Journal of Urban Economics 111 (2019) 1–13

Contents lists available at ScienceDirect

Journal of Urban Economics

journal homepage: www.elsevier.com/locate/jue

About the origin of cities☆

André de Palma[a,*], Yorgos Y. Papageorgiou[b], Jacques-François Thisse[c,e,f], Philip Ushchev[d,e]

[a] ENS Cachan, CREST, University Paris-Saclay, France
[b] McMaster University, Hamilton, Ontario L8S 4K1, Canada
[c] CORE-UCLouvain, Belgium
[d] Monash University, Australia
[e] HSE University, Russian Federation
[f] CEPR, UK

ARTICLE INFO

JEL classification:
R12
R14

Keywords:
Central place
Spatial externality
Congestion
Urban hierarchy

ABSTRACT

We provide a bare–bones framework that uncovers the circumstances which lead either to the emergence of equally-spaced and equally-sized central places or to a hierarchy of central places. We show how these patterns reflect the preferences of agents and the efficiency of transportation and communication technologies. With one population of homogeneous individuals, the economy is characterized by a uniform distribution or by a periodic distribution of central places having the same size. The interaction between two distinct populations may give rise to a hierarchy of central places with one or several primate cities.

1. Introduction

Ever since (Christaller, 1933), economists, geographers, and regional scientists have shown that cities form hierarchical systems displaying some regularity (see, e.g., Gabaix and Ioannides, 2004). This begs the following two questions. First, what are the conditions for systems of cities and urban hierarchies to emerge in a featureless space? Second, how can this process be derived within a microeconomic model from the interplay between local and global interactions among agents pursuing their own interests? In this paper, we develop and use a simple framework in response to these challenging questions.

The term *central place* is used in this paper as shorthand for sites that display peaks in the spatial distribution of economic agents. Our aim is to show how a pattern of equally-spaced central places having the same size and a hierarchy of central places characterized by the existence of large and small cities, may emerge from a symmetry-breaking process (Matsuyama, 2008). To do this, we follow the main thread of spatial economics by assuming that the distribution of individuals is the outcome of the interplay between agglomeration forces and dispersion forces. However, we differ from the gallery of existing models in two major respects.

First, most existing models use a geometry characterized by some kind of boundedness condition and this matters for the nature of the equilibrium outcome. For example, Beckmann (1976) showed that land use and social interactions give rise to a bell-shaped distribution of individuals over a compact interval. His result is driven by this assumption as the peak tends to vanish when this interval becomes arbitrarily wide. Mossay and Picard (2011) revisit Beckmann's model when individuals are distributed over a circle and show that the equilibrium involves multiple outcomes, having any odd number of identical and evenly spaced cities. So, the choice of a particular geometry may critically affect the results. To rule this out, we assume away any "first-nature" asymmetries by working with the simplest possible seamless and unbounded space, i.e., the *real line*. The most natural way to think of this assumption is to suppose that land, which is locally scarce, is available everywhere. We then run the following thought experiment: what are the spatial patterns that emerge from the interaction among agents? More specifically, we determine the conditions that force a perfectly dispersed population

☆ We wish to thank G. Duranton, two referees, H. Koster, L. Limonov, F. Moizeau, P. Mossay, Y. Murata, W. Strange, M. Turner, and participants to the ITEA conference (Madrid 2017), SMU Conference on Urban and Regional Economics (Singapore 2018), Universities of Copenhagen and Roma 3, and the Interdisciplinary Leontief Centre for their comments. We are especially grateful to A. Baranov for his outstanding lectures in functional analysis. The first author would like to thank ANR (Elitisme) for financial support. The last two authors acknowledge the financial support of the Russian Science Foundation under the grant N°18-18-00253.

