Indirect Evidence of Network Effects in a System of Cities

Journal Title
XX(X):1–9

DOI: 10.1177/ToBeAssigned
www.sagepub.com/
SAGE

Author

Abstract
We propose a simple model of urban growth for systems of cities, which investigate in particular the role of physical networks on interdependence of growth rates. Under the assumption of stochastic independence, a generalized non-linear formulation of recursive population growth captures spatial interactions between cities. At the second order, feedback of physical network, introduced as an abstraction of transportation network, is introduced as influencing average growth rates. Model exploration and calibration using large-scale computation and specific algorithms, yield typical characteristics of spatial interaction, such as decay distance, and their evolution in time under non-stationarity hypothesis. Furthermore, network effects are revealed by a fit improvement when adding network module.

Keywords
Urban Systems, Urban Growth, Spatial Interactions, Network Effects

Introduction

Modeling Urban Growth

Understanding processes driving urban growth is more crucial than ever, as urban population recently crossed the symbolic proportion of half world population .Future of world economies and sustainability of future societies seem to be deeply interlinked with the dynamics of urban systems.A better knowledge of how cities differentiate, interact and grow is thus a relevant topic both theoretically and for application. Many disciplines have studied models of urban growth with different objectives and taking different aspects into account. Economics are still reluctant to include spatial interactions in their models Krugman (1998), whereas geography fails to embrace a certain level of complexity. The example of this two disciplines shows how it is difficult to make

Email:

bridges, as it needed exceptional minds to translate from one to the other (as P. Hall did for Von Thunen work Taylor (2016))

Urban Growth and Spatial Interaction

Some approach of Urban Growth insist on the role of space and spatial interactions. Bretagnolle et al. (2000) already proposed a spatial extension of the Gibrat model. The gravity-based interaction model that Sanders (1992) use to apply concept of Synergetics to cities (getting indeed out of the scope of synergetics by taking expectancies and getting rid of master equation and the probabilistic formulation of trajectories) is also close to this idea of interdependent urban growth, contained physically in the phenomenon of migration between cities. A more refined extension with economic cycles and innovation waves was developed by Favaro and Pumain (2011), yielding a system dynamics version of the core of Simpop models Pumain (2012), which are an other approach to the role of spatial interaction in the growth of urban system

Urban Growth and Networks

We work on simple territorial systems that are country-wide city systems, and more particularly French cities, on a time scale corresponding to that spatial scale, i.e. two last centuries. Taking into account physical networks can improve the understanding of city growth within that system in two ways : a qualitative one, for which the extended model would exhibit qualitative features corresponding to stylized facts empirically observed but that more basic models do not manage to reproduce, and a quantitative way, in the sense that model extension improves explained variance further than the mechanic improvement due to the introduction of supplementary degrees of freedom. If at least one of these is unveiled in our particular case, the evidence will support the theory at these scale and in this context.

The rest of this paper is organized as follows : our model is introduced and formally described in next section ; we then describe results obtained through exploration and calibration of our model on data for french cities, in particular the unveiling of network effects significantly influencing growth processes. We finally discuss interpretations of these results and implications for planning.

Model Description

From Gibrat to Marius : the dilemma of formulation

Some confusion may arise when surveying at stochastic and deterministic models of urban growth. To what extent is a proposed model "complex" and is the simulation of stochasticity necessary ? Concerning Gibrat model and most of its extensions, independence assumptions produce a totally predictable behavior and thus not complex in the sense of exhibiting emergence*. In particular, the full distribution of random growth

*taking e.g. weak emergence, i.e. computational emergence introduced by Bedau models can be analytically at any time , and in the case of studying only first moment, a simple recurrence relation avoids to proceed to any Monte-Carlo simulation. Under these assumptions, it is natural to work with a deterministic model, as it is done by Cottineau with the Marius model .

