

A generative model for urban firm networks

J. Raimbault^{1,2,3,*}, N. Zdanowska^{1,3}

¹ Centre for Advanced Spatial Analysis, UCL

² UPS CNRS 3611 ISC-PIF

³ UMR CNRS 8504 Géographie-cités

* `juste.raimbault@polytechnique.edu`

Abstract

The emergence of entangled urban networks is a salient feature of globalisation processes. Understanding the drivers behind the growth of such networks, and in particular urban firm networks, is crucial for the economic resilience of urban systems. We introduce in this paper a generative network model for firm networks at the urban area level, which includes several complementary processes which are the economic size of urban areas at origin and destination, industrial sector proximity between firms, the strength of links from the past as well as the geographical and socio-cultural distance. An empirical network analysis on European firm ownership data confirms a role for each factor. We then simulate network growth for synthetic systems of cities, unveiling stylized facts such as a transition from a local to a global regime or a maximal integration achieved at an intermediate interaction range. We finally calibrate the model on the European network, outperforming statistical models and showing a strong role of path-dependency. Potential applications of the model include the study of mitigation policies for exogenous shocks such as economic crisis.

1 Introduction

The intensification of globalisation of the economy since the late 20th century resulted in the rise of the network society [1] and the emergence of a world-city system [2], characterized by highly interconnected urban centres at the global level [3]. These interactions can be of different nature and determined by economic, socio-cultural, or geopolitical proximity implying spatial and non spatial patterns [4]. Interactions of economic nature between transnational firms, are together with trade among the most representative of the current economic trends [2]. Understanding the drivers of growth and geographical properties of such a urban network, can be used to foster innovation in specific cities and to shape public policies for local industrial clusters [5]. Straightening these city-interactions can be crucial for revitalising certain geographical areas [6], by developing strategic industries for global insertion of regions [7], cities [8] and countries [9]. The functioning of such a city-network and the position of cities within these networks [8] can also permit to infer the consequences of future exogenous shocks such as an economic crisis, a single market collapse or a post-Brexit scenario for the European Union.

The use of network models to understand processes underlying the emergence and dynamics of such urban networks has already been widely tackled in the literature. [10] propose a generative urban network model at the regional scale, combining geographical factors with network topological properties. Air transportation networks can also be understood as urban networks, and [11] reviews several models for such networks. Trade network have been extensively studied from complex networks perspective, but are generally driven at the country-level [12] because of lack of city-level data. Input-output models, mostly used in regional science to represent flows between geographical areas, are considered in some cases as a type of urban generative model [13]. In this stream of research, spatial interaction models [14] can form the basis of urban network models for abstract flows [15] or physical infrastructure [16].

In this paper we choose to focus on urban networks of firms to question how can we understand and capture geographical and economic processes as agglomeration economies [17] characteristic to spatial economy [18] within a generative model. The specific question we tackle is the complementarity of different processes to generate such urban networks. One originality of the present paper is to approach these issues within the system of cities framework [19], where the position and dynamics of cities in the socio-economic world-wide space can be considered by their interactions with other cities [20]. We examine the interactions of European cities within firm linkages defined by corporate ownership links. Since transnational firms structure is one of the determinant of the global economic space, it provides one accurate proxy to unveil geographical structures [21].

More precisely, our contribution relies on the following points: (i) we propose an empirical analysis of the firm ownership urban network for the European Union, based on the Bureau Van Dijk’s *AMADEUS* database; (ii) we introduce a generative network model to simulate the growth of such linkages at urban areas scale, which combines multiple factors influencing link formation, as economic intensity of origin and destination of urban areas, industrial sector proximity, the strength of previous links, and geographical and socio-cultural distance; (iii) the model allows us to compare the effect of different factors on the final network structure, which is extensively studied through model sensitivity analysis and exploration; and (iv) we calibrate the model on the empirical network.

The rest of this paper is organized as follows: we first proceed to an empirical analysis of the network studied. We then describe the generative model for urban networks, and its sensitivity analysis and exploration on synthetic data. We then parametrize and calibrate the model on real European firm ownership data. We finally discuss implications of our results and possible extensions.

2 Empirical network analysis

Companies from the AMADEUS database are georeferenced using the Geonames database (using postcode or address depending on the information available). We then join this data with the GHS-FUA dataset for Functional Urban Areas [22], which are the most relevant objects for applying our model. Starting from a firm-level dataset of 1,562,862 nodes and 1,866,936 links, we obtain a directed network of 573 urban areas and 9323 ownership links. Weight of links are obtained by computing the owned share of turnover at destination, i.e. $w_{ij} = \sum_{k \in i, l \in j} p_{kl} \cdot T_l$ where p_{kl} of the ownership share of company l by company k and T_l the turnover of company l . Node attributes include an economic size of urban areas, that we compute as the sum of turnovers of companies within the area. This quantity is highly correlated with area-level GDP included in the GHS database (Spearman correlation $\rho = 0.71$) and also with population ($\rho = 0.60$), so we expect to capture size effects while keeping the consistence of a single main data source. An other attribute of nodes is the industrial sector composition, expressed as a proportion of turnover in the area associated to a given industrial sector. Detailed sectors of firms are available up to the four digits of the NACE classification in the raw data. We consider the first level classification (21 categories) and compute sector proportions accordingly. This allows us to define an industrial proximity matrix between urban areas, taking a cosine similarity between vectors as $s_{ij} = \sum_{k=1}^K S_{ik} \cdot S_{jk}$ where S_{ik} are sector proportions.