* Corresponding author.
E-mail addresses: andre.depalma@ens-cachan.fr, adepalma@riskdesign.com (A. de Palma), yorgos@mcmaster.ca (Y.Y. Papageorgiou), jacques.thisse@uclouvain.be (J.-F. Thisse), ph.ushchev@gmail.com, filipp.ushchev@monash.edu, fuschev@hse.ru (P. Ushchev).

https://doi.org/10.1016/j.jue.2019.01.006
Received 21 August 2018; Received in revised form 16 January 2019
Available online 14 March 2019
0094-1190/

to become a regular or a hierarchical pattern of settlements. In such a context, the spatial equilibria obtained are the solutions of Fredholm integral equations, a concept which does not belong to the toolbox of urban economists.[1]

Second, while most existing models emphasize specific effects which are seldom combined within the same framework, we focus on two general effects, namely, a *local congestion effect* and an *exponential decay externality*. The former depends on the local population density and the latter, which aims to capture different types of interaction across dispersed individuals such as flows of people or of information and knowledge, depends on the whole distribution of individuals over space. Since we do not specify the details of these two effects, we consider a bare–bones framework in which both effects are captured by *reduced forms* which are consistent with different narratives.[2] Although our setting has no explicit market, it is our contention that the interplay between the above two external effects takes on board several of the market and non-market interactions which are at work in urban systems.

In this paper, we build on Papageorgiou and Smith (1983) who maintained that the spatial distribution of identical individuals is determined by their attitudes regarding human interaction and by the properties of the spatial interaction field itself. Their analysis was limited to the necessary and sufficient conditions that render a uniform distribution of identical agents unstable. Our scope is much broader in that we find the necessary and sufficient conditions for the emergence of a flat or uneven population pattern in a setting where locations and flows are determined simultaneously. More specifically, in the case of one population of identical individuals, we show that, when individuals put a relatively high weight on spatial interaction as opposed to the local congestion effect, *the economic landscape is formed by the concatenation of identical human settlements*. How big are these areas and how packed are people depend on the attitude of agents toward congestion and human interaction. In particular, we are able to determine the conditions under which the population distribution is close or far away from the uniform distribution. Importantly, *the transition from the uniform distribution to an uneven distribution*, which is reminiscent of the early stages of urbanization, *is welfare-enhancing*. What is more, a growing population affects the equilibrium utility level according to the way the population is distributed. Individual welfare decreases when the equilibrium outcome involves a dispersed population because in this case the congestion effect is dominant. By contrast, the equilibrium utility level increases with the population size as long as central places exist.

We also show that an improvement in spatial interaction possibilities through new communication devices or more efficient transportation has a non-monotone impact on the density of central places. Initially, as the interactions among individuals become easier, their distribution takes the form of a denser system of central places. However, beyond some threshold, the additional gains associated with growing interactions are more than compensated by the costs generated by a rising congestion. In other words, technological devices that make interactions among individuals easier lead, first, to a more compact urban system and, then, to a sparser packing of central places. This is reminiscent of the bell-shaped curve of spatial development, which states that agents are dispersed when transport costs are either high or low and agglomerated when transport costs take on intermediate values (Krugman and Venables, 1995; Puga, 1999; Ottaviano et al., 2002). In those papers, like in ours, *there is more agglomeration when spatial frictions are neither too strong nor too weak*.

The most enduring problem in spatial economics is probably the existence of an urban system involving large and medium-sized cities, towns, and villages. To deal with this problem, we must extend our baseline model in which central places have the same size. A natural extension is to introduce heterogeneity across individuals by working with *two populations* such that individuals are homogeneous within each population but heterogeneous between populations, and to impose a complete spatial interdependence between the two populations. In this context, we show that *a hierarchy of tiered central places may emerge as an equilibrium outcome*. Furthermore, the share of each population in an urban area varies with its location. Thus, we may conclude that the central places are here the outcome of a richer pattern of interactions, which occur *between* and not only *within* populations, than in the case of a single population.

