# Unveiling co-evolutionary patterns in systems of cities: a systematic exploration of the SimpopNet model

*Chapter proposal for Theories and Models of Urbanization*

Juste Raimbault[1,2]
[1] UPS CNRS 3611 ISC-PIF
[2] UMR CNRS 8504 Géographie-cités

**Abstract**

**Keywords :** *Co-evolution; Networks and Territories*

## 1 Introduction

A considerable gain in knowledge can be observed, from the conceptual or thematic description of a model, to its mathematical formalisation, its implementation, its systematic exploration, up to its exploration in deep with the help of specific meta-heuristics. Our postulate, that is a consequence of both our positioning (see **??** on simulation) and experiments of which previously developed models are part, is that it is important, but furthermore of a *qualitative* nature, in the sense that the nature of knowledge follows abrupt transitions during the advance of the investigation in this continuum.

The SimpopNet model introduced by **?**, which is to our knowledge the only co-evolution model in the perspective of the evolutive urban theory. Its behavior was however not systematically explored, what makes it a good candidate for our approach.

### 1.1 Description of the model

We briefly reformulate the model, following the notations for the formalization of the interaction model in **??**, since a certain number of parameters and processes are similar. Cities grow following a specification that rejoins equation **??**, i.e. in this specific case

$$\mu_i(t+1) - \mu_i(t) = \mu_i(t) \cdot \frac{\lambda^\beta}{N} \sum_j \frac{V_{ij}}{< V_{ij} >}$$

where the potential is of the form $V_{ij} = \mu_j / d_{ij}^\beta$ and $V_{ii} = 0$, and $\beta$ is a parameter for the distance decay and $\lambda$ shape parameter for the decay function. We thus find our formulation, with $r_0 = 0$ and $w_G = \lambda^\beta \cdot N$. Since $\lambda$ gives the typical distance of interaction, it will be noted $d_G$ in the following, and $\beta$ will be noted $\gamma_G$ (it is indeed a level of hierarchy as a function of distance).

The network growths at each time step through a process that can be seen as a potential breakdown (as described in chapter **??**): a couple of cities is chosen, the first according to populations with a hierarchy $\gamma_N$ (i.e. with a probability proportional to ${\mu_i}^{\gamma_N}$) and the second following interaction forces $\mu_i \mu_j / d_{ij}^\beta$ with the same hierarchy $\gamma_N$. A link is then created if the network is not efficient enough, i.e. if $d_{ij}/d_{ij}^{(N)} > \theta_N$. The links created at a date $t$ have a speed $v(t)$, which will depend on current transportation technologies. The creation of new intersections to yield a planar graph is only done for links with a similar speed.

In order to study a stylized version of the model, we consider a configuration such that $v(t > 0) = v_0$ and $v(0) = 1$ (the initial model considers three values for speed that correspond to the reality of transportation technologies between 1830 and 2000).

### 1.2 Perspectives

We can put the structure of this model into perspective. Some modeling choices are not in direct consistency with the application it is used for: for example, such a precision in the parametrization of dates and speeds (historical dates from 1800 to 2000 and speed that approximatively corresponds to transportation technologies) makes it a hybrid model, and should correspond to an application on a real spatial configuration. In a synthetic configuration as used in the model, these parameters have a sense only if we know the behavior of simulated dynamics, and in particular the role of the spatial configuration, i.e. if we are able to differentiate effects linked to the dynamics from effects linked to the initial spatial configuration.

Furthermore, the use of the interaction model without the endogenous Gibrat term would be difficultly adaptable to an application of the model on real data since the values we obtained in the precedent studies of interaction models, but stays relevant in a stylized model, in order to understand the interaction processes in an isolated way, as we will do later (keeping in mind that this knowledge does not necessarily describes the coupled behavior, since the interaction between processes can lead to the emergence of new behaviors).

The formulation of the potential, given above, as $(\lambda/d_{ij})^{\beta}$, implies that $\lambda$ captures both the weight of the potential and the shape of the decreasing function, but imposes a dependence between these two effects, on the contrary to the specification we use previously. It furthermore does not allow an interpretation in terms of limit flows[1].

Finally, rules allowing variable values for $v(t)$ and the non-planarity mechanism[2], allows the introduction of a tunnel effect, which is as we recall is the absence of interaction of an infrastructure traversing a territory with it. The effect is however exogenous since explicitly specified in model rules, on the contrary to the interaction model with feedback of flows, in which the variations of $w_N$ and $d_N$ should capture an endogenous tunnel effect. The introduction of specific indicators to measure it would be an interesting development direction, but we stay here at considering the hierarchy of centralities which is already a good indicator for it[3].

## 2 Methods

### 2.1 Spatial configuration

An important aspect for understanding co-evolution processes implied in this model is the role of the initial spatial configuration in emerging patterns observed. We therefore apply the methodology developed in **??**, which allows to extend the sensitivity analysis of a model to spatial meta-parameters[4].

**Generation of synthetic configurations** A synthetic system of cities is constructed the following way (see Appendix **??** for the notion of synthetic data, calibrated at the first and the second order). A fixed number $N$ of cities is uniformly distributed in space, under the constraint of a minimal distance between

[1]The weight parameter in our model in **??** gives indeed the value of the flow when the distance attenuation goes to infinity and for all the population.

[2]When a new link is constructed, it does create intersections only with links of similar speed.

[3]Indeed, a highly hierarchical distribution of accessibilities means that there exists a small number of cities very accessible and a large number with a low accessibility. If main cities reasonably cover the space, then their links necessarily ignore the overflown cities with low accessibility, otherwise the distribution would be less hierarchical.

[4]We recall that in our case a meta-parameter is a parameter allowing to generate an initial configuration upstream of the model.

each, and their population is attributed following a rank-size law which parameters $P_m$ and $\alpha$ can be adjusted (the distribution of city sizes in the initial model corresponds to $\alpha \simeq 0.68$ with $R^2 = 0.98$).

A skeleton of network is created by progressive connection: the algorithm connects cities two by two by closest neighbour in terms of euclidian distance, and then iteratively selects randomly a cluster and connects it perpendicularly to the closest link outside the cluster. The network is then extended by the creation of local shortcuts, through a repetition $n_s$ times of the random selection of a city according to populations, and its connection to a neighbour in a radius $r_s$ under conditions of a maximal degree $d_s$. The final network is then made planar.

This process creates networks that visually correspond (in terms of the order of magnitude of the number of loops, and their spatial range) to the initialization of the model, knowing that a single instance of the network does not allow to determine distributions of topological parameters for which a more precise calibration could be done.

### 2.1.1 Indicators

A crucial aspect of the study of simulation models is the definition of relevant indicators, particularly in the case of synthetic models where it is not possible to produce outputs that are directly linked to data for example. Very general stylized facts, as aiming at producing an urban hierarchy or a network hierarchy, are relatively limited. Moreover, the hierarchy is mechanically produced by most models including aggregation processes. We therefore need more elaborated indicators to understand the dynamics of the system. These indicators must in particular give elements of answer to the following questions:

- types of systems of cities produced by the model;
- change in time of the organization of the system of cities;
- typical profiles of trajectories;
- ability to "produce some co-evolution".

In order to concentrate on the ability of the model to produce trajectories that are both diverse and complex, and for example its ability to produce bifurcations that would manifest as inversions in ranks, and also its ability to capture different aspects of co-evolutive dynamics, we propose a set of indicators, including for example lagged correlation measures in the spirit of causality regimes exhibited in **??**, or a correlation measure as a function of distance, to understand the role of spatial interactions in the coupling of trajectories. Given a variable $X_i(t)$ defined for each city and in time (that will be the population or centrality measures for example), we define the following indicators.

- Indicators characterizing the distribution of $X_i$ in time: hierarchy (slope of the least squares adjustment of $X_i$ as a function of rank) $\alpha(t)$, entropy of the distribution $\varepsilon(t)$, descriptive statistics (average $\hat{\mathbb{E}[X]}(t)$ and standard deviation $\hat{\sigma}(t)$).
- Rank correlation between the initial time and the final time, which translates the quantity of change in the hierarchy during the evolution of the system, and is defined by $\rho_r = \hat{\rho}\left[rg(X_i(t=0)), rg(X_i(t=t_f))\right]$, where $rg(X_i)$ is the rank of $X_i$ among all values.
- Diversity of trajectories $\mathcal{D}\left[X_i\right]$, which captures a diversity of time series profiles for the considered variable. With $\tilde{X}_i(t) \in [0;1]$ the trajectories that have been individually rescaled, it is defined by

$$\mathcal{D}\left[X_i\right] = \frac{2}{N \cdot (N-1)} \sum_{i<j} \left( \frac{1}{T} \int_t \left( \tilde{X}_i(t) - \tilde{X}_j(t) \right)^2 \right)^{\frac{1}{2}}$$

- Changes in direction of trajectories $\mathcal{C}[X_i]$, that we take as the number of inflexion points. In the context of such a type of model, which mainly produces monotonous trajectories, this indicator witnesses in a certain way of a "complexity" of trajectories.

- Correlations as a function of distance, to understand the way the effect of distance is translated at the macroscopic scale. The profile of this function, regarding interaction distance parameters included in the model, will translate the tendency of the model to lead to the emergence of one level of interaction or the other. It is computed as

$$\rho_d = \hat{\rho}\left[(X(\vec{x}_k, Y(\vec{x}_{k'}))\right]$$

where $X_i, Y_i$ are the two variables considered and $(k, k')$ the set of couples such that $\|\vec{x}_k - \vec{x}_{k'}\| - d \leq \varepsilon$ with $\varepsilon$ a tolerance threshold (in practice taken to regroup couples by distance deciles).

- Lagged correlations between the variations of variables, to identify causality patterns between variables $X$ and $Y$. The patterns $\hat{\rho}_\tau$ for all variables, and for $\tau$ lag or anticipation, must be understood in the sense of potential regimes, explored in **??**.

$$\rho_\tau = \hat{\rho}\left[\Delta X(t-\tau), \Delta Y(t)\right]$$

These indicators are used on the following variables:

- populations $\mu_i(t)$,
- closeness centralities

$$c_i(t) = \frac{1}{N-1}\sum_{i\neq j}\frac{1}{d_{ij}(t)}$$

which capture the position within the urban system,

- accessibilities

$$X_i = \frac{1}{\sum_k \mu_k}\sum_{i\neq j} P_j \exp\left(-d_{ij}(t)/d_G\right)$$

which capture the insertion within the urban system.

We furthermore introduce diverse indicators for network topology, to understand the final forms produced by the heuristic: diameter, average path length, average betweenness centrality and its level of hierarchy, average performance, total length, as they have been defined in **??**.

# 3 Results

## 3.1 Experience plan

Given an initial spatial configuration (i.e. a value of meta-parameters), we establish the behavior of indicators by exploring a grid of the parameter space. The number of parameters being low and the objective being a first grasp of the model behavior, in particular if it is able to produce co-evolution dynamics, we do not use more elaborated exploration methods. The parameters are $(d_G, \gamma_G, \gamma_N, \theta_N, v_0)$ and meta-parameters $(N_S, \alpha_S, d_S, n_S)$. We take also the meta-parameters into account in order to understand the sensitivity of the model to space.

We explore a grid of 16 configurations of meta-parameters, 324 configurations of parameters, and 30 random replications, what corresponds to 155520 simulations. They are executed on a computation grid with the intermediary of OpenMole[5].

[5] Simulation results are available at `http://dx.doi.org/10.7910/DVN/RW8S36`.

Since the model is stochastic, it is important to control the convergence of indicators, that will be more or less easy depending on their variability. To quantify the variability of an indicator $X$ regarding stochasticity, we use a measure similar to the one used in **??**, given by $v\left[X\right] = \hat{\mathbb{E}}\left[X\right]/\hat{\sigma}\left[X\right]$ with basic estimators for the expectance and the standard deviation. On the full set of replications, we obtain for all indicators given previously, a median for the ratio $v\left[X\right]$ estimated within replications, estimated on all parameter values, which takes a minimal value of 3.94, for the average accessibility at final time, what witnesses a low stochastic variability. We can furthermore use this value to estimate the level of convergence: it corresponds to a 95% confidence interval around the mean of relative size 0.18 (under the assumption of a normal distribution of the average), i.e. a good convergence. This aspect is crucial for the robustness of results.

### 3.2 Sensitivity to space

The Table 1 give values of $\tilde{d}$ for 16 configurations of meta-parameters[6], in comparison to an arbitrary reference configuration (first column). The hierarchy within the initial system of cities appears as the stronger determinant of variability, since all configurations with $\alpha_S = 1.5$ give values larger than 1.7, what witnesses a very strong sensitivity relative to this hierarchy.

Then, the number of cities plays a non negligible secondary role , giving the stronger effects of space. Thus, it is crucial to keep in mind this role of the initial configuration during the analysis of phase diagrams. To stay within the same spirit than the model that was initially proposed, we will however comment a phase diagram for a given spatial configuration. The study of the extended model with integration of meta-parameters to which it is sensitive at their full extent is beyond the reach of this auxiliary analysis.

Table 1: **Sensitivity to space of the SimpopNet model.** Each column corresponds to an instance of the phase diagram, for which meta-parameters are given, with the relative distance to an arbitrary reference diagram. As inputs we have the meta-parameters $N_S, \alpha_S, d_S, n_S$ and as outputs of simulations the distance $\tilde{d}$.

| | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| $N_S$ | 40 | 40 | 40 | 40 | 40 | 40 | 40 | 40 | 80 | 80 | 80 | 80 | 80 | 80 | 80 | 80 |
| $\alpha_S$ | 0.5 | 0.5 | 0.5 | 0.5 | 1.5 | 1.5 | 1.5 | 1.5 | 0.5 | 0.5 | 0.5 | 0.5 | 1.5 | 1.5 | 1.5 | 1.5 |
| $d_S$ | 5 | 5 | 10 | 10 | 5 | 5 | 10 | 10 | 5 | 5 | 10 | 10 | 5 | 5 | 10 | 10 |
| $n_S$ | 10 | 30 | 10 | 30 | 10 | 30 | 10 | 30 | 10 | 30 | 10 | 30 | 10 | 30 | 10 | 30 |
| $\tilde{d}$ | 0 | 0.05 | 0.26 | 0.21 | 1.79 | 1.80 | 1.79 | 1.72 | 0.44 | 0.36 | 0.42 | 0.42 | 2.25 | 2.23 | 2.24 | 2.21 |

### 3.3 Model behavior

The Fig. 1 reports the behavior of the model according to a selection among the diverse indicators given above. We comment a particular spatial configuration which corresponds to a low hierarchical system with a network having only local shortcuts, given by meta-parameters $N_S = 80, \alpha_S = 0.5, d_S = 10, n_S = 30$, which are the values giving configurations that are the most similar to the one of the initial model. Complete plots are available in Appendix **??**.

The values taken by the entropy for centralities (first panel of Fig. 1), as a function of time, for $\gamma_N = 2.5$ and $v_0 = 110$, exhibit different regimes depending on $d_G$ and $\gamma_G$. A low hierarchy leads to an entropy stabilizing in time, what corresponds to a certain uniformization of distances. On the contrary, a strong hierarchy produces a regime with a minimum, and then an increase of disparities in time.

This variety of behaviors can be found again with the rank correlation $\rho_R$, that we show here for the population variable, as a function of $d_G$. It has a low sensitivity to $\theta_N$ and $\gamma_N$ (see Appendix **??**), but strongly varies as a function of $d_G$ and $\gamma_G$: interactions at a higher distance induce systematically a larger

[6] The definition of the relative measure of sensitivity, given in **??**, is for two phase diagrams $f_1, f_2$ and $d$ euclidian distance, $\tilde{d} = 2d(f_1, f_2)/(\mathrm{Var}[f_1] + \mathrm{Var}[f_2])$.

number of changes in the hierarchy of populations. These can occur when the hierarchy of distance is low. To summarize, the increase of the range of interactions will diminish the inertia of trajectories of the system of cities, whereas the increase of their hierarchy will increase it. This is relatively credible from a thematic point of view: longer and uniform interactions have more chances to make individual trajectories change.

The behavior of correlation indicators is shown in Fig. 2. Concerning the effect of distance on correlations between variables, i.e. the evolution of $\rho_d$, it is interesting to note that an increase of $d_G$ systematically diminishes the levels of correlation, what corresponds to the complexification that we previously showed. As expected, $\rho_d\left[d\right]$ decreases as a function of distance, and exhibits non zero values for the correlation between population and centrality for a high hierarchy $\gamma_G$, what shows that simultaneous adaptation regimes are rare in this model.

### 3.4 Causality regimes

Finally, by studying $\rho_\tau$ (Fig. 2, bottom panel), we observe that causality regimes in the sense of **??** are not very varied (as the Fig. **??** in Appendix **??** confirms it for a broader range of parameters). The population is systematically caused by the centrality, but there exists no regime in which we observe the contrary. This is a logic of an effect of reinforcement of hierarchy by centrality, but not a configuration with circular causalities, and thus not a co-evolution properly speaking as we defined in the statistical sense.

This brief exploration allows us to say that this model captures urban trajectories of a certain complexity, but that it does apparently not reproduces co-evolution regimes.

### 3.5 PSE algorithm

## 4 Discussion

## 5 Conclusion

We have thus in this section introduced the tools to understand trajectories produced by a co-evolution model, and tested these on the SimpopNet model.

In the following, we will explore in the same spirit a co-evolutive extension of the interaction model developed in **??**, and will aim at establishing to what extent it is able to capture co-evolutive dynamics.

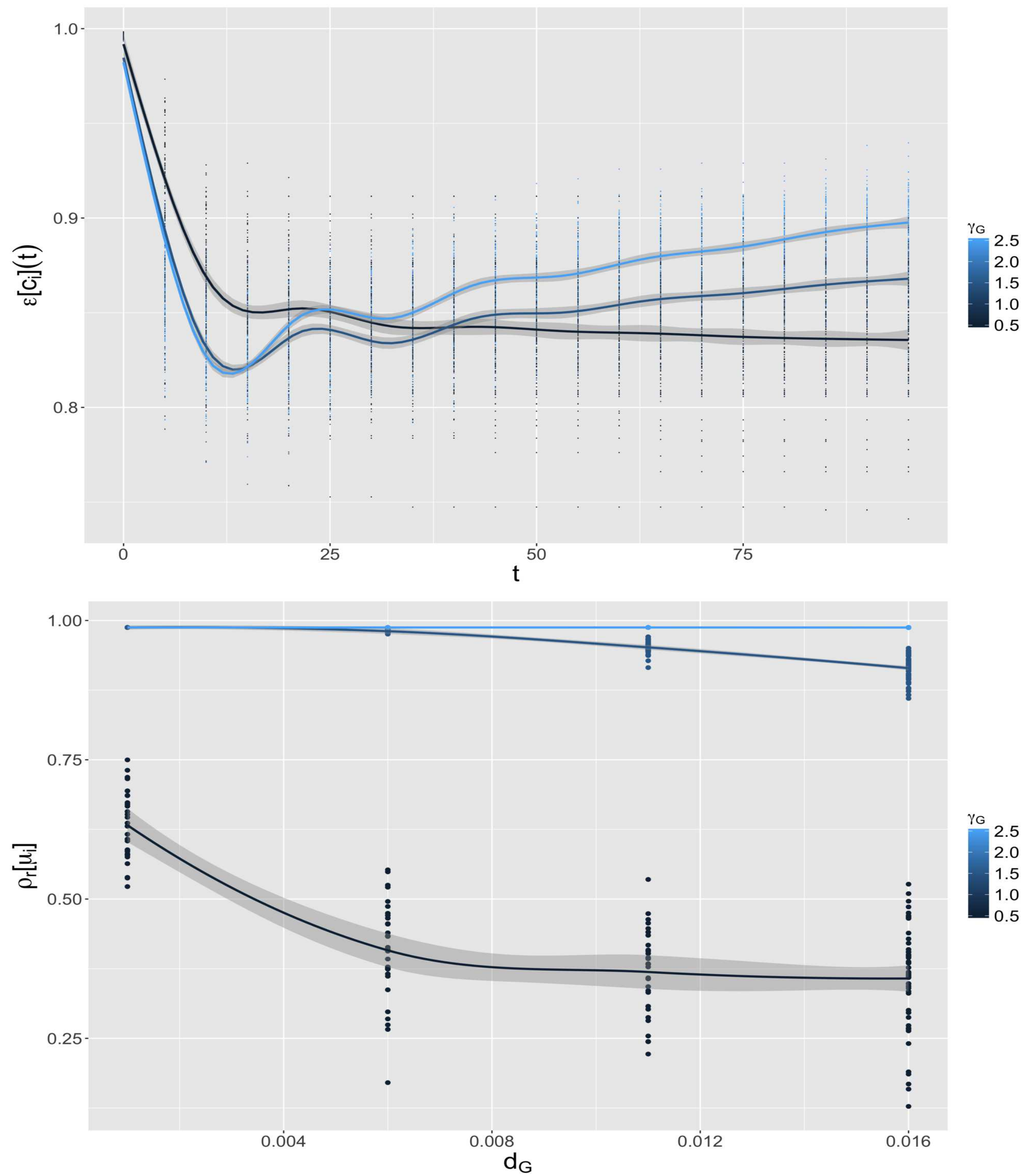


Figure 1: **Model behavior for the spatial configuration** $N_S = 80, \alpha_S = 0.5, d_S = 10, n_S = 30$**.** (*Top*) Temporal trajectories of the entropy for closeness centralities, for $\gamma_N = 2.5$, $v_0 = 110$, $d_G = 0.016$, $\theta_N = 11$, as a function of $\gamma_G$ (color); (*Bottom*) Rank correlation for population, as a function of $d_G$ and of $\gamma_G$ (color), for $\theta_N = 11$, $\gamma_N = 2.5$.

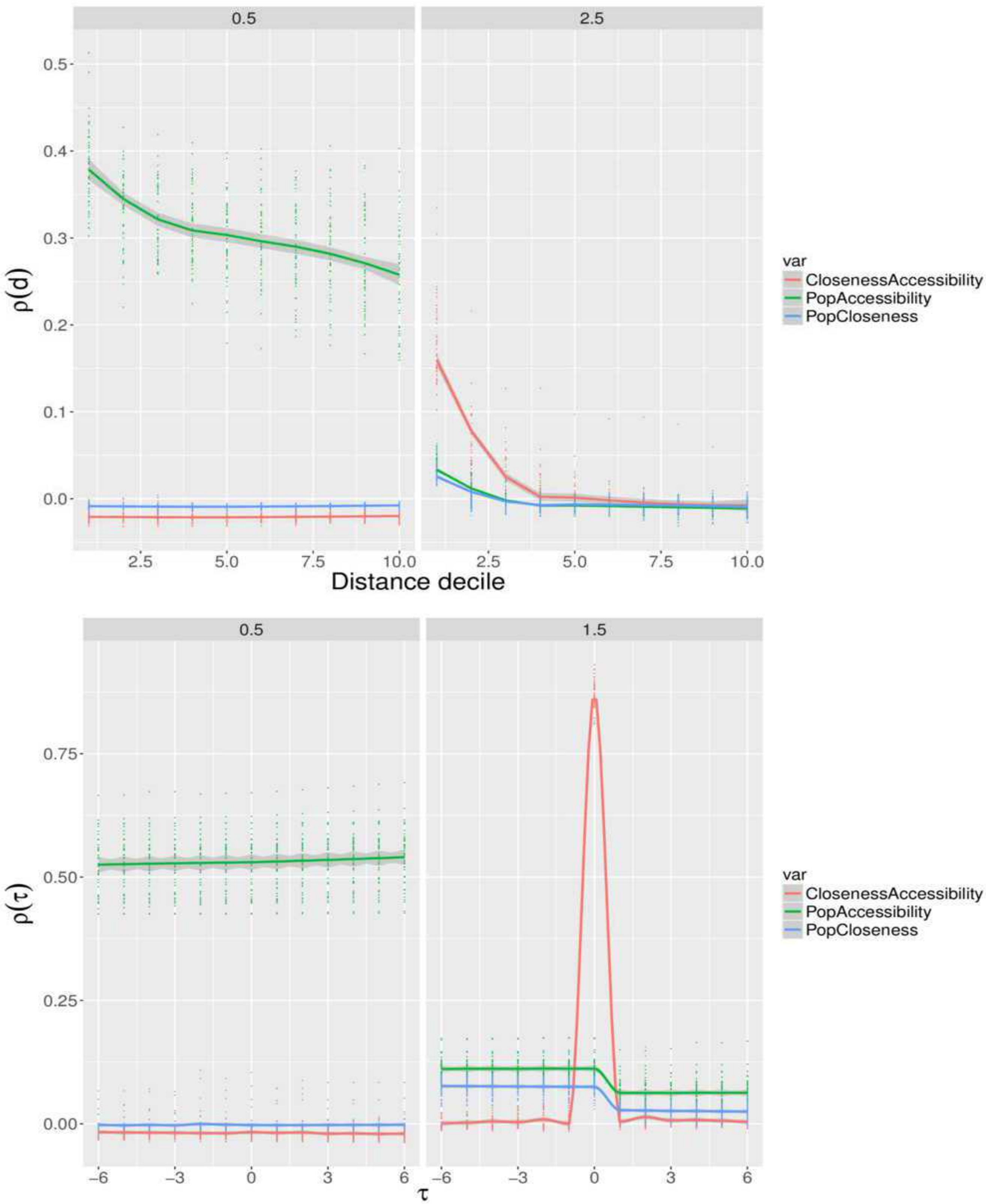


Figure 2: **Correlations in the model for the spatial configuration** $N_S = 80, \alpha_S = 0.5, d_S = 10, n_S = 30$**.** (*Top*) Correlations as a function of distance, for couples of variables (color), for $\gamma_N = 2.5$, $\theta_N = 21$, $v_0 = 10$, and for $d_G$ (columns) and $\gamma_G$ (rows) variables; (*Bottom*) Lagged correlations for the same parameters.