The urban network has heavy tail weighted degree and edge weight distributions, as shown in Fig. 1. We fit power law and log-normal distributions, including minimal value cutoff following [23], using the `powerlaw` R package [24]. For both, log-normal distributions appear to be a better fit ($\mu = 18.8$, $\sigma = 2.3$ for weighted degree; $\mu = 13.8$, $\sigma = 2.8$ for edge weight), and include a

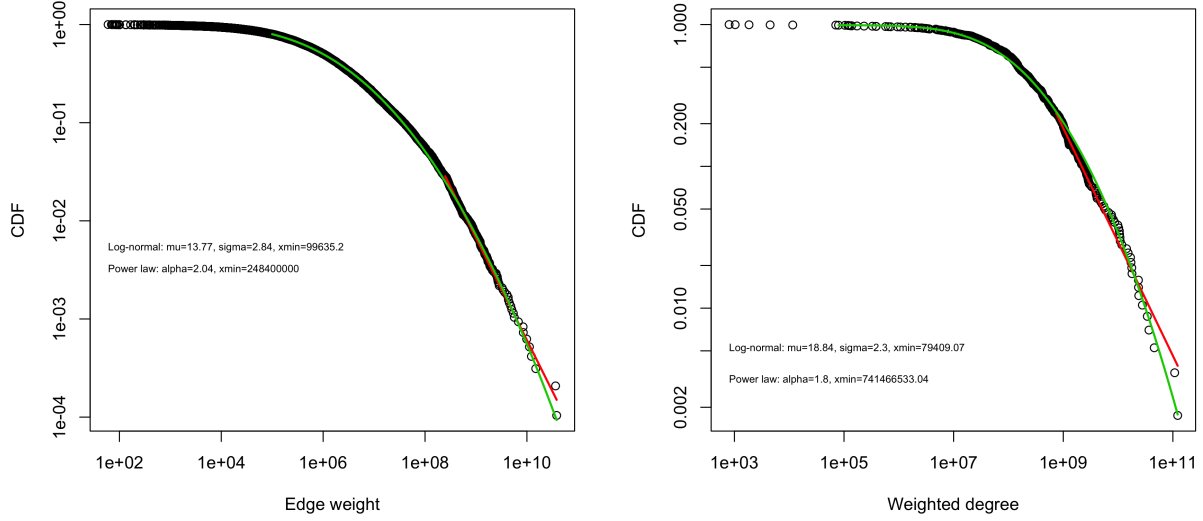


Figure 1: Distributions of network properties. (*Left*) Cumulated distribution for edge weight, with power law fit (red) and log-normal fit (green); (*Right*) Cumulated distribution for weighted degree.

much broader range of the empirical distributions with lower estimated cut-offs. These heavy tail properties suggest the relevance of a generative model including self-reinforcing processes which are known to produce such distributions.

Regarding the community structure of the network, different modularity maximisation algorithms (run on the undirected corresponding network) give a same undirected modularity of 0.38, and a directed modularity [25] of 0.34 for the greedy algorithm of [26] and 0.36 for the Louvain algorithm [27], corresponding to 36 communities (resp. 15). In comparison, with 29 countries, the classification of urban areas by country has a modularity of 0.32. This means that the structure of flows present certain regional patterns. Indeed, a null model assigning random labels with 30 communities, bootstrap 1000 times, yields an average modularity of 0.049 ± 0.002 , confirming the statistical significance of these communities.

To have a first insight into factors playing a role in the formation of links, we proceed to a statistical analysis similar to unconstrained spatial interaction models [28]. We test five different statistical models, including progressively the processes candidate to be included in the network generative model. The most general statistical model is

$$\log(w_{ij}) = \log(d_{ij}) + \log(T_i) + \log(T_j) + \log(s_{ij}) + \alpha_{c_i, c_j} + \varepsilon_{ij}$$

where α_{c_i, c_j} is a fixed effect term for the couple of countries c_i, c_j . We use a multiplicative form by taking logarithms as it is usually done for spatial interaction models. Furthermore, origin and/or destination constrained models are not relevant in our case as links include not only turnover but also a share of ownership. Estimation results are given in Table 1. We also fit a Poisson generalized linear model with a logarithmic link on the rounded weights [29]. We find that distance alone (model 1) has a significant negative effect but with an explained variance close to zero. The second model adds fixed effects by pair of countries, for which around a third are significant, and with a better R^2 of 0.17. Adding origin and destination economic size (turnovers T_i and T_j , model 3) slightly improves the explanative power, but the best model among the linear ones is the model

4 with all factors and fixed effects, both in terms of explained variance ($R^2 = 0.31$) and Akaike information criterion. The latter indicates the absence of overfitting. The Poisson model with all variables is much better regarding pseudo- R^2 and provides much more significant estimates (all fixed effects significant and much smaller standard deviations). Across all models, we find consistent qualitative stylized facts: (i) distance has a negative influence; (ii) economic sizes have a positive influence, with the size of the origin being more important (what is consistent with ownership links, strongly driven by decision at the headquarters); (iii) industrial proximity has a positive influence, i.e. similarity is more important than complementarity for this dimension; (iv) an important part of country fixed-effects are significant; and (v) the order of magnitude of the influence of each factor is roughly the same, in the sense that no process totally dominates the others. This suggests the relevance of including all these different dimensions in the generative network model.

Table 1: Ordinary Least Square estimations of statistical models explaining $\log(w_{ij})$ (standard errors in parenthesis). ***, ** and * respectively indicate 0.01, 0.05 and 0.1 levels of significance. For the fixed effect by origin-destination pairs of countries, we give the proportion of significant coefficients (where $p < 0.1$). Models (1) to (4) are ordinary least square models, while model (5) is a Poisson generalized linear model with logarithmic link. We compare model performance using R^2 , Mean Square Error, and Akaike Information Criterion. For the Poisson model, AIC is not given as it is not comparable, and the pseudo- R^2 is given either.

Model	(1)	(2)	(3)	(4)	(5)
$\log(d_{ij})$	-0.06** (0.03)	-0.11** (0.05)	-0.41*** (0.02)	-0.35*** (0.04)	-0.26*** ($5 \cdot 10^{-6}$)
$\log(T_i)$			0.56*** (0.01)	0.56*** (0.01)	0.79*** ($2 \cdot 10^{-6}$)
$\log(T_j)$			0.39*** (0.01)	0.39*** (0.01)	0.66*** ($1.5 \cdot 10^{-6}$)
$\log(s_{ij})$				0.19*** (0.07)	0.53*** ($9 \cdot 10^{-6}$)
Countries		28.3%		26.5%	100%
R^2	0.00059	0.17	0.21	0.31	0.56
MSE	7.75	6.84	6.10	5.33	8.72
AIC	44304	43578	42131	41917	

3 Model

3.1 Rationale

The model is expected to capture a single macroscopic urban scale, i.e. the level of the urban system where basic entities are cities. Links are induced by underlying firms but these are not explicit. Cities are defined by their economic size, but also by an industrial sector profile. As [4] point out, a combination of several distances may play a role in establishing linkages. Geographical distance, industrial proximity and complementarity will typically be equally significant [30]. Therefore, we combine several processes driving network growth: (i) geographical proximity (in our case the crow-fly distance, but it can be generalized to any type of accessibility); (ii) geopolitical proximity, which captures the higher propensity to establish links within existing submarkets despite the European common market (confirmed relevant by the fixed effect regression described above); (iii) city economic size, which has furthermore scaling laws properties; (iv) industrial proximity in terms of firm sector composition; (v) previous linkages (influence of history). This last point is crucial to capture accumulated effects and include path-dependency, which fits well the use of a simulation model instead of a more simple statistical model.

3.2 Model description

Let $1 \leq i \leq N$ cities defined at time t by their economic structure such that $E_i(t)$ is the total economic volume (GDP) for city i and $S_{ik}(t)$ is a matrix giving economic volume proportions within each economic sector k , assuming K economic sectors. Formally, the model creates links between cities which are characterized by their economic size E_i (GDP) and economic structure S_{ik} in terms of activity sectors (probability distribution of firms within K sectors). Starting with an existing network with no links, the model iteratively adds links, following a probability given by a generalized Cobb-Douglas function [31] as

$$p_{ij} \propto \left(\frac{E_i}{E}\right)^{\gamma_F} \cdot \left(\frac{E_j}{E}\right)^{\gamma_T} \cdot \left(\frac{w_{ij}}{W}\right)^{\gamma_W} \cdot s(S_{ik}, S_{jk})^{\gamma_S} \cdot \exp(-\gamma_G \cdot d_{ij}) \cdot \exp(-\gamma_D \cdot g_{ij}) \quad (1)$$

where $E = \sum_k E_k$, $W = \sum_{i,j} w_{ij}$ weights of previous links, $s(S_{ik}, S_{jk})$ is a similarity function between activity sectors S_{ik} and S_{jk} taken as a cosine similarity, d_{ij} euclidian distance, and g_{ij} a socio-cultural distance. In the case of an already existing link between two areas, the weight of the latter is incremented by a unit volume w_0 . The model is stopped either when a final time is reached or when the cumulated volume of links reaches a target quantity. This formulation can be considered both as an economic utility model and a generalisation of spatial interaction models.

Note that: (i) we consider an asymmetric influence of sizes, assuming that link directions are important (similarly the similarity function s may be taken as asymmetric); (ii) the influence of previous links is similar to a preferential attachment process; (iii) we do not renormalize the exponents to 1 (which is different than renormalising the probability), to ensure to include convex functions. The “geopolitical/sociocultural” distance remains abstract and should be parametrised (see setup below) and estimated in the data-driven setting. We assume no evolution of economic size in this simple version of the model, which means that adjusted parameter are valid on a restricted time period.

3.3 Model setup

Besides model dynamics, its initialisation (or parametrisation on real data) regarding the spatial distribution, sizes, countries and industrial sector compositions, is also important to understand model behavior. In addition [32] showed that the initial spatial configuration can have significant impacts on model outcomes. We therefore consider two versions: one with a synthetic system of cities which can be stochastically repeated, and one other based on real data.

Synthetic setup The synthetic system of cities is generated to be at the scale of a continent with properties similar to Europe: (i) $N = 700$ cities with an empirical power-law from the Global Human Settlements database for economic sizes ($\alpha = 1.1$), distributed randomly into a uniform and isotropic square space of width $d_{max} = 3000km$; (ii) clustered into $C = 20$ countries using a k-means algorithm (this remains consistent with non-correlated city sizes at the scale of the countries as our cities are randomly distributed in space [33]); (iii) with synthetic distribution of sectors following log-normal distributions adjusted in a way that large cities have more high-value activities and are more diverse. More precisely for this last point, assuming sectors are distributed continuously along a one-dimensional axis, cities are assumed to have a sector distribution with a log-normal density function. The largest city is constrained to have a standard deviation and a mode of $1/2$ and the smallest $1/K$, with a linear interpolation between the two for other cities. Parameters of the log-normal law are obtained by solving the corresponding two equations system

for each city, and each distribution is discretised and renormalised to be a vector of probabilities. In synthetic experiments, the network is generated from an empty initial network until reaching $t_f = 1500$, with a unit link volume of $w_0 = 1$.

Real setup The real setup is done using the empirical network analyzed above. Urban areas are distributed according to their positions in the actual geographical space, and economic size and sector composition attributes are used. The socio-economic distance is constructed using information from the statistical modeling, by considering fixed-effect coefficients estimated for each pair of countries. Initial network is empty, and the real links are taken as targets for the generated network. We constrain the total volume of flows by setting $w_0 = \sum_{ij} w_{ij}/t_f$ when the number of time steps t_f is given as a parameter.

3.4 Model parameters and indicators

The model parameters are the Cobb-Douglas weights $\gamma_F, \gamma_T, \gamma_W, \gamma_s$ and distance decay parameters γ_G, γ_D . The model also includes path-dependency by integrating the influence of previous links. Therefore it can not be reduced to a simple statistical model. Finally, the role of space introduces even further complexity, yielding a generative model, which has to be simulated.

We consider several macroscopic indicators to quantify the generated network. Firstly, geographical structures are captured by (i) internationalisation (modularity of countries in the network); (ii) metropolisation (correlation between weighted degree and city size). Two other thematic indicators, which are regionalisation (correlation between length and flow of links, stratified by size of extremities), and specialisation (correlation between sector proximity and flow of links, stratified by size of extremities), are not included in this study for the sake of simplicity as each has a high dimension. We also consider several network indicators as Louvain modularity, sizes of maximal modularity communities, degree and link weight distributions (described by their average, hierarchy, and entropy), and correlations between degree and city size, and between link weight and distance.

4 Model exploration and calibration

The model is implemented in NetLogo, which provides a good compromise between performance and interactive prototyping and exploration [34]. Numerical experiments are done by integrating the model into the OpenMOLE software for model exploration [35], which provides sensitivity analysis and calibration methods and a transparent access to high-performance computing infrastructures. Model source code, aggregated network data (raw initial firm network can not be made available for data ownership reasons), and results are available on the open Git repository of the project at <https://github.com/JusteRaimbault/ABMCitiesFirms>. Large simulation result files are available on the dataverse repository at <https://doi.org/10.7910/DVN/UPX23S>. We show in Fig. 2 examples of generated networks, with very contrasted patterns obtained with different spatial interaction ranges. Low values of the gravity decay parameter yield local networks around main cities, while long interaction ranges yield more integrated networks. We also show networks simulated in the real setting.