Depending on the type of interaction between the two populations and the parameters that characterize each population, the spatial economy may display a whole range of distinct patterns. In the first one, the spatial equilibrium involves two uniformly distributed populations. In this case, the interactions between the two populations are too weak for the blending of these populations to generate a steady-state that departs from what we observe with each population separately. The second pattern involves a periodic distribution. The third pattern leads to richer implications. Indeed, the superimposition of the two populations yields *an urban system in which one, two, or several central places are at the top of the urban system*. Although the distribution of the total population may exhibit one primate city and smaller cities whose size decreases with the distance to the top city, as argued by Christaller (1933) and Lösch (1940), the urban system may also involve several large cities, together with intermediate cities whose size need not decrease with the distance to the large cities. Two points are worth stressing here. First, the two populations must interact for a hierarchy of central places to emerge at the equilibrium outcome. Second, the two population distributions are interdependent and determined simultaneously. In this context, the equilibrium distribution of each population differs from the distribution that would emerge if this population were alone.

Thus, despite its great simplicity, our setting allows us to account for a rich set of urban patterns. Equally important, it provides a background for two of the main features of classical central place theory: (i) a pattern of equidistant and identical settlements and (ii) a hierarchy of cities. Our analysis also shows that there is no reason to expect the urban system to obey the rigid, pyramidal structure assumed by Christaller and Lösch. Instead, the urban system may involve two or several large urban agglomerations at the top of the urban hierarchy. It is also worth stressing that our setting differs from most urban system models, which assume away spatial frictions across cities whose locations are unspecified. To the best of our knowledge, no existing contribution has been able to generate a range of results comparable to ours within a microeconomic framework as parsimonious as ours.

Related literature. How to place our contribution in the field of spatial economics? Central place theory, based on the pioneering work of Christaller (1933) and Lösch (1940), aims to explain why economic activities are distributed over a system of cities in which the number of activities performed in a city rises with its size, and where cities having the same size are equally spaced (Mulligan et al., 2012). However, the bulk of classical central place theory has been directed towards identifying geometric conditions under which a superimposition of regular structures is possible. Here lies the Achilles Heel of the regional science approach to central place theory. These considerations are only interesting if they are based on solid microeconomic foundations. Unfortunately, such geometric analyses have not been rationalized through the choices made by optimizing agents.

New economic geography has the opposite problem. It relies on a full-fledged general equilibrium setup (Krugman, 1991; Fujita et al., 1999a). But since most models typically use a two-region setting, they do not permit to deal with the emergence of settlements and the formation of a hierarchy of settlements. There are exceptions, however. Fujita and Mori (1997) use an economic geography setting to show that cities are created at more or less equal distances when the total population grows in an economy with one manufacturing sector. Fujita et al. (1999b)

[1] Helsley (1990) is a noticeable exception.

[2] For surveys of the various agglomeration and dispersion forces used in the literature, see Duranton and Puga (2004); Fujita and Thisse (2013), and Behrens and Robert-Nicoud (2015).

extend this setting to a multisector economy and show how a hierarchy of cities may emerge. Tabuchi and Thisse (2011) also use an economic geography setting to prove that decreasing transportation costs foster the emergence of big cities supplying two goods, which coexist with small and specialized cities in which only one good is produced.

The spatial competition approach à la Hotelling has attracted less attention. Eaton and Lipsey (1982) appealed to multipurpose shopping as an economic foundation for the existence of clusters in which spatially dispersed consumers buy different goods available in dispersed shops. That consumers combine their purchases to reduce travel costs creates demand externalities which firms can exploit by locating close to firms selling other goods. Unfortunately, this approach becomes very quickly intractable. Indeed, with several goods, trip-chaining implies a complex structure of substitution between outlets. Hsu (2012) proposed to circumvent this difficulty by assuming that firms deliver their products to consumers. In this case, a consumer buys each product from the firm charging the lowest delivered price regardless of where the providers of the other goods are. Hsu showed that there exists an equilibrium that displays the hierarchical structure of central place theory, that is, all the goods available in a central place of a given order are also available in all central places of higher order. In these two papers, distance between buyers and sellers is the main driver.

In urban economics, (Henderson, 1974; 1988) has developed a compelling approach to describe the formation of urban systems in which cities are specialized. However, Henderson assumed that shipping goods between cities is costless. The same applies to more recent and richer models of urban systems such as (Behrens and Robert-Nicoud, 2015) and (Davis and Dingel, 2019). The various attempts made to explain the urban hierarchy through the Zipf Law also ignore where cities are located. This is not an innocuous assumption because cities are typically anchored in specific locations, while trade flows are gravitational in nature. Quantitative spatial models, such as (Diamond, 2016) and (Gaubert, 2018), provide a possible way out by combining within and between-cities frictions. However, these models aim to replicate real-world urban systems and to assess the impact of various shocks or counterfactuals.

Finally, a handful of papers seek to explain how one or several employment centers emerge within a city. They all rely on the trade-off between the benefits that firms enjoy when they are located close to each other and the higher wage firms must pay to workers to compensate them for the higher land rent and the longer commutes caused by the agglomeration of firms (Ogawa and Fujita, 1980; Fujita and Ogawa, 1982; Berliant et al., 2002; Lucas and Rossi-Hansberg, 2002. These papers share the same fundamental idea as us since workers compete for land while firms seek proximity. Like us, they consider gravity-like reduced forms in which spatial interactions are presumed. However, we pursue a different objective and use a different modeling strategy.

In the next section, we present our baseline model with one population of homogeneous individuals and discuss various interpretations of the agglomeration and dispersion forces. We then study the equilibrium outcome and determine the necessary and sufficient conditions for a regular pattern of identical settlements to emerge in a world which is otherwise featureless. In Section 3, we consider the case of two populations such that agents interact both within and between populations. The blending of the two distributions shows how a hierarchy of cities hosting agents belonging to both populations may come into being. Section 4 concludes.

2. Agglomeration in a homogeneous world

In this section, we aim to determine the conditions for a non-uniform distribution of identical individuals to be a stable outcome when individuals are free to choose where to locate.

2.1. The model

The economy features a continuum of identical individuals and a one-dimensional, featureless, and unbounded space. Therefore, ex post spatial differentiation (if any) is not driven by ex ante differences between locations or individuals. Each individual has a single choice variable, i.e., location. Formally, the location set is represented by the real line $\mathbb{R} \equiv (-\infty, \infty)$, with locations $x, y \in \mathbb{R}$. The real line is the simplest possible homogeneous space on which we may expect to obtain the emergence of a regular pattern of cities independently of any exogenous considerations.

Our aim is to determine how people distribute themselves across locations. Formally, we consider the set of *population distributions*, denoted by $n(x)$, which are defined by the non-negative, piecewise continuous, and bounded mappings from $\mathbb{R}$ to $\mathbb{R}_+$ with a finite mean $\bar{n}$:

$$\bar{n} \equiv \lim_{b\to\infty} \frac{1}{2b} \int_{-b}^{b} n(x)\mathrm{d}x < \infty.$$

Clearly, the uniform distribution, $n(x) = \bar{n}$ for all $x \in \mathbb{R}$, satisfies these conditions. That said, the following warning is in order: the analysis here is more complex than in standard models because we work in functional spaces, not in finite-dimensional spaces.

The utility $u(x)$ of an individual at x is linear in the following two variables:[3] the population $n(x)$ and a *spatial externality* $E(x)$ whose origin stems from interactions with other individuals:

$$u(x) = E(x) - \alpha n(x). \tag{1}$$

In line with urban economics we assume that the utility decreases with $n(x)$ because a higher number of residents amounts to reducing the individual lot size $1/n(x)$. In this case, α is positive and a high (low) value of α means that individuals focus more (less) on what is going on at the local level than on the various benefits generated by the interactions with the rest of the population.

In this paper, interaction across individuals is expressed through a spatial externality generated by the sum of each individual's interaction with all the others in $\mathbb{R}$. Following the literature, we assume that the interaction flow between individuals located respectively at x and y decays according to an exponential function $\exp\{-\beta|x-y|\}$ where $\beta > 0$ is an inverse measure of the spatial frictions that reduce the interaction among spatially dispersed individuals (Fujita and Ogawa, 1982; Lucas and Rossi-Hansberg, 2002; Desmet and Rossi-Hansberg, 2014).[4] Therefore, β is a spatial impedance parameter and the spatial externality at x is defined as

$$E(x) = \int_{\mathbb{R}} \exp\{-\beta|x-y|\} n(y)\mathrm{d}y > 0, \tag{2}$$

which varies with the population distribution. Since $n(\cdot)$ is piecewise continuous and bounded, $E(\cdot)$ is continuous and bounded. It is as if an individual were at the source of a flow which diffuses at a decreasing rate across space. At the same time, an individual at x receives an interaction flow emitted by those established at y, which is equal to $n(y)\exp\{-\beta|x-y|\}$. It follows that the bilateral interaction between x and y is given by the *exponential gravity* $n(x)n(y)\exp\{-\beta|x-y|\}$. Smith (1978) showed that this gravity rule is equivalent to a cost-efficiency principle, which states that a low-cost interaction pattern is more likely to be observed than a high-cost pattern.

That humans are "social animals" is perhaps the most basic justification of the need for interaction among individuals, while distance is an

[3] Using linear utility functions is fairly standard in urban economics. An example of such preferences is given by the following indirect utility: $w(x) - s(x)R(x)$, where $w(x)$ is the wage, $s(x)$ the land consumption, and $R(x)$ the land rent at x.

[4] Ogawa and Fujita (1980) show that using a linear decay function such as $a - \beta|x-y|$ simplifies the analysis. This does not hold true here because the equilibrium is the solution of a differential–difference equation which is difficult to solve.

impediment to interaction. Likewise, an individual often learns more by being close to a large pool of individuals. In this case, $\exp\{-\beta|x-y|\}n(y)$ describes the amount of information and knowledge transferred from the population at y to an individual established at x. Observe that $\exp\{-\beta|x-y|\}$ also has the nature of a distance-increasing iceberg transport cost (Fujita et al., 1999a; Rossi-Hansberg, 2005). In this case, $E(x)$ has the nature of a market potential of location x. Using the unavoidable CES, $E(x)$ is the aggregate consumption of a continuum of varieties differentiated by the places where they are produced, when the same quantity of each variety is shipped from its place of origin to its destination x (Head and Mayer, 2004).[5] For all these reasons, it is reasonable to view β as an inverse measure of the efficiency of the technology that permits the spatial externality to unfold across space. Examples include communication devices and transportation means.

We acknowledge that the general is preferable to the particular. However, it is impossible to pin down the spatial equilibrium without using a specific functional form for preferences. In this respect, a linear utility allows us to determine explicitly the equilibrium distribution for all admissible values of the parameters of the model. Furthermore, even though our setting involves no markets, it captures a fundamental feature of spatial economics: the most preferred location of an individual depends on where the others are set up. The specification (1) of preferences also captures the tension between two key ideas of urban economics, i.e., people prefer shorter trips to longer trips and more space than less space. Therefore, (1) may be viewed as a *reduced form* that captures the trade-off between the benefit generated by spatial interaction with the entire population and the congestion cost associated with the crowding out of locations. To illustrate, consider the example of a bell-shaped distribution, such as that obtained by Beckmann (1976). Individuals located near the maximizer of the distribution enjoy the benefits of a high potential of interactions because distances are on average shorter, but they also bear higher congestion costs associated with a denser local population.

2.2. The equilibrium population distribution

The population distribution $n^*(x)$ is a *spatial equilibrium* when the following three conditions hold:

(i) all individuals enjoy the same utility level u^*:

$$E^*(x) - \alpha n^*(x) = u^* \tag{3}$$

for any x such that $n^*(x) > 0$, where $E^*(x)$ is given by $E(x)$ evaluated at $n^*(x)$;

(ii) $n^*(x) = 0$ when

$$E^*(x) - \alpha n^*(x) < u^*;$$

(iii) the mean of $n^*(x)$ over $\mathbb{R}$ is equal to the constant $\bar{n} > 0$. The last condition amounts to the population constraint.

Before proceeding any further, it is worth mentioning the following properties of a spatial equilibrium: First, as in many economic geography models, *the uniform distribution* $n(x) = \bar{n}$ *is always a spatial equilibrium* because in this case the spatial externality is constant and equal to $2\bar{n}/\beta$. However, this equilibrium need not be unique and, if not, it need not be stable.

Second, although all individuals are mobile, *the spatial equilibrium involves no unpopulated areas*, unlike what we observe in the core-periphery model with land (Ottaviano et al., 2002). More precisely, there exists no spatial equilibrium such that $n^*(x) = 0$ over an interval $[x_1, x_2]$ with $x_1 < x_2$. The proof is tedious and relegated to the Supplementary Material (Result 1). Intuitively, if individuals care more about congestion than about interaction, an unpopulated area should be attractive to these people. Conversely, if individuals value interaction more than congestion, the existence of an empty area reduces the level of interaction, which is not compensated by a nil congestion effect.

Last, any equilibrium distribution $n^*(x)$ is *continuous* over $\mathbb{R}$. Indeed, assume that $n^*(x)$ jumps downwards at $x = x_0$, which implies that $n^*(x) > 0$ over some interval $(\hat{x}, x_0)$ with $\hat{x} < x_0$. Over this interval, $u(x) = u^*$ must hold by definition of a spatial equilibrium. However, since $E^*(x)$ is continuous, the right-hand side of the equilibrium condition $E^*(x) = u(x) + \alpha n^*(x)$ is also continuous. Hence, the utility level $u(x)$ must jump upwards at $x = x_0$, implying $u(x) > u^*$ over some interval $(x_0, \bar{x})$ with $x_0 < \bar{x}$, a contradiction. An upward jump of $n^*(x)$ is ruled out along the same lines. Consequently, without loss of generality, we may focus on continuous distributions.

Our results are driven by the interplay between the parameters α and β. In addition, the parameter:

$$\phi \equiv \frac{2}{\alpha\beta} > 0 \tag{4}$$

is often sufficient to characterize the equilibrium. Roughly speaking, ϕ blends the two forces that affect individuals' well-being; ϕ is low when the spatial decay factor β is high – and thus the spatial externality very localized –, the marginal disutility of congestion α is high, or both. Intuitively, a low value of ϕ is associated with an economy in which individual individuals "think and act locally," while a high value of ϕ is associated with individuals who are globally oriented. The communication or transportation technology and the attitude of individuals toward local and global interactions are critical for the nature of the spatial equilibrium.

Combining (3) with (2) and (1) shows that any solution to

$$n(x) = \frac{1}{\alpha}\left[\int_{\mathbb{R}} \exp\{-\beta|x-y|\}n(y)\mathrm{d}y - u^*\right] \tag{5}$$

is a spatial equilibrium, where u^* is an unknown constant. The expression (5) is known as a *Fredholm integral equation of the second kind*: the unknown function $n(x)$ appears inside and outside the integral (Kolmogorov and Fomin, 2012, ch. 2, p. 74). In what follows, we determine the closed-form solution of this equation.

The great merit of economic geography is to explain the emergence of agglomeration as a symmetry-breaking mechanism. This mechanism relies on the interplay between increasing returns in production, transportation costs, and preference for variety in a two-region setting (see Baldwin et al., 2003, for a synthesis of the main results). We adopt the same approach in the following proposition.

Proposition 1. *Assume the utility function (1) and the spatial externality (2). The uniform distribution is the unique spatial equilibrium if and only if* $\phi < 1$. *Furthermore, this equilibrium is stable.*

We give here a sketch of the proof; details are relegated to Appendix A. When $\phi < 1$, the integral operator in the right-hand side of (5) is a contraction mapping. Therefore, the Banach fixed-point theorem implies that the uniform distribution is the unique solution to (5) (Kolmogorov and Fomin, 2012, ch. 2, p. 66). Regarding our concept of stability, the reader is also referred to Appendix A.

Proposition 1 shows that, when the communication technology is inefficient (β is high), or individuals are much affected by local congestion (α is high), or both ($\phi < 1$), *the world is flat*. In other words, there is no agglomeration in a world populated by locally-oriented individuals. This echoes what we observe in the core-periphery model: when transportation costs are high, workers are uniformly distributed between the two regions while firms produce mainly for their regional market. But what happens when $\phi \geq 1$? The answer is given below.

Proposition 2. *Assume the utility function (1) and the spatial externality (2). If* $\phi > 1$, *all non-uniform equilibria are given by*

$$n^*(x) = \bar{n} + A\sin\left(\beta\sqrt{\phi - 1}\,x\right), \tag{6}$$

[5] There exists a rich literature on social networks, which could be used to provide a wider range of justifications for what we call an interaction field (see, e.g., Jackson et al., 2017).

where A is an arbitrary positive constant such that $A \le \bar{n}$.[6]

The formal argument goes as follows (see Appendix B for more details). When $\phi > 1$, the Eq. (5) is equivalent to a special case of the Sturm-Liouville problem in which the differential operator is $-\mathrm{d}^2/\mathrm{d}x^2$ (Dunford and Schwartz, 1988, ch. XIII, p. 1291). The eigenfunctions of this operator are given (up to adding $\bar{n}$) by the right-hand side of (6).

It remains to study the borderline case where $\phi = 1$. Taking pointwise the limit under $\phi \to 1_+$ on the right-hand side of (6) shows that, in this case, $n^*(x) = \bar{n}$ is the unique spatial equilibrium.

The intuition behind Proposition 2 is as follows. When individuals become more globally oriented, the world ceases to be flat. Indeed, when β is sufficiently small, the spatial externality allows using a linear approximation: $\exp\{-\beta|x-y|\} \approx 1 - \beta|x-y|$. When β becomes even smaller, the range of spatial interaction – i.e., the domain where $1 - \beta|x-y| > 0$ – widens and tends to infinity, which captures the idea that spatial interaction changes its nature from local to global. Consequently, some locations are more populated than others. Thus, like in Krugman (1991), the process of agglomeration is the result of a symmetry-breaking mechanism that operates when the cost of interaction across space is relatively low. A major difference with his work is that the location set which sustains the interaction flows involves here a continuum of locations. Furthermore, the emergence of agglomeration depends on the relative value of two parameters, i.e., α and β. In other words, the individual attitude toward land consumption also matters for the nature of the spatial equilibrium. This is reminiscent of Helpman (1998) and Tabuchi (1998) who show that introducing land consumption in the core-periphery model may reverse Krugman's conclusions.

In what follows, we study the properties that hold for all spatial equilibria. The population distribution (6) is maximized at the *central places* given by

$$x^k_{\max} = \pm\left(\frac{1}{2} + 2k\right)\frac{\pi}{\beta}\sqrt{\frac{1}{\phi - 1}} \qquad k = 0, 1, .$$

It is minimized at

$$x^k_{\min} = \pm\left(\frac{3}{2} + 2k\right)\frac{\pi}{\beta}\sqrt{\frac{1}{\phi - 1}} \qquad k = 1, 2, \dots,$$

and equal to the mean $\bar{n}$ at

$$\bar{x}^k = \pm(1 + 2k)\frac{\pi}{\beta}\sqrt{\frac{1}{\phi - 1}} \qquad k = 0, 1, \dots$$

Since $|x^k_{\max}|$ is arbitrarily large when ϕ is arbitrarily close to 1_+, the central places are pushed toward $+\infty$ or $-\infty$ when $\phi \to 1_+$, which is consistent with Proposition 1. Moreover, the transition from the uniform pattern to a landscape with central places is continuous, while it is discontinuous in Krugman (1991).

The limit of the urban area is reached at the distance $\left|x^k_{\max} - x^k_{\min}\right|$ from the central place. When the distance exceeds this threshold, the population distribution starts rising and takes again its highest value at the next central place. To be precise, the peaks and troughs of the spatial distribution of individuals are distributed with a period equal to

$$T = \frac{2\pi}{\beta}\sqrt{\frac{1}{\phi - 1}}.$$

Therefore, the population distribution displays periodically distributed oscillations. The period T, hence the location of population peaks, is independent of A while the overall shape of the equilibrium distribution remains the same for all $A \in (0, \bar{n}]$.

There exists a continuum of spatial equilibria. More specifically, one equilibrium is associated with each admissible value of A. The constant A plays here the role of the constant of integration in the solution to a differential equation. What changes with the value of A is the amplitude of the oscillations of the equilibrium distribution, that is, the size of central places. Since the least populated areas in the real world have no or little activity ($n^*(x) \approx 0$), we find it intuitively plausible to focus on the case $A \approx \bar{n}$. This implies that the population is close to $2\bar{n}$ in the vicinity of the peaks of $n^*(x)$. This intuitive argument turns out to be remarkably consistent with using stability as an equilibrium selection device. To be precise, we have the following result (see Appendix C for the proof).

Proposition 3. *Assume the utility function (1) and the spatial externality (2). If $\phi > 1$, then the equilibrium (6) is stable when $A = \bar{n}$ and unstable otherwise.*

In other words, when $\phi > 1$, there exists a unique stable spatial equilibrium given by

$$n^*(x) = \bar{n}\left[1 + \sin\left(\beta\sqrt{\phi - 1}\, x\right)\right]. \tag{7}$$

Because (7) describes a periodic function, *the economic landscape is formed by a succession of identical monocentric cities whose population size is equal to $\bar{n}T$*, while *the distance between central places is equal to T*. The population distribution at a central place $x^k_{\max}$ is equal to $2\bar{n}$. As a result, changing the parameters α or β affects the population and physical sizes of cities. That space is unbounded allows the period T to respond to all shocks without affecting the population peak $2\bar{n}$. As suggested by the work of Beckmann (1976) and Mossay and Picard (2011), this ceases to hold when space is bounded. Furthermore, *the population distribution decreases around every central place $x^k_{\max}$*, a pattern that concurs with what urban economics predicts. The individual lot size at x depends on the amount of land available. If the land supply is equal to 1, the lot size at x, which is given by $1/n^*(x)$, increases with the distance to the nearest central place, which also agrees with the monocentric city model. As expected, the population size of cities increases with the population mean, but their physical size T is independent of $\bar{n}$. Therefore, a higher mean leads to a configuration in which individuals enjoy a smaller lot size and a higher level of spatial interaction that keep T unchanged.

The above results are illustrated by Fig. 1, where we have set $x^0_{\max} = 0$ without loss of generality. The dotted horizontal line represents the uniform distribution $n(x) = \bar{n}$; the brown and red lines describe two unstable equilibria, while the blue line represents (7). The period T is the distance between two adjacent extrema of n^*.

The distance between two central places decreases when local congestion matters less to individuals (α decreases). The impact of the spatial decay factor β is more involved. Differentiating T with respect to β shows that the distance between two neighboring central places first decreases and, then, increases when β increases from 0 to $2/\alpha$; the turning point is reached at $\beta = 1/\alpha$. Indeed, when β is very large (or very small), individuals focus on their local environment (or benefit more or less equally from the others regardless of their locations). In both cases, an individual cares less about where the others are located, but for different reasons. By contrast, when β takes on intermediate values, the intensity of the spatial externality depends strongly on the whole distribution of individuals. In short, *the packing of activities is the densest, and the size of cities the smallest, for intