Models Coupling Urban Growth and Transportation Network Growth : An Algorithmic Systematic Review Approach

Juste Raimbault

UMR CNRS 8504 - Géographie-Cités, Paris, France

UMR-T IFSTTAR 9403 - LVMT, Champs-sur-Marne, France

Abstract

Taking as an object of study the co-evolution of land-use and transportation infrastructure, a broad bibliographical study suggests the quasi-absence of quantitatve models of simulation that would integrate both network and urban growth. This absence may be due to scientific disciplines involved that do not easily interact and are rather self-centered. We propose an algorithmic systematic review to give quantitative elements of answer to his question. A formal algorithm to retrieve corpuses of references from initial keywords, based on text-mining, is developed and implemented. We study its convergence prperties and do a sensitivity analysis. We then apply it on queries representing typically of our question, and the study of retrieven corpuses tends to confirm the hypothesis.

1 - Introduction

Transportation networks and urban land-use are known to be strongly coupled components of urban systems at many scales [Bretagnolle, 2009] . One common approach is to consider them as co-evolving rather than misleading interpretations such as the dangerous *myth of structuring effect of transportation [Offner, 1993]* . A question rapidly arising is the existence of models that endogeneize this co-evolution, taking into account simultaneously urban and network growth. We try to answer it using an algorithmic systematic review. The rest of the paper is organized as follows : after a brief state of the art of existing literature, we present the approach and formalize the algorithm, which results are then presented and discussed.

1-1 Modeling Interactions between Urban Growth and Network Growth : State of the Art

Land-Use Transportation Interaction Models

A wide class of models that have been developed essentially for planning purposes, which are the so-called *Land-use Transportation Interaction Models*, is a first type answering our research question. See recent reviews [Chang, 2006, Iacono et al., 2008, Wegener and Furst, 2004] to get an idea of the heterogeneity of included approaches, that exist for more than 30 years ago. Recent models with diverse refinments are still developed today, such as [Delons et al., 2008] which includes housing market for Paris area. Diverse aspects of the same system can be translated into many models (see e.g. [Wegener et al., 1991]), and traffic, residential and employment dynamics, resulting land-use evolution, influenced also by a static transportation network, are generally taken into account.

Network Growth Approaches

On the contrary, many economic literature has done the opposite of previous models, i.e. trying to reproduce network growth given assumptions on the urban landscape, as reviewed in [Zhang and Levinson, 2007] . In [Xie and Levinson, 2009] , economic empirical studies are positioned within other network growth approaches, such as work by physicists giving model of geometrical network growth [Barth elemy and Flammini, 2008].. Analogy with biological networks was also done, reproducing typical robustness properties of transportation networks [Tero et al., 2010] .

Hybrid Approaches

Very few approaches coupling urban growth and network growth can be found in the literature. [Barth elemy and Flammini, 2009] couple density evolution with network growth in a toy model, obtaining theoretical results. In [Raimbault et al., 2014] , a simple Cellular Automaton coupled with an evolutive network reproduces stylized facts of *human settlements* described by LE CORBUSIER. At a smaller scale, [Achibet et al., 2014] propose a model of co-evolution between roads and buildings, following geometrical rules. These approaches stay however limited and exceptional.

1-2 Bibliometrical Approach

Literature review is a crucial preliminary step for any scientific work and its quality and extent may have a dramatic impact on perspectives for research question and objectives, and systematic review techniques have been developed, from qualitative review techniques to quantitative meta-analyses allowing to produce new results by combining existing studies[Rucker, 2012] . Missing some references can be considered as a scientific mistake in the context of emerging information systems[Lissack, 2013]. Using a systematic technique should be an advantage to tackle our issue.

Indeed, observing the form of the bibliography obtained in previous section raises some hypothesis. We see that all components are present for co-evolutive models to exist but that a lack of communication seems to impeed it. As it was shown in [Commenges, 2013] for the concept of mobility, for which a “small world of actors” relatively closed invented a notion *ad hoc*, using models without accurate knowledge of a more general scientific context, we could be in an analog case for the type of models we are interested in. Restricted interactions between scientific fields working on the same objects but with different purposes, backgrounds and at different scales, could be at the origin of the relative absence of co-evolving models. We propose an algorithmic method described in the following to enlight this issue.

2 - Description of the Algorithm

2-1 Paradigm

While classical bibliometrical studies are based on citation networks [Newman, 2013] or co-autorship networks [Sarigo l et al., 2014] we propose to use a novel paradigm based on text-processing introduced in [Chavalarias and Cointet, 2013] . It was used to retrieve a cartography of scientific disciplines based on the content of their study, excluding possible partitioning due to socially separated domains. For our question, it has a particular interest, as we want to understand the structure of researches on the subject. The algorithm proceeds by iterations to obtain a stabilized corpus from initial keywords.

2-2 Formalization

Let A be an alphabet, A^* corresponding words and $T = \bigcup_{k \in \mathbb{N}} A^{*k}$ texts of finite length on it. A reference is for the algorithm a record with text fields representing title, abstract and keywords. Set of references at iteration *n* will be denoted $C_n \subset T^3$. We assume the existence of a set of keywords K_n, initial keywords being K_0. An iteration goes as follows :

- A raw intermediate corpus R_n is obtained through a catalog request providing previous keywords K_{n-1}.
- Overall corpus is actualized by $C_n = C_{n-1} \cup R_n$
- New keywords K_n are extracted from corpus through Natural Language Processing treatment, given a parameter N_k fixing the number of keywords.

The algorithm stops when cardinal of corpus becomes stable or a user-defined maximal number of iterations has been reached. Fig. 1 shows the global workflow.

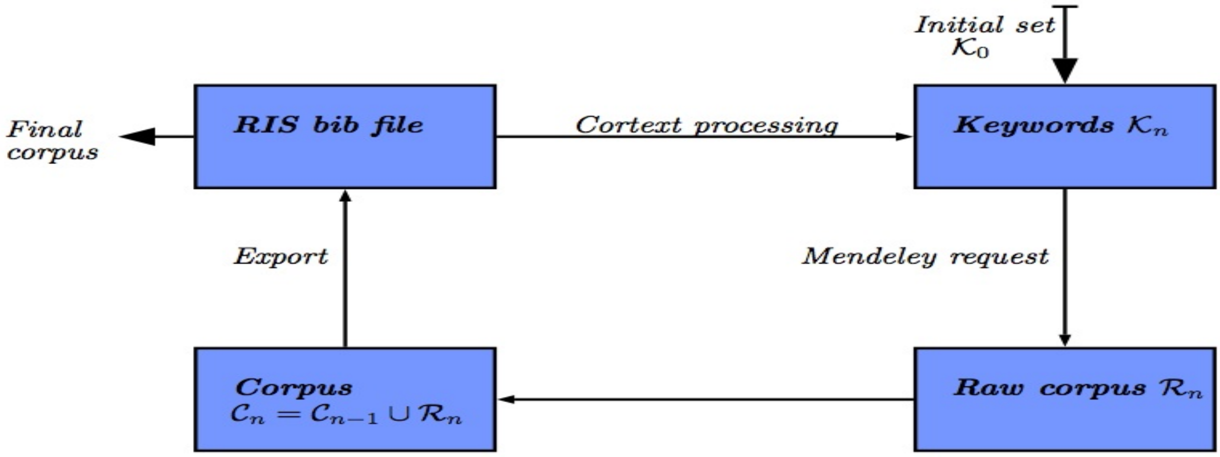

Figure 1: Global workflow of the algorithm. We provide also implementation details : catalog request is done through Mendeley API ; final state of corpuses are RIS files.

3 - Results

3-1 Implementation

General Implementation

Because of the heterogeneity of operations required by the algorithm (references organisation, catalog requests, text procesing), it was found a reasonable choice to implement it in Java. Source code and binaries are available on the Github repository of the project[1].

Catalog Requests

Given a set of keywords, we need to extract a corpus of articles from a bibliographical database. It has to be done automatically and records must have abstract record populated as text mining is mostly done on it. Therefore, it was chosen to use Mendeley reference manager software [Mendeley,

1 at `https://github.com/JusteRaimbault/CityNetwork/tree/master/Models/Biblio/AlgoSR/AlgoSRJavaApp`

2015] as it provide an open access to a flexible API that allows such catalog requests.

Natural Language Processing

Keyword extraction is done through Natural Language Processing (NLP) techniques, following the workflow given in [Chavalarias and Cointet, 2013] . Although powerful and flexible libraries exist for current operations[2], the elaborated workflow of the paper would be painful to implement and is furthermore already made available by the authors on the dedicated website of the *CorText* project[3], plateform with which we interacted automatically for the language processing step.

3-2 Convergence and Sensitivity Analysis

It is not possible to formally show the convergence of the algorithm as it will depend on the unknown structure of request results and keywords extraction. We need thus to study empirically its convergence. For different initial request and many values of parameter N_k, we studied the number of references as a function of the iteration. Good convergence properties but various sensitivities to N_k were found as presented in Fig. 2. We also studied the internal lexical consistence of final corpuses as a function of leyword number. As expected, small number gave mor consistent corpuses, but the variability when increasing was reasonable, what gives a clue on the relevance of the approach that retrieves always stable corpuses.

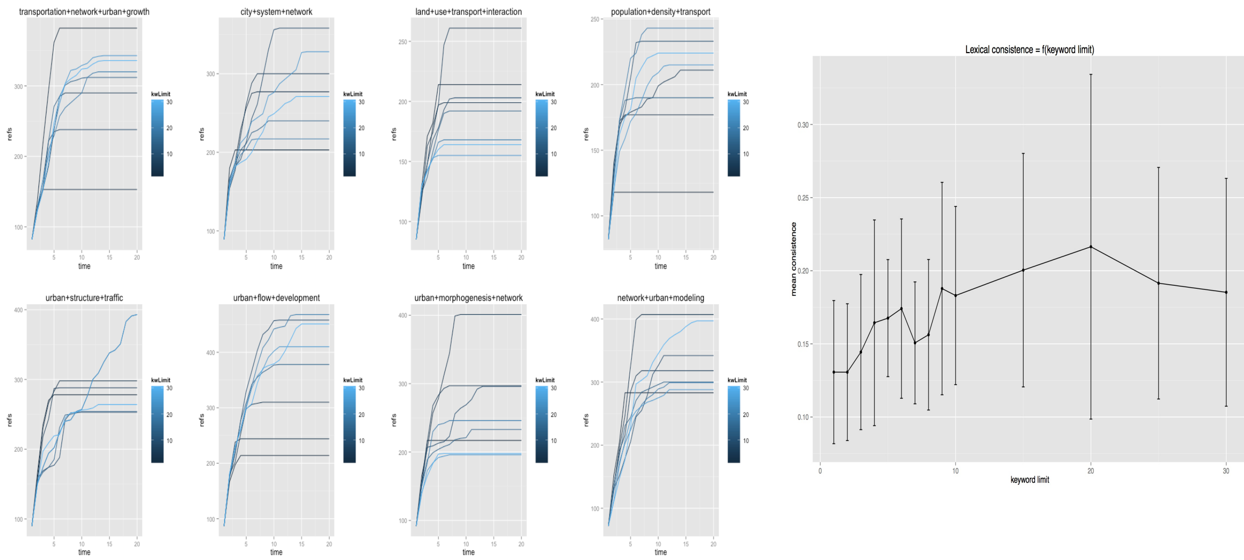

Fig. 2 : Convergence and sensitivity analysis. Left : Plots of number of references as a function of iteration, for various queries linked to our thema, for various values of N_k (from 2 to 30). We obtain a rapid convergence for most cases, around 10 iterations needed. Final number of references appears to be very sensitive to keyword number depending on queries, what seems logical since encountered landscape should strongly vary depending on terms. *Right :* Mean lexical consistence and standard error bars for various queries, as a function of keyword number. Lexical consistence is defined though co-occurrences of keywords by, with N final number of keywords, f final step, and $c(i)$ co-occurrences in references, k = 2/N(N-1) \sum_{i,j \in \mathcal{K}_f}{|c(i)-c(j)|}$. The stability confirms the consistence of retriven corpuses.

3-3 Application

Once the algorithm is validated, we can apply it to our question. We start from 5 different initial

2 see e.g. Java library by The Stanford Natural Language Processing Group at `http://nlp.stanford.edu/software/corenlp.shtml`, or Python library NLTK at `http://www.nltk.org/`.

3 `http://manager.cortext.net`

requests that were manually extracted from the various domains identified in the manual bibliography (that are “city system network”, “land use transport interaction”, “network urban modeling”, “population density transport”, “transportation network urban growth”). We take the weakest assumption on parameter $N_k = 30$, as it should contrain less domains explorations and increase consistence of final results.

After having retrieven corpuses, we study their lexical distances as an indicator to answer our initial question. If they are indeed very far one from eachother, it goes in the direction of the assumption made in section 2, i.e. that discipline self-centering may be at the origin of the non-existence of co-evolutive models. We show in Table 1 values of relative lexical proximity, that appear to be significantly low, what confirms our assumption.

Corpuses	1	2	3	4	5
1	-	0.00	0.03	0.01	0.03
2	0.00	-	0.04	0.00	0.02
3	0.03	0.04	-	0.01	0.07
4	0.01	0.00	0.01	-	0.05
5	0.03	0.02	0.07	0.05	-

Table 1 : Symmetric matrix of lexical proximities between final corpuses, defined as the sum of overall final keywords co-occuring between corpuses, normalized by number of final keywords (100). We obtain very low values, confirming that corpuses are significantly far.

Further work is planned towards the construction of citation networks through an automatic access to Google Scholar that provides backward citations. The confrontation of inter-cluster coefficients on the citation network for the different corpuses with our lexical consistence results are an essential aspect of a further validation of our results.

Conclusion

The dramatic absence of models simulating the co-evolution of transportation networks and urban land-use, confirmed through a state-of-the-art covering many domain, may be due to the absence of communication between scientific disciplines studying different aspects of that problems. We have proposed an algorithmic method to give elements of answers through text-mining-based corpus extraction. First numerical results seem to confirm the assumption. However, such a work can not have a sense alone, but should come as a back-up for qualitative studies such as the one lead in [Commenges, 2013] , where free questionnaires with historical actors of modeling provide highly relevant information.

References

[Achibet et al., 2014] Achibet, M., Balev, S., Dutot, A., and Olivier, D. (2014). A model of road network and buildings extension co-evolution. Procedia Computer Science, 32:828–833.

[Barth elemy and Flammini, 2008] Barth elemy, M. and Flammini, A. (2008). Modeling urban street patterns. Physical review letters, 100(13):138702.

[Barth□ elemy and Flammini, 2009] Barth□ elemy, M. and Flammini, A. (2009). Co-evolution of density and topol- ogy in a simple model of city formation. Networks and spatial economics, 9(3):401–425.

[Bretagnolle, 2009] Bretagnolle, A. (2009). Villes et r□ eseaux de transport : des interactions dans la longue dur□ ee, France, Europe, E□ tats-Unis. Hdr, Universit□ e Panth□ eon-Sorbonne - Paris I.

[Chang, 2006] Chang, J. S. (2006). Models of the relationship between transport and land-use: A review. Trans- port Reviews, 26(3):325–350.

[Chavalarias and Cointet, 2013] Chavalarias, D. and Cointet, J.-P. (2013). Phylomemetic patterns in science evolution—the rise and fall of scientific fields. Plos One, 8(2):e54847.

[Commenges, 2013] Commenges, H. (2013). The invention of daily mobility. Performative aspects of the instru- ments of economics of transportation. Theses, Universit□ e Paris-Diderot - Paris VII.

[Delons et al., 2008] Delons, J., Coulombel, N., and Leurent, F. (2008). PIRANDELLO an integrated transport and land-use model for the Paris area.

[Iacono et al., 2008] Iacono, M., Levinson, D., and El-Geneidy, A. (2008). Models of transportation and land use change: a guide to the territory. Journal of Planning Literature, 22(4):323–340.

[Lissack, 2013] Lissack, M. (2013). Subliminal influence or plagiarism by negligence ? the slodderwetenschap of ignoring the internet. Journal of Academic Ethics.

[Mendeley, 2015] Mendeley (2015). Mendeley reference manager. http://www.mendeley.com/.

[Newman, 2013] Newman, M. E. J. (2013). Prediction of highly cited papers. ArXiv e-prints.

[Offner, 1993] Offner, J.-M. (1993). Les "effets structurants" du transport: mythe politique, mystification scien- tifique. Espace g□ eographique, 22(3):233–242.

[Raimbault et al., 2014] Raimbault, J., Banos, A., and Doursat, R. (2014). A hybrid network/grid model of ur- ban morphogenesis and optimization. In Proceedings of the 4th International Conference on Complex Systems and Applications (ICCSA 2014), June 23-26, 2014, Universit□ e de Normandie, Le Havre, France; M. A. Aziz- Alaoui, C. Bertelle, X. Z. Liu, D. Olivier, eds.: pp. 51-60.

[Rucker, 2012] Rucker, G. (2012). Network meta-analysis, electrical networks and graph theory. Research Synthe- sis Methods, 3(4):312–324.

[Sarigo□ l et al., 2014] Sarigo□ l, E., Pfitzner, R., Scholtes, I., Garas, A., and Schweitzer, F. (2014). Predicting Scien- tific Success Based on Coauthorship Networks. ArXiv e-prints.

[Tero et al., 2010] Tero, A., Takagi, S., Saigusa, T., Ito, K., Bebber, D. P., Fricker, M. D., Yumiki, K., Kobayashi, R., and Nakagaki, T. (2010). Rules for biologically inspired adaptive network design. Science, 327(5964):439–442.

[Wegener and Fu□ rst, 2004] Wegener, M. and Fu□ rst, F. (2004). Land-use transport interaction: state of the art. Available at SSRN 1434678.

[Wegener et al., 1991] Wegener, M., Mackett, R. L., and Simmonds, D. C. (1991). One city, three models: com- parison of land-use/transport policy simulation models for dortmund. Transport Reviews, 11(2):107–129.

[Xie and Levinson, 2009] Xie, F. and Levinson, D. (2009). Modeling the growth of transportation networks: A comprehensive review. Networks and Spatial Economics, 9(3):291–307.

[Zhang and Levinson, 2007] Zhang, L. and Levinson, D. (2007). The economics of transportation network growth. In Essays on transport economics, pages 317–339. Springer.