4.1 Synthetic city systems

We first study extensively the behavior of the model on synthetic systems of cities, to isolate its intrinsic behavior independently of a contingent geographical situation.

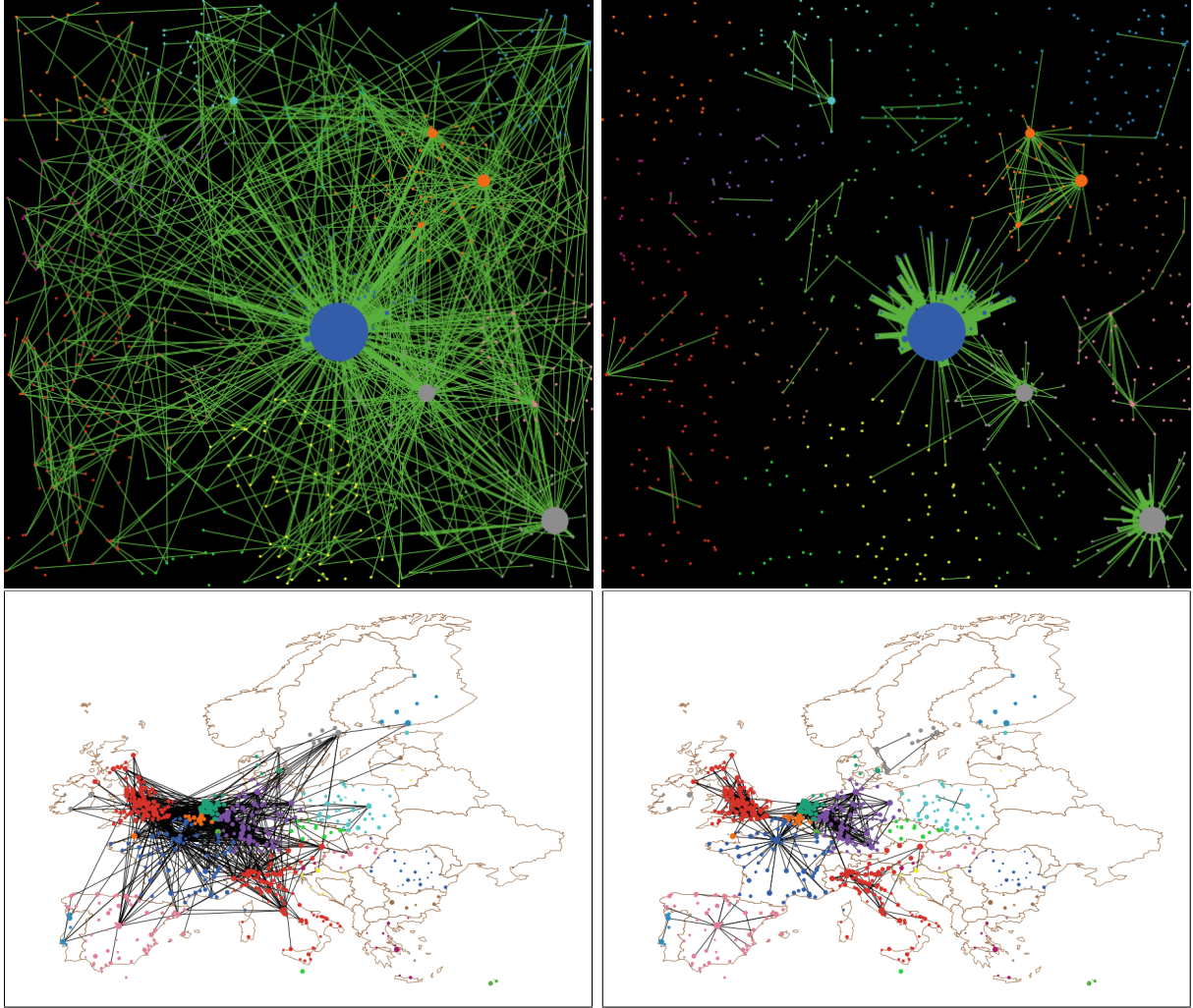


Figure 2: Example of simulated networks at $t = 1500$, for a synthetic system of cities (Top row) and the European urban network (Bottom row). , with a high (resp. low) gravity decay parameter for the left column (resp. right).

Table 2: Saltelli sensitivity indices, for indicators in rows and parameters in columns. We give for each pair the first order index (F) and the total order index (T).

	γ_G		γ_D		γ_S		γ_W		γ_F		γ_T	
	F	T	F	T	F	T	F	T	F	T	F	T
Internationalisation	0.2	0.3	0.7	0.7	0.001	0.009	$4 \cdot 10^{-4}$	0.007	0.03	0.04	0.02	0.04
Metropolisation	0.02	0.1	0.02	0.2	0.002	0.1	0.001	0.09	0.2	0.6	0.3	0.6
Modularity	0.3	0.4	0.6	0.6	0.004	0.02	$3 \cdot 10^{-4}$	0.01	0.005	0.03	0.002	0.03
Avg. com. size	0.008	0.09	0.01	0.1	0.002	0.07	0.003	0.04	0.3	0.6	0.4	0.6
Degree entropy	0.006	0.02	0.003	0.02	0.006	0.03	0.008	0.02	0.5	0.5	0.5	0.5
Weight entropy	0.04	0.1	0.03	0.1	0.008	0.08	0.01	0.07	0.4	0.5	0.4	0.5

Sensitivity analysis A global sensitivity analysis first unveils an aggregated influence of parameters on model indicators. We apply Saltelli’s Global Sensitivity method introduced by [36]. This produces values for each parameter and each output of a first order sensitivity index, defined as $Var[\mathbb{E}_{\mathbf{X}_{\sim i}}(Y_j|X_i)]/Var[Y]$, for parameter X_i and output Y_j , $\mathbf{X}_{\sim i}$ being all other parameters. This captures the total influence of X_i all other things being equal. The total order index is given by $\mathbb{E}_{\mathbf{X}_{\sim i}}[Var(Y_j|\mathbf{X}_{\sim i})]/Var[Y_j]$ and captures non-homogeneity in behavior and interaction between parameters. Indices were estimated with a Sobol sequence sampling of 20,000 model runs [37]. We give in Table 2 estimated results. We find that indicators related to modularity (internationalisation and optimal modularity) are mostly influenced by geographical parameters. On the contrary, indicators of network structure are driven by origin and destination sizes. Sectors and link history have a second order by not negligible role. Altogether, the analysis witnesses of strong interaction effects between parameters, since for several indicators the total order index is one order of magnitude larger than the first order index. This means that the generative model captures some complex behavior of the system. The very different important parameters depending on outputs studied confirms the relevance of a more exhaustive grid exploration.

One factor sampling A first numerical experiment of one-factor sampling on all Cobb-Douglas parameters and 100 stochastic repetitions confirms a good statistical convergence (average Sharpe ratios for indicators all larger than 5). We use thus 20 repetitions in following larger experiments (in the case of a normal distribution, a 95% confidence interval on the average is of size $2\sigma \cdot 1.96/\sqrt{n}$, what gives $n = 16$ repetitions for a CI of width of the standard deviation). Behavior of some indicators are shown in Fig. 3. For example, internationalization and the correlation between economic size and weighted degree both show a phase transition as a function of γ_G , from a local to a global regime. The sector proximity parameter γ_S influences the internationalization linearly, but induces a maximum for the correlation between size and degree, which can be interpreted as a setting where size and the total volume in a city are the most correlated.

Grid exploration The model behaviour is then studied with a grid sampling for parameters and 20 repetitions (the computations run on the European Grid Infrastructure, and are equivalent to 2.5 years of CPU time). Distance decay and sector parameters are varied with 10 steps, other parameters with three. The influence of gravity decay parameters is confirmed when plotting the internationalization index against the distance, which exhibits an exponential decay as shown in Fig. 4. It interacts with the role of sectors and the size of the origin, with a more significant effect of sector proximity when size has the largest influence (third panel).

Other indicators exhibit non-trivial patterns. For example, when considering average com-

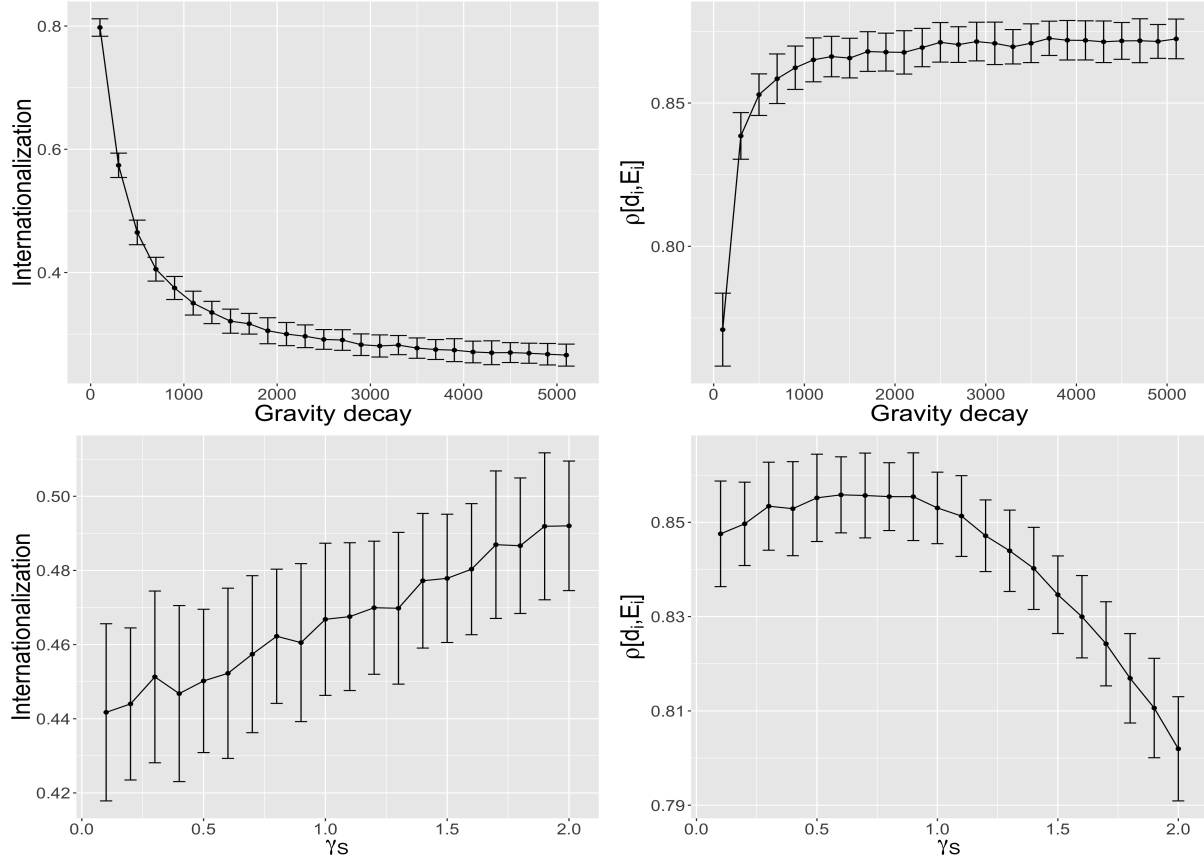


Figure 3: (Top Left) Internationalisation index decreases exponentially with gravity decay; (Top Right) Correlation between city weighted degree and size. Both plots show a transition from a local to a global regime. (Bottom Left) Internationalisation varies linearly with sector proximity γ_S ; (Bottom Right) Correlation between degree and size exhibits a maximum, witnessing an intermediate regime where size is the most important.

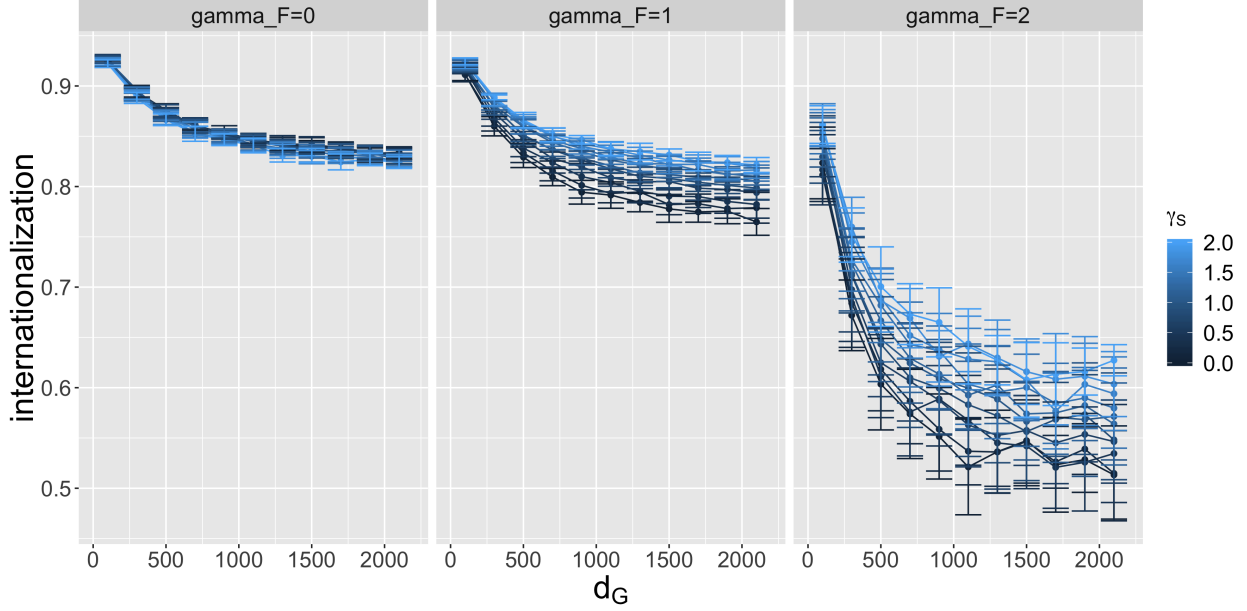


Figure 4: The transition as a function of interaction range depends on the influence of origin size γ_F ; sector proximity γ_S plays a role only for a large influence of the origin.

munity size of the final network as shown in Fig. 5, we obtain a maximal integration in term of communities at an intermediate value of the gravity decay, what can be interpreted as the emergence of a regional regime. The size of communities is largely influenced by the value of the elasticities for the similarity function and the ratio of the economic output of the area. In particular, we observe a qualitative inversion of the role of γ_S when introducing an effect of the origin (switching γ_F from 0 to 1, first panel to second panel). The maximal community size disappears when $\gamma_F = 2$, implying a regime with small local communities where large cities mostly connect with there hinterland. These experiments reveal a complex interplay between processes and how the model produces diverse stylised facts.

4.2 Model calibration on the European urban network

We then apply the model to a real system of cities by calibrating it on the European ownership network. Model setup is done following the real setup described above. The number of time steps t_f is left as an additional parameter, which in a sense determines the granularity of the cumulative network generation process. The objective functions for calibration are the average mean squared error on logarithms of weights given by $\varepsilon_L = \frac{1}{N^2} \sum_{i,j} (\log w_{ij} - \log \hat{w}_{ij})^2$ if \hat{w}_{ij} are the simulated weights, and the logarithm of the mean squared error $\varepsilon_M = \log \left(\frac{1}{N^2} \sum_{i,j} (w_{ij} - \hat{w}_{ij})^2 \right)$. [16] showed that these two are complementary to calibrate urban systems models. Calibration is done using a genetic NSGA2 algorithm [38], which is particularly suited for such a bi-objective calibration of a simulation model. Stochasticity is taken into account by adding the estimation confidence as an additional objective, and we filter the final population to have at least 20 samples. We use a population of 200 individuals and run the algorithm for 20,000 generations, on 1000 islands in parallel.

We show in Fig. 6 the calibration result, as a Pareto front of compromise points between the two objectives. As the mean squared error on logarithm is the same than the one for statistical models,

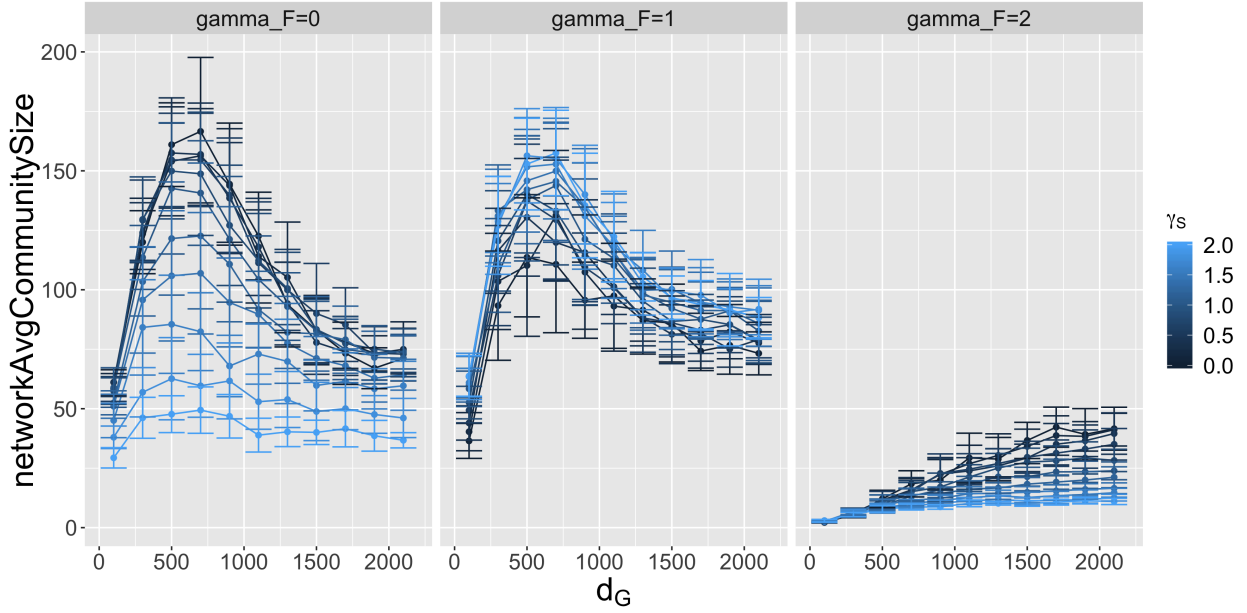


Figure 5: Maximal integration in term of community size is achieved at an intermediate value of d_G : emergence of a regional regime; Maximal size depends on the role of sectors γ_S , in a decreasing way when origin size is deactivated, and increasing way when $\gamma_F = 1$; This regime disappear when origin size influence is too large.

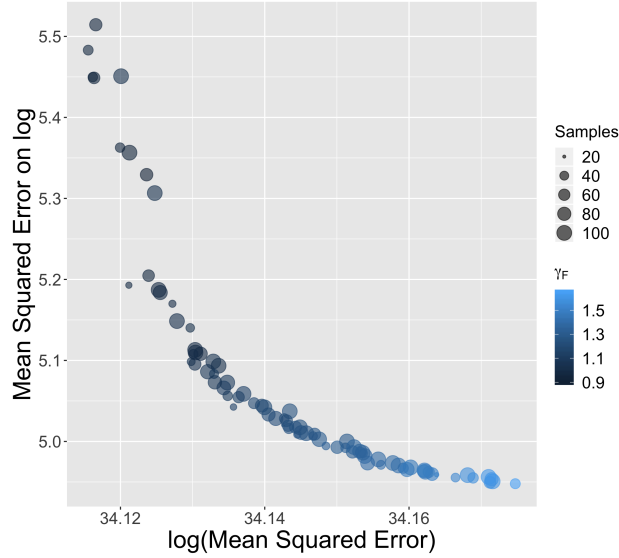


Figure 6: Pareto front obtained for the bi-objective model calibration for ε_L and ε_M objectives. ε_L can be compared to the MSE of statistical models fitted above, and the generative model outperform these for several parametrizations (lowest MSE 5.33 for statistical models). Point size gives the number of stochastic samples and color the value of γ_F .

it can be directly compared. The other objective would correspond to statistical models directly on variables: therefore, our model covers in a sense a set of intermediate models which make a compromise between the two objectives, and is therefore much more flexible. The values for ε_L are for a large proportion of points lower than for statistical models, our model outperforming these in this predictive sense. As shown in Fig. 6, increasing values for γ_F lead to better performance regarding ε_L : increasing the importance of the origin will provide a better fit on all order of magnitudes of link weights (while ε_M favors mostly the largest links given the fat-tail distribution).

When considering optimal points with $\varepsilon_L < 5.0$, the adjusted parameters are: long range gravity interactions and socio-economic interactions given by $d_G = 1/\gamma_G = 5205 \pm 1775$ and $d_D = 1/\gamma_D = 109 \pm 29$ (the maximal g_{ij} is 11.8); a stronger influence of destination than origin as $\gamma_F = 1.45 \pm 0.09$ and $\gamma_T = 2.0 \pm 0.06$; a comparable influence of sectors as $\gamma_S = 1.75 \pm 0.17$; a fine time granularity as $t_f = 4960 \pm 109$; and a very strong influence of previous links as $\gamma_W = 7.99 \pm 3.77$. This last point furthermore shows in particular a role of the path-dependency parameter, confirming the relevance of this complex generative model including path-dependency. The interaction of this process with others implies an inversion of the relative importance of origin and destination (on the contrary to statistical models). The fact that calibrated interaction distances are large is consistent with long range economic interactions within Europe. Besides these stylized facts, these calibration results altogether demonstrate the feasibility of model application to real data.

5 Discussion

This paper aimed at presenting a generative model for urban networks defined by interactions between firms based on synthetic data, simulated via the OpenMole platform and calibrated on real data on ownership linkages of firms for Europe. The simulation on a synthetic system of cities provides an broad knowledge on model behavior in its parameter space, and its calibration suggests its potential application to policy issues. This kind of model can indeed have practical applications in the future in terms of testing the effect of exogenous shocks in the socio-economic structure as the upcoming effect of Brexit on the European market.

Even in such a simple model (close to directly tractable stationary state) the behaviour is highly non-linear in many dimensions. Our study shows how crucial model exploration is to overcome hidden parameters (deactivated mechanisms or default parameter values). This paper emphasis on the fact that exploration of intrinsic dynamics on synthetic data is of great importance should be driven before an application on real data so as to put aside the geographical effects from model dynamics.

Several possible extensions of the current work can be undertaken in the future as a co-evolution model with an evolution of city sizes or a model with firm agents leading to a multi-scale agent-based model. A more precise study of the role of path dependency can also be undertaken in order to understand the creation of the future links. Other formulations of the model might be taken into consideration as other formulation of the combination of factors or multi-objective optimisation depending on sectors using Pareto fronts.

References

- [1] Manuel Castells. The rise of the network society. *Blackwell Publishers, Cambridge MA*, 2000.
- [2] Peter J Taylor. Specification of the world city network. *Geographical analysis*, 33(2):181–194, 2001.
- [3] Saskia Sassen. The global city: New york, london, tokyo. *Princeton University Press, New Jersey*, page 480, 1991.
- [4] Kirsten Martinus and Thomas J Sigler. Global city clusters: theorizing spatial and non-spatial proximity in inter-urban firm networks. *Regional studies*, 52(8):1041–1052, 2018.
- [5] Ekaterina Turkina, Ari Van Assche, and Raja Kali. Structure and evolution of global cluster networks: evidence from the aerospace industry. *Journal of Economic Geography*, 16(6):1211–1234, 2016.
- [6] Jennifer Clark, John Harrison, and Ernest Miguelez. Connecting cities, revitalizing regions: the centrality of cities to regional development. *Regional Studies*, 52(8):1025–1028, 2018.
- [7] Stuart Dawley, Danny Mackinnon, and Robert Pollock. Creating strategic couplings in global production networks: regional institutions and lead firm investment in the humber region, uk. *Journal of Economic Geography*, 2019.
- [8] Johannes Glückler and Robert Panitz. Relational upgrading in global value networks. *Journal of Economic Geography*, 16(6):1161–1185, 2016.
- [9] Kirsten Martinus, Thomas Sigler, Iacopo Iacopini, and Ben Derudder. The brokerage role of small states and territories in global corporate networks. *Growth and Change*, 2019.
- [10] Liang Dai, Ben Derudder, Xingjian Liu, Lei Ye, and Xuejun Duan. Simulating infrastructure networks in the yangtze river delta (china) using generative urban network models. *Belgeo. Revue belge de géographie*, (2), 2016.
- [11] Massimiliano Zanin and Fabrizio Lillo. Modelling the air transport with complex networks: A short review. *The European Physical Journal Special Topics*, 215(1):5–21, 2013.
- [12] Diego Garlaschelli and Maria I Loffredo. Structure and evolution of the world trade network. *Physica A: Statistical Mechanics and its Applications*, 355(1):138–144, 2005.
- [13] Yu-xian Jin and Alan Wilson. Generation of integrated multispatial input-output models of cities (gimimoc) i: Initial stage. *Papers in Regional Science*, 72(4):351–367, 1993.
- [14] Adam Dennett and Alan Wilson. A multilevel spatial interaction modelling framework for estimating interregional migration in europe. *Environment and Planning A*, 45(6):1491–1507, 2013.
- [15] Liang Dai, Ben Derudder, and Xingjian Liu. Generative network models for simulating urban networks, the case of inter-city transport network in southeast asia. *Cybergeog: European Journal of Geography*, 2016.
- [16] Juste Raimbault. Indirect evidence of network effects in a system of cities. *Environment and Planning B: Urban Analytics and City Science*, 47(1):138–155, 2020.

- [17] Edward L. Glaeser. Agglomeration economics. *The University of Chicago Press*, 2011.
- [18] Masahisa Fujita, Paul Krugman, and Anthony J. Venables. The spatial economy: cities, regions, and international trade. *Cambridge: MIT Press*, 2001.
- [19] Brian JL Berry. Cities as systems within systems of cities. *Papers in regional science*, 13(1):146–163, 1964.
- [20] Denise Pumain. An evolutionary theory of urban systems. In *International and Transnational Perspectives on Urban Systems*, pages 3–18. Springer, 2018.
- [21] Natalia Zdanowska. Exploring cities of Central and Eastern Europe within transnational company networks: the core-periphery effect. *arXiv e-prints*, page arXiv:1910.14652, Oct 2019.
- [22] A.J. Florczyk, C. Corbane, M. Schiavina, M. Pesaresi, L. Maffenini, M. Melchiorri, P. Politis, F. Sabo, S. Freire, D. Ehrlich, T. Kemper, P. Tommasi, D. Airaghi, and L. Zanchetta. Ghs urban centre database 2015, multitemporal and multidimensional attributes, r2019a. Technical report, European Commission, Joint Research Centre (JRC), 2019.
- [23] Aaron Clauset, Cosma Rohilla Shalizi, and Mark EJ Newman. Power-law distributions in empirical data. *SIAM review*, 51(4):661–703, 2009.
- [24] Colin S. Gillespie. Fitting heavy tailed distributions: The powerLaw package. *Journal of Statistical Software*, 64(2):1–16, 2015.
- [25] Vincenzo Nicosia, Giuseppe Mangioni, Vincenza Carchiolo, and Michele Malgeri. Extending the definition of modularity to directed graphs with overlapping communities. *Journal of Statistical Mechanics: Theory and Experiment*, 2009(03):P03024, 2009.
- [26] Aaron Clauset, Mark EJ Newman, and Cristopher Moore. Finding community structure in very large networks. *Physical review E*, 70(6):066111, 2004.
- [27] Vincent D Blondel, Jean-Loup Guillaume, Renaud Lambiotte, and Etienne Lefebvre. Fast unfolding of communities in large networks. *Journal of statistical mechanics: theory and experiment*, 2008(10):P10008, 2008.
- [28] Alan Geoffrey Wilson. Some new forms of spatial interaction model: A review. *Transportation Research*, 9(2-3):167–179, 1975.
- [29] Robin Flowerdew and Andrew Lovett. Fitting constrained poisson regression models to interurban migration flows. *Geographical Analysis*, 20(4):297–307, 1988.
- [30] Clémentine Cottineau and Elsa Arcaute. The nested structure of urban business clusters. *Applied Network Science*, 5(1):1–20, 2020.
- [31] Gabriel Eduard Vilcu. A geometric perspective on the generalized cobb–douglas production functions. *Applied Mathematics Letters*, 24(5):777–783, 2011.
- [32] Juste Raimbault, Clémentine Cottineau, Marion Le Texier, Florent Le Nechet, and Romain Reuillon. Space matters: Extending sensitivity analysis to initial spatial conditions in geosimulation models. *Journal of Artificial Societies and Social Simulation*, 22(4):10, 2019.

- [33] Filippo Simini and Charlotte James. Testing heaps law for cities using administrative and gridded population data sets. *EPJ Data Science*, 8(1):1–13, 2019.
- [34] Steven Railsback, Daniel Ayllón, Uta Berger, Volker Grimm, Steven Lytinen, Colin Sheppard, and Jan Thiele. Improving execution speed of models implemented in netlogo. *Journal of Artificial Societies and Social Simulation*, 20(1), 2017.
- [35] Romain Reuillon, Mathieu Leclaire, and Sebastien Rey-Coyrehourcq. Openmole, a workflow engine specifically tailored for the distributed exploration of simulation models. *Future Generation Computer Systems*, 29(8):1981–1990, 2013.
- [36] Andrea Saltelli, Marco Ratto, Terry Andres, Francesca Campolongo, Jessica Cariboni, Debora Gatelli, Michaela Saisana, and Stefano Tarantola. *Global sensitivity analysis: the primer*. John Wiley & Sons, 2008.
- [37] Andrea Saltelli, Paola Annoni, Ivano Azzini, Francesca Campolongo, Marco Ratto, and Stefano Tarantola. Variance based sensitivity analysis of model output. design and estimator for the total sensitivity index. *Computer Physics Communications*, 181(2):259–270, 2010.
- [38] Kalyanmoy Deb, Amrit Pratap, Sameer Agarwal, and TAMT Meyarivan. A fast and elitist multiobjective genetic algorithm: Nsga-ii. *IEEE transactions on evolutionary computation*, 6(2):182–197, 2002.