematic aspects that can be derived and the extradisciplinary dimension of each cluster. Cluster of words present variation in terms of number of nodes, from 77 for the biggest to 23 for the smallest. They are presented in decreasing order.

The first cluster corresponds to the research on worlds cities. The urban agents at the centre of this approach are private transnational firms (“firm”, “advanced producer services”, “company”). These agents are associated with the terms describing the modern capitalism (“globalization”, “finance”, “foreign direct investment”). The geographical scope of most research in this cluster clearly becomes apparent in the term “global scale” and a number of city names that correspond to major metropolises (“Beijing”, “Brussels”, “Guangzhou”, “Hong Kong”, “London”, “Los Angeles”, “Mexico City”, “Miami”, “New York”, “Paris”, “Shanghai”, “Tokyo”). The profusion of place names seems to indicate that this research trend is mainly empirically oriented rather than theoretic. From a methodological point of view, the presence of “ranking” and “position” indicates the importance of benchmarking studies as an outcome of this approach. The vocabulary of network analysis also clearly shows up with the presence of basic notions from graph theory such as “node”, “link”, “matrix”,

Figure 3: Graph of co-occurrences of the vocabulary of urban system theory



“centrality” and “connectivity”. But these words are very general and can also be used in a metaphorical manner. Graph theory has been imported by geographers and regional scientists in the 1960s (Haggett, Chorley, 1969; Kansky, 1963) and is now widely used. Consequently, it seems not appropriate to label this an ‘import’ of concepts.

A second cluster appears to be focused thematically around the notions of “governance”, “health” and “project”. The urban agents mentioned are local authority, “local government”, “community” and “planer”, which indicates that public stakeholders and collective entities are at the centre of this kind of approach. Several terms refer to countries (“Italy”, “France”, “Canada”) and to Europe, indicating that this field is not theoretical but rather oriented to empirical studies or practice. There is no clear mention of methodological terms except “evaluation” and “interview” which indicate that this cluster refers to benchmarking studies and qualitative approaches. From a thematic point, most of the words refer to social and economic well-being (“cooperation”, “culture”, “education”) as well as environmental aspects of cities (“energy”, “climate change”, “sustainability”). This cluster seems to emanate from literature on urban policies applied to group of cities.

The third cluster includes the vocabulary of the analysis of city size distributions. The inhabitants of cities are the basic element of this kind of analysis (“city population”, “city size”) which focuses mainly on the hierarchical structure of urban systems (“megacity”, “large city”, “small city”). In terms of

Table 1 : The clusters of the vocabulary of urban system theories

Cluster	Key-phrases by order of importance									
1	world city network	world city	firm	connectivity	global city	london	globalization	position	shanghai	new york
2	governance	health	european healthy cities network	healthy city	project	sustainability	evaluation	equity	life	resilience
3	law	zipf	city size distribution	city size	size distribution	power law	exponent	gibrat	self	property
4	polycentric city	polycentric urban region	urban region	spatial planning	randstad	identity	city region	netherlands	complementarity	polycentrism
5	simulation	cellular automata	agent	complex urban system	urban growth	rule	ca model	scenario	complexity	modeling
6	global economy	restructuring	sector	urban form	global urban system	housing	census	county	employment	migration
7	cost	agglomeration economy	industry	population growth	urban population	return	new economic geography	proximity	deconcentration	wage
8	network structure	hub	core	yangtze river delta	pearl river delta	amsterdam	strength	periphery	overview	new method
9	moscow	settlement system	tendency	rest	russia	capital city	european city	conception	berlin	identification

scope, these studies seem to focus on national and regional scales. One can see names of countries (“Brazil”, “Canada”, and “Mexico”) as well as the mention of “regional urban system”. Noun phrases related to methodology or theoretical frameworks are “central place”, “entropy”, “complex system”, “fractal dimension”, “gibrat”, “gravity model”, “power law”, “property”, “rank size rule”, “scaling law” and “zipf”. It clearly shows the dominance of statistical approach and mathematical formulations in the study of the size distribution of cities. The importance of the term “law” in the cluster clearly indicates the nomothetic dimension of this field of research. In this case, it is also difficult to prove a clear enrichment by other disciplines because the gravity model and Zipf’s law have long been used in geography and regional science.

The “polycentricity” and “polycentric urban region” cluster seems to mainly focus on stakeholders as the presence of “policy maker” shows. The important weight of “municipality” in the network seems to confirm this focus but we must be careful because this word can also be used to describe the municipal territory. An important part of the noun phrases of this cluster belongs to the lexical field of public action (“regional planning”, “regional development”, “synergy”, “cooperation”) which indicates that this approach is mostly policy oriented. This subfield also studies individuals by focusing on their mobility (“commuting”, “travel”) which is consistent with the planning orientation of this cluster. Intermediate scales are at the centre of the research agenda with “regional scale”, “metropolitan region” and “urban region” as well as with “Randstad”, the Dutch region where the biggest cities of the country are gathered which is the prototypical example of the polycentric urban region. The methodological basis of these studies is not revealed by our analysis.

In the cluster organised around the terms “complex urban system” and “complexity”, the basic entities studied by those employing this approach are

clearly identifiable (“agent”, “consumer”, “household”, “individual”). The interest for parameters set at the micro scale and elementary interactions between individuals is very visible in the vocabulary with terms such as “behaviour”, “choice” and “decision”. There is no mention of any city, region or country, which seems to indicate that this approach is theoretical rather than empirical. Methods and tools are at the centre of this approach as the profusion of terms related to modelling and simulation shows. One can see the two different methods of simulation: “abm” (for agent based model) and “multi agent system” on one hand, and “ca model” (for “cellular automata”, also in the graph) on the other hand. These methods are associated with a particular terminology: “rule”, “scenario”, “prediction”. All these noun phrases show clearly the interdisciplinary background of this research line, parallel to the computational turn in social science. According to Sanders (2014) the two inspirations in the simulation of systems of cities are physics, with the works on dissipative structures and synergetic, and computer science and artificial intelligence that notably created the tools.

With “global economy” and “global urban system” among its most representative terms, the sixth cluster call to mind the first one. But its scope is not limited to the global scale as this list of phrases shows: “global urban system”, “national economy”, “regional economy”; as well as the presence of terms referring to countries (“Australia”, “Japan”, “Chinese city”). The presence of “migration” and “census” seems to indicate an interest in demographic variables, but expressions such as “wealth”, “sector”, and “housing” are more related to economic phenomena. Terms related to methods are “regression analysis” and “statistic”, which are both very common in geography and regional science. Given the general dimension of this cluster, we can hypothesise that it corresponds to clusters of terms shared by the different approaches.

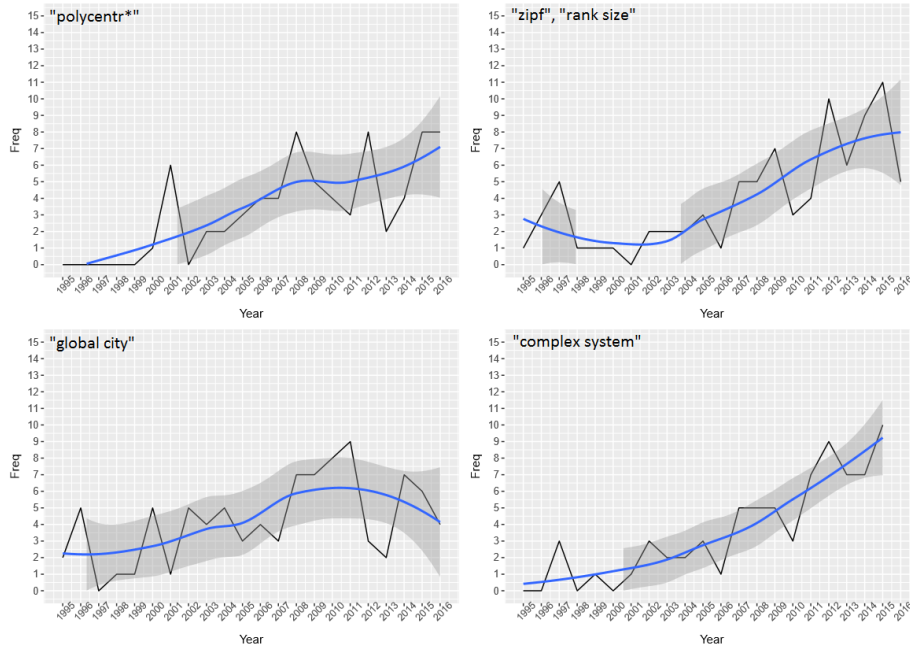
The seventh cluster can be easily entitled because of the presence of a term referring directly to a research trend: the “new economic geography”. The noun phrases “census data”, “commuter”, “worker” and “population growth” indicating the interest for the elementary entities of the urban phenomenon. As with the ‘complex urban system’ approach, there is no mention of any toponyms, which indicate the theoretical focus of this approach. Interdisciplinarity can be assessed through the cohabitation of two lexical fields, the one referring to urban form and dynamics (“large metropolitan area”, “spatial concentration”, “diffusion”, “agglomeration economies”) and the vocabulary from the mainstream economic research (“industry”, “inequality”, “wage”, “cost”, “return”).

The two last clusters gather a very limited number of terms which seems to indicate that they are related to a very small and specific field of research rather than an actual research trend.

We selected terms representative of four of the main clusters to study their dynamics over the period 1995-2016 (Figure 4). The frequency on the Y-axis represents the number of papers in which they appear and not their absolute frequency of appearance in the corpus. The general growth tendency observed in Figure 2 can be observed for most of the evolution of the use of key-phrases. This tendency does not concern all the concepts and patterns of growth can

be very different. The notion of polycentricity have experienced a first wave of interest in the beginning of the 2000s, leading to a first peak around the year 2008, at the period when the notion was embraced by many planning strategies in European countries. Since this period, the notion remains used in the literature with a rather stable pattern. In the case of the literature on the size distribution of cities, it appears that after an increase around 1996-1997, the notion witnessed a downturn which was however followed by an important increase in the second half of the 2000s. Quite opposite, the notion of “global city” was widespread in the 1990s, but seems to decrease nowadays after a peak around 2010-2011. Finally, the most prominent growth pattern among these notions that we analysed is obtained for the interdisciplinary notion of “complex system” that experienced an almost continuous growth over the past 20 years.

Figure 4: Temporal evolution of the use of some central notions in the corpus

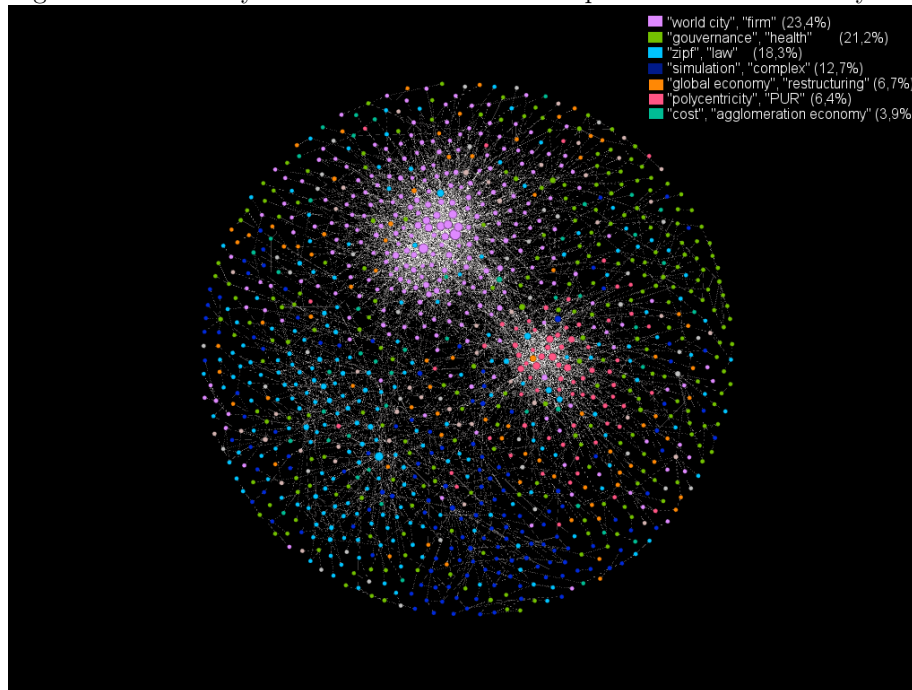


3.2 Subfields and citation patterns

To evaluate the research approach of each single publication, we designed a program based on a scoring method. For example, if a paper mentioned noun phrases from the cluster dealing with simulation 20 times, and 10 times noun phrases from the cluster about size distribution of cities, the paper will be considered as belonging to the simulation approach. We visualise the result in Figure 5. Each node corresponds to a paper or a book chapter, its colour

indicates which cluster of words (according to 3.1) is the most represented in its title and abstract, and where the links between the nodes are the citations.

Figure 5: Vocabulary and citation network of the publications on urban system



The literature on world cities and the one on polycentric urban regions are gathered into dense clusters of citations organized around the seminal papers of these subfields. These papers formulate the research agenda, new methodologies or give first empirical insights leading to the new approach. In the case of the world literature, these papers are for example those of Beaverstock, Smith, Taylor (2000), Taylor (2001) and of Alderson and Beckfield (2004). And for the polycentric urban region research, the most central papers are those of Batten (1995), Kloosterman and Musterd (2001) and Parr (2004). These two clusters are related to each other. Some of the publications about city size distribution and scaling patterns of cities manifest a slight community structure around the paper of Bettencourt et al. (2007) but nothing comparable with the subfields mentioned above. The governance literature is quite spread and does not show strong coherence.

In terms of interdisciplinarity, the world city cluster appears as self-centred with very few references to other approaches. The cluster of publications on polycentric urban regions is very dense, with many contributions originating from other approaches, such as the city size distribution and the governance subfield. The other subfields are not self-centred and are more likely to cite other approaches. The two clusters on size distribution of cities and simulation

are well connected as well as the size distribution approach and the economic one.

4 Conclusion

The literature addressing relationships between cities has become rather fragmented both in terms of vocabulary and citation patterns. The analysis of the vocabulary in the papers belonging to this literature allowed us to reveal nine clusters or subfields in this literature. These can be grouped according to three main foci in the literature: the studies that are focusing on theory, the ones that are empirical oriented, and the ones that are (planning) policy oriented. Two research clusters belong to the latter: the group of publications about governance and the one about polycentric urban regions. In the vocabulary of these two approaches, public organisation and stakeholders clearly show up. The cluster corresponding to the simulation approach manifests exclusively an interest in theory and methodology while the world city literature appears as empirical oriented with a profusion of city names. Studies on city-size distributions manifest a hybrid character, with many phrases referring to mathematical modelling and at the same time several country names. In terms of citation patterns this hybrid character becomes evident in the relatively greater abundance of connections to other subfields such as the simulation approach or the more economic cluster. Publications on world cities and polycentric urban regions form very dense clusters, meaning that they are strongly interwoven yet distinct. The former is very self-centred, the latter builds much more on contributions from other subfields. In terms of interdisciplinarity the integration of concepts and methods from complexity sciences is clearly visible in the cluster about city-size distributions, which mention scaling patterns, as well as in the cluster about simulation given the references to cellular automata and agent based models. The vocabulary of mainstream economics also appears obviously in the cluster representing the New Economic Geography.

The bibliometric method developed in this paper has proved its capacity to study a set of publications by exploring both their content and their citation behaviour. First, the multi-step delineation procedure with adaptive queries and learning process allowed to collect a set of papers and book chapters with a minimum of noise. Then, the mapping of key-phrases has revealed effectively the different schools of thought working on urban systems. Studying the vocabulary of papers allowed us to assess the methodology, scope, thematic favoured by each of the approaches and whether they are mostly theoretical or empirical.

Further research will be undertaken to refine the tagging process of papers by taking into account the relevance score of the key-phrases, presumably making the results even more accurate. The outcome of that would. Moreover, the temporal evolution of the size of clusters, their citation patterns and relations with other clusters will be further investigated to see which literature is growing and declining over the period, and whether this is related to the extent to which they are interdisciplinary by developing relations with other subfields.

Acknowledgement

The authors thank Bijan Ranjbar-Sahraei from the AIDA project for his valuable advice on the bibliometric method. This work has benefited from a VIDI grant provided by the Netherlands Organisation for Scientific Research (NWO).

Bibliography

Alderson A., Beckfield J., 2006, Globalization and the World City System: Preliminary Results from a Longitudinal Data Set , in: Cities in Globalization, Practices, Policies and Theories. Peter Taylor, Ben Derudder, Pieter Saey, and Frank Witlox.

Alderson A. S., Beckfield J., 2004, Power and Position in the World City System , American Journal of Sociology, vol.109, num.4, 811-851.

Auerbach F., 1913, Das Gesetz der Bevlkerungskonzentration. , Petermanns Geographische Mitteilungen, vol.59, 7476.

Batten D. F., 1995, Network Cities: Creative Urban Agglomerations for the 21st Century , Urban Studies, vol.32, num.2, 313-327.

Batty M., 2013, The New Science of Cities. MIT Press, 519 p .

Beaverstock J. V., Smith R. G., Taylor P. J., 2000, World-City Network: A New Metageography? , Annals of the Association of American Geographers, vol.90, num.1, 123-134.

Bergeaud A., Potiron Y., Raimbault J., 2017, Classifying patents based on their semantic content , PLOS ONE, vol.12, num.4, e0176310.

Berry B. J. L., 1964, Cities as Systems Within Systems of Cities , Papers in Regional Science, vol.13, num.1, 147-163.

Bettencourt L. M. A., Lobo J., Helbing D., Khnert C., West G. B., 2007, Growth, innovation, scaling, and the pace of life in cities , Proceedings of the National Academy of Sciences, vol.104, num.17, 7301-7306.

Bourne L. S., Simmons J. W., 1978, Systems of cities: readings on structure, growth, and policy. Oxford University Press, 596 p .

Callon M., Courtial J.-P., Turner W. A., Bauin S., 1983, From translations to problematic networks: An introduction to co-word analysis , Social Science Information, vol.22, num.2, 191-235.

Camagni R. P., Salone C., 1993, Network Urban Structures in Northern Italy: Elements for a Theoretical Framework , Urban Studies, vol.30, num.6, 1053-1064.

Capello R., 2000, The City Network Paradigm: Measuring Urban Network Externalities , Urban Studies, vol.37, num.11, 1925-1945.

Chavalarias D., Cointet J.-P., 2013, Phylomemetic Patterns in Science EvolutionThe Rise and Fall of Scientific Fields , PLOS ONE, vol.8, num.2, e54847.

Christaller W., 1933, Die zentralen Orte in Sddeutschland: Eine konomisch-geographische Untersuchung ber die Gesetzmssigkeit der Verbreitung und Entwicklung der Siedlungen mit stdtischen Funktionen. University Microfilms, 352 p .

Ducruet C., Beauguitte L., 2013, Spatial Science and Network Science: Review and Outcomes of a Complex Relationship , Networks and Spatial Economics, vol.14, num.3-4, 297316.

Ducruet C., Lugo I., 2013, Cities and Transport Networks in Shipping and Logistics Research , The Asian Journal of Shipping and Logistics, vol.29, num.2, 145-166.

van Eck N. J., Waltman L., 2014, "Visualizing Bibliometric Networks", 285320in: Y. Ding, R. Rousseau, and D. Wolfram (ed.), Measuring Scholarly Impact. Cham, Springer International Publishing.

Fujita M., Krugman P., Mori T., 1999, On the evolution of hierarchical urban systems1 , European Economic Review, vol.43, num.2, 209251.

Gibrat R., 1931, Les inegalits conomiques. Recueil Sirey, 296 p .

Haggett P., Chorley R. J., 1969, Network Analysis in Geography. Edward Arnold, 368 p .

Hicks D., 1999, The difficulty of achieving full coverage of international social science literature and the bibliometric consequences , Scientometrics, vol.44, num.2, 193-215.

Kamalski J., Kirby A., 2012, Bibliometrics and urban knowledge transfer , Cities, vol.29, Supplement 2, S3-S8.

Kansky K. J., 1963, Structure of transportation networks: relationships between network geometry and regional characteristics. University of Chicago., 186 p .

Kloosterman R. C., Musterd S., 2001, The Polycentric Urban Region: Towards a Research Agenda , Urban Studies, vol.38, num.4, 623-633.

Knox P. L., 1995, World Cities in a World-System. Cambridge University Press, 356 p .

Latour B., Woolgar S., 2013, Laboratory Life: The Construction of Scientific Facts. Princeton University Press, 295 p .

Leydesdorff L., 1998, Theories of citation? , Scientometrics, vol.43, num.1, 5-25.

Leydesdorff L., Zhou P., 2007, Nanotechnology as a field of science: Its delineation in terms of journals and patents , Scientometrics, vol.70, num.3, 693-713.

Liu Z., 2005, Visualizing the intellectual structure in urban studies: A journal co-citation analysis (1992-2002) , Scientometrics, vol.62, num.3, 385402.

van Meeteren M., 2016, From polycentricity to renovated urban systems theory: explaining Belgian settlement geographies. Ghent University.

van Meeteren M., Poorthuis A., Derudder B., Witlox F., 2015, Pacifying Babels Tower: A scientometric analysis of polycentricity in urban research , Urban Studies.

Meijers E., 2007, From central place to network model: theory and evidence of a paradigm change , Tijdschrift voor economische en sociale geografie, vol.98, num.2, 245259.

Meijers E., Romein A., 2003, Realizing Potential: Building Regional Organizing Capacity in Polycentric Urban Regions , European Urban and Regional Studies, vol.10, num.2, 173-186.

- Milanez D. H., Noyons E., Faria L. I. L. de., 2016, A delineating procedure to retrieve relevant publication data in research areas: the case of nanocellulose , *Scientometrics*, vol.107, num.2, 627-643.
- Parr J., 2004, The Polycentric Urban Region: A Closer Inspection , *Regional Studies*, vol.38, num.3, 231-240.
- Parr J. B., 2014, The Regional Economy, Spatial Structure and Regional Urban Systems , *Regional Studies*, vol.48, num.12, 1926-1938.
- Pflieger G., Rozenblat C., 2010, Introduction. Urban Networks and Network Theory: The City as the Connector of Multiple Networks , *Urban Studies*, vol.47, num.13, 2723-2735.
- Pred A., 1977, City Systems in Advanced Economies: Past Growth, Present Processes, and Future Development Options. Wiley, 256 p .
- Pumain D., 1997, Pour une thorie évolutive des villes , *Espace géographique*, vol.26, num.2, 119-134.
- Pumain D., 2006, Hierarchy in natural and social sciences. Dordrecht, The Netherlands.
- Pumain D., 2011, "Systems of Cities and Levels of Organisation" 225-249 in: P. Bourguin and A. Lesne (d.), *Morphogenesis*. Springer Berlin Heidelberg.
- Raimbault J., 2016, Indirect Bibliometrics by Complex Network Analysis <https://halshs.archives-ouvertes.fr/halshs-01328058>
- Sanders L., 2014, Trois décennies de modélisation des systèmes de villes: sources d'inspiration, concepts, formalisations, Three decades of modeling systems of cities: sources of inspiration, concepts, formalization , *Revue d'économie Régionale et Urbaine*, vol.décembre, num.5, 833-856.
- Sanders L., Pumain D., Mathian H., Gurin-Pace F., Bura S., 1997, SIM-POP: A Multiagent System for the Study of Urbanism , *Environment and Planning B: Planning and Design*, vol.24, num.2, 287-305.
- Sassen S., 1991, *The Global City*: New York, London, Tokyo. Princeton University Press, 397 p .
- Taylor P. J., 2001, Specification of the world city network , *Geographical analysis*, vol.33, num.2, 181-194.
- Taylor P. J., Derudder B., 2015, *World City Network: A Global Urban Analysis*. Routledge, 251 p .
- Waltman L., van Eck N. J., 2012, A new methodology for constructing a publication-level classification system of science , *Journal of the American Society for Information Science and Technology*, vol.63, num.12, 2378-2392.
- Wang H., He Q., Liu X., Zhuang Y., Hong S., 2012, Global urbanization research from 1991 to 2009: A systematic research review , *Landscape and Urban Planning*, vol.104, num.34, 299-309.
- Zipf G. K., 1949, *Human behavior and the principle of least effort: an introduction to human ecology*. Addison-Wesley Press, 596 p .
- Zitt M., 2015, Meso-level retrieval: IR-bibliometrics interplay and hybrid citation-words methods in scientific fields delineation , *Scientometrics*, vol.102, num.3, 2223-2245.
- Zitt M., Bassecoulard E., 2006, Delineating complex scientific fields by an hybrid lexical-citation method: An application to nanosciences , *Information*

Processing and Management, vol.42, num.6, 1513-1531.

Appendix 1 : The ‘urban system classics’

1	Alderson, Beckfield, 2004	Power and Position in the World City System
2	Berry, 1964	Cities as Systems Within Systems of Cities
3	Bettencourt, 2013	The Origins of Scaling in Cities
4	Camagni, Salone, 1993	Network Urban Structures in Northern Italy: Elements for a Theoretical Framework
5	Capello, 2000	The City Network Paradigm: Measuring Urban Network Externalities
6	Castells, 1992	The Informational City: Economic Restructuring and Urban Development
7	Cattan, Rozenblat, Pumain, Saint-Julien, 1994	Le système des villes européennes
8	Christaller, 1933, 1966	Die zentralen Orte in Süddeutschland: Eine ökonomisch-geographische Untersuchung über die Gesetzmässigkeit der Verbreitung und Entwicklung der Siedlungen mit städtischen Funktionen
9	Corrêa, 2006	Estudos sobre a rede urbana
10	Dematteis, Bonavero, 1997	Il sistema urbano italiano nello spazio unificato europeo
11	Derudder, Taylor, Witlox, Catalano, 2003	Hierarchical Tendencies and Regional Patterns in the World City Network: A Global Urban Analysis of 234 Cities
12	Friedmann, 1986	The World City Hypothesis
13	Gabaix, 1999	Zipf’s Law for Cities: An Explanation
14	Guerin-Pace, 1995	Rank-Size Distribution and the Process of Urban Growth
15	Hall, Pain, 2006	The Polycentric Metropolis: Learning from Mega-City Regions in Europe
16	Harris, Ullman, 1945	The Nature of Cities
17	Henderson, 1974	The Sizes and Types of Cities
18	Kloosterman, Musterd, 2001	The Polycentric Urban Region: Towards a Research Agenda
19	Krugman, 1990	Increasing returns and economic geography
20	Parr, 1985	A note on the size distribution of cities over time
21	Pumain, 1982	La Dynamique des villes
22	Pumain, Moriconi-Ebrard, 1997	City size distributions and metropolisation
23	Rosen, Resnick, 1979	The Size Distribution of Cities: An Examination of the Pareto Law and Primacy
24	Sassen, 1991	The Global City: New York, London, Tokyo
25	Taylor, 2001	Specification of the world city network
26	Veltz, 1996	Mondialisation, villes et territoires: une économie d’archipel
27	Zipf, 1949	Human behavior and the principle of least effort: an introduction to human ecology