Model description

We choose to work on a deterministic extension of the Gibrat model, what is equivalent to consider only expectancies in time as explained in the previous subsection. Let $\vec{P}(t) = (P_i(t))_i$ be the population of cities in time. Under Gibrat independence assumptions, we have

$$\text{Cov}[P_i(t), P_j(t)] = 0$$

A linear extended version would write $\vec{P}(t+1) = \mathbf{R} \cdot \vec{P}(t)$ where $\mathbf{R}$ is an independent random matrix of growth rates (identity in the initial case). It yields directly thanks to the independence assumptions that $\mathbb{E}\left[\vec{P}(t+1)\right] = \mathbb{E}[\mathbf{R}] \cdot \mathbb{E}\left[\vec{P}\right](t)$. We generalize this linear relation to a non-linear relation that allows to be more consistent regarding some thematic considerations, by taking with denoting $\vec{\mu}(t) = \mathbb{E}\left[\vec{P}(t)\right]$, the relation $\vec{\mu}(t+1) = f(\vec{\mu}(t))$ (note that in that case, stochastic and deterministic versions are not equivalent anymore[†]). In our case, we take

$$f(\vec{\mu}) = r_0 \cdot \mathbf{Id} \cdot \vec{\mu} + \mathbf{G}\left(\vec{\mu}\right) \cdot \mathbf{1} + \mathbf{N}\left(\vec{\mu}\right) \tag{1}$$

with $G_{ij} = w_G \cdot \frac{V_{ij}}{<V_{ij}>}$ such that interaction potential follow a gravity-type function given by

$$V_{ij} = \left(\frac{\mu_i \mu_j}{\sum {\mu_k}^2}\right)^{\gamma_G} \cdot \exp\left(-d_{ij}/d_G\right) \tag{2}$$

and the network effect term is given by

$$N_i = w_N \cdot \sum_{kl} \left(\frac{\mu_k \mu_l}{\sum \mu}\right)^{\gamma_N} \exp\left(-d_{kl,i}\right)/d_N \tag{3}$$

where $d_{kl,i}$ is distance to shortest path between k, l computed with slope impedance $(Z = (1 + \alpha/\alpha_0)^{n_0}$ with $\alpha_0 \simeq 3)$. The first component is the pure Gibrat model, that we obtain by setting the weights $w_G = w_N = 0$. The second component captures direct interdependencies between cities, under the form of a separable gravity potential such as the one used in Sanders (1992). The rationale for the third term, aimed at capturing network effects by expressing a feedback of network flow between cities k, l on the city i. Intuitively, a demographic and economic flow physically transiting through a city or in its surroundings is expected to influence its development (through intermediate stops

[†]precisely because of the non-linearity

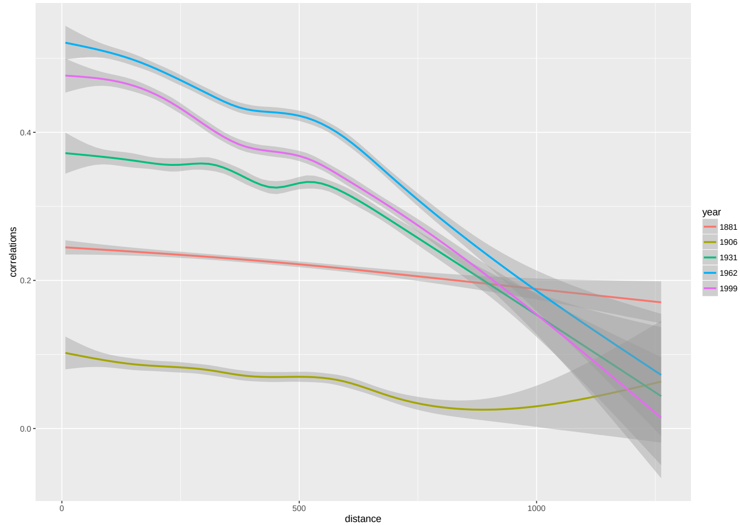

Figure 1. Time-series correlations as a function of distance. Solid line correspond to smoothed correlations, computed between each normalized population time-series, on successive periods. More precisely, we consider overlapping 50 years time-windows finishing respectively in (1881,1906,1931,1962,1999) and compute on each, for each couple of cities (i,j), an estimated correlation $\hat{\rho}_{ij} = \rho\left[\Delta\tilde{P}_i, \Delta\tilde{P}_j\right]$

e.g.), this effect being of course dependent on the transportation mode since a high speed line with few stops will induce a *tunnel effect* .Under assumption of non-stationarity, temporal evolution of fitted parameters corresponding to this feedback should therefore contain information on the evolution of transportation modes.

Model Parameter Space

Data

Population data We work with the Pumain-INED historical database for French Cities, which give populations of *Aires Urbaines* (INSEE definition) at time intervals of mostly 5 years, from 1830 to 1995. The latest version of the database, described in Pumain and Riandey (1986) integrates the definition of Urban Areas, allowing to follow them on long time-period, according to Bretagnole's long time cities ontology Bretagnolle (2009) (that constructs a definition of cities as evolving entities which boundaries are not fixed in time).

Stylized facts Basic stylized facts can be extracted from such a database, as it has already been widely explored in the literature . We show in figure 1 mean time-series correlation as a function of distance

Physical flows As stated above, this modeling exercise focuses on exploring the role of physical flows, whatever the effective shape of the network. We do not need for this reason network data which is furthermore not easily available at different time periods, and physical flows are assumed to take the geographical shortest path that include terrain slope (to avoid geographical absurdities such as cities with a difficult access having an overestimated growth rate). Using the 1km resolution Digital Elevation Model openly available from IGN , we construct an impedance field of the form

$$Z = \left(1 + \frac{\alpha}{\alpha_0}\right)^{n_0}$$

We took fixed parameter values $\alpha_0 = 3$ (corresponding to approximatively a 5% slope) and $n_0 = 3$. Supplementary Material S1 justifies this choice by investigating the sensibility of paths to these parameters.

A semi-parametrized model Our model is assumed as hybrid as it relies on semi-parametrization on real data. It could be possible to be a full toy-model, initial configuration and physical environment being constructed as synthetic data. As Raimbault (2016a) points out, it should even be a step in an extensive study of model, using synthetic data to unveil sensibility of dynamics to meta-parameters defining setup. This enterprise is however out of the scope of this paper, as we aim here to extract advanced stylized facts from a dataset, and we focus therefore on the semi-parametrized version of the model.

Model Evaluation

We work on an explanatory rather than an exploratory model, and indicators to evaluate model outputs are therefore not linked to a performance of trajectories or obtained final states, but to a distance to phenomenon we want to explain, i.e. the data. We use therefore the following complementary indicators :

- Logarithms of mean-square error
- Mean-square error on logarithms

Results

Implementation

Data preprocessing, result processing and models profiling are implemented in R. For performances reasons and an easier integration into the OpenMole software for model exploration described by Reuillon et al. (2013), a `scala` version was also developed. The typical question of trade-off between implementation performance and interoperability appeared quickly as an issue, as a blind exploration and calibration can difficultly provide useful thematic conclusions for that kind of model. Finding an improvement in model fit among one parameter dimension is significant if the geographical situation is visualized and the improvement is confirmed as reasonable and not an absurdity.

Figure 2. Example of output of the model. The graphical interface allows to explore interactively on which cities changes operate after a parameter change, what is necessary to interpret raw calibration results.

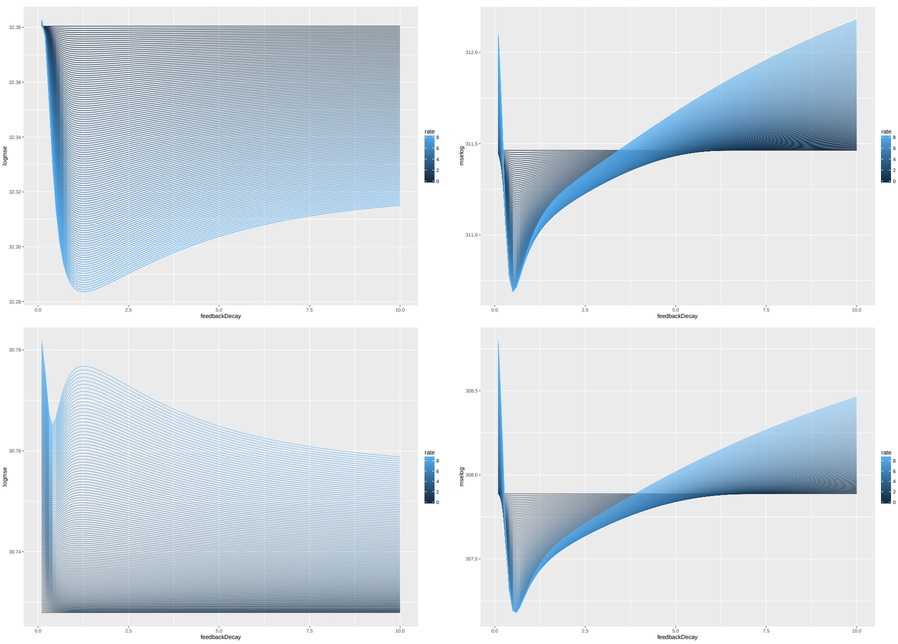

Figure 3. Evidence of network effects revealed by model exploration. Feedback with fixed gravity : first evidences of network effects ; confirmed with effect of α_0

Model Exploration

Model Calibration

Calibration using Genetic Algorithms The optimization problem associated to model calibration does not present features allowing an easy solving (closed-form of a likelihood function, convexity or sparsity of the optimization problem, etc.), we must rely on alternative techniques to solve it. Brute force grid search is rapidly limited by the dimensionality curse

Quantifying overfitting : Empirical AIC A large part of statistical models allow to compute tools for model comparison and selection, in particular to take into account possible overfitting due to additional degrees of freedom. The Akaike Information Criterion provides the gain in information between two models, and many extensions have been proposed since .Note that cross-validation type of methods are not suitable to our case because of the small size of the dataset. We can formalize the proposed method based on the intuitive idea of approaching models of simulation by statistical models and

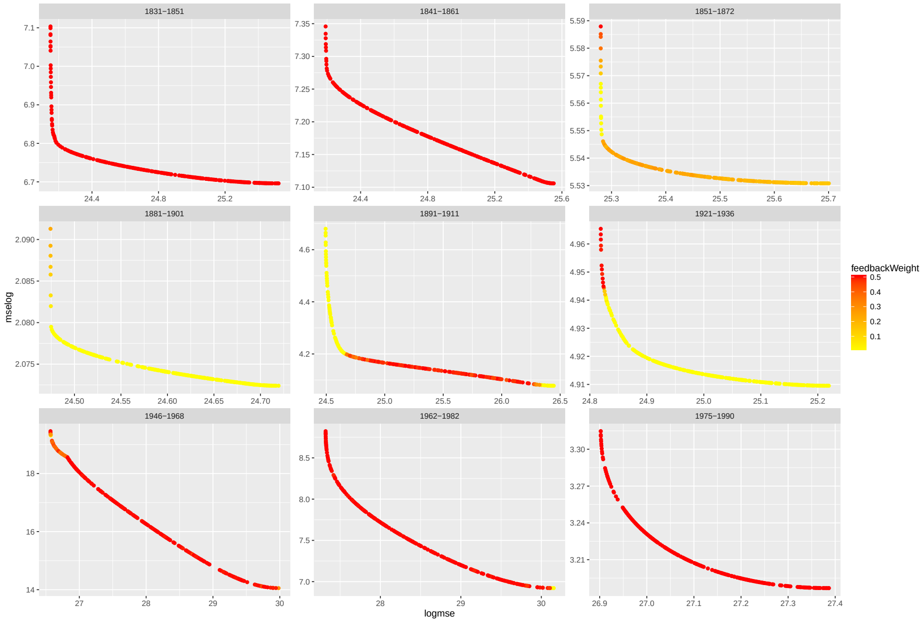

Figure 4. Pareto fronts for bi-objective calibration. Steady-state populations are obtained using Island Genetic Algorithm embedded in `OpenMole`, with parameters, and on independent time periods as detailed in text. We show corresponding parameter values for w_N, for the gravity-only model (Top) and the full model (Bottom). Full graphs are available in Supplementary materials.

using the corresponding AIC under certain validity conditions. Let (X, Y) be the data and observations. Computational models are functions $(X, \alpha_k) \mapsto M^{(k)}_{\alpha_k}(X)$ mapping data values to a random variable. What is seen as data and parameters is somehow arbitrary but is separated in the formulation as corresponding dimensions will have different roles. We assume the model has been fitted to data in the sense that an heuristic has been used to approximate $\alpha^*_k = \text{argmin}_{\alpha_k} \left\| M^{(k)}_{\alpha_k}(X) - Y \right\|$. The gain of information between two computational models is not directly accessible and we propose an indirect way, through the fitting of statistical models. For all computational model, a large set of statistical models with similar degree of freedom are fitted. Let S_k be the statistical models fitting best $M^{(k)}_{\alpha^*_k}(X)$. We can compute

$$\Delta D_{KL}\left(M^{(1)}|M^{(2)}\right) = \Delta D_{KL}\left(S^{(1)}|S^{(2)}\right) + \left[\Delta D_{KL}\left(S^{(2)}|M^{(1)}\right) + \Delta D_{KL}\left(S^{(1)}|M^{(2)}\right)\right]$$

Under certain assumptions, the order of magnitude of second term appears to be negligible. More precisely, with $s^{(k)} = M^{(k)} - S^{(k)}$, we have

$$\left\| \int f \log\left(\frac{S^{(k)}}{M^{(k')}}\right) \right\| = \left\| \int f \log\left(1 + \frac{s^{(k)}}{M^{(k')}}\right) \right\|$$

Discussion

Theoretical implications

We propose to support our hypothesis that *physical transportation networks are necessary to explain the morphogenesis of territorial systems* (aka *Network Necessity*) by showing on a relatively simple case that the integration of physical networks into some models effectively increase their explanative power (being careful on the precise definition of model improvement to avoid overfitting).

Methodological implications

Further developments

Specificity of the Urban System The model has not yet been tested on other urban systems and other temporalities.

Towards co-evolutive models of cities and transportation networks Our focus on network effects remains quite limited since (i) we do not consider an effective infrastructure but abstract flows only, and (ii) we do not take into account the possible network evolution, due to technical progresses (Bretagnolle et al. (2000)) and infrastructure growth in time. An ambitious but necessary development would be the inclusion of both in a model of co-evolution between urban growth and transportation network growth, in order to investigate to what extent the refinement of network spatial structure and network dynamics can improve the explanation of urban system dynamics. It has been shown by Raimbault (2016b) that disciplinary compartmentalization may be at the origin of the relative absence of such type of models in the literature.

Conclusion

Acknowledgements

The author would like to thank

References

Bretagnolle A (2009) *Villes et réseaux de transport : des interactions dans la longue durée, France, Europe, États-Unis*. Hdr, Université Panthéon-Sorbonne - Paris I. URL http://tel.archives-ouvertes.fr/tel-00459720.

Bretagnolle A, Mathian H, Pumain D and Rozenblat C (2000) Long-term dynamics of european towns and cities: towards a spatial model of urban growth. *Cybergeo: European Journal of Geography* .

Favaro JM and Pumain D (2011) Gibrat revisited: An urban growth model incorporating spatial interaction and innovation cycles. *Geographical Analysis* 43(3): 261–286.

Figure 5. Densities of shortest paths, computed with the 1225 shortest between all 50 cities included in the dataset, for different values of parameters (α_0, n_0).

Figure 6. Errors on shortest paths using the geographical path heuristic. (a) Distribution of integrated distance between paths. (b) Density maps for both heuristic and real paths.

Krugman P (1998) Space: the final frontier. *The Journal of Economic Perspectives* 12(2): 161–174.

Pumain D (2012) Multi-agent system modelling for urban systems: The series of simpop models. In: *Agent-based models of geographical systems*. Springer, pp. 721–738.

Pumain D and Riandey B (1986) Le fichier de l'ined. *Espace, populations, sociétés* 4(2): 269–277.

Raimbault J (2016a) Generation of correlated synthetic data. In: *Actes des Journees de Rochebrune 2016*.

Raimbault J (2016b) Models coupling urban growth and transportation network growth: An algorithmic systematic review approach. *arXiv preprint arXiv:1605.08888* .

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.

Sanders L (1992) *Système de villes et synergétique*. Economica.

Taylor PJ (2016) A polymath in city studies. In: *Sir Peter Hall: Pioneer in Regional Planning, Transport and Urban Geography*. Springer, pp. 11–20.

Physical Flows Parametrization

We show in Fig. 5 footprints of shortest paths, revealing how different values make the path more or less realistic regarding the elevation map. A rather rough validation can also been done by comparing these with the actual current shortest paths by the road network. These are computed using the road network dataset provided by and which construction is detailed in . For each path, we can define a measure of geographical distance as, given $(\vec{l_i}(s), \vec{l_j}(s))$ normalized linear parametrization of two paths (p_i, p_j), by

$$d(p_i, p_j) = \int_{s=0}^{1} \left\| \vec{l_i}(s) - \vec{l_j}(s) \right\| ds$$

We show in Fig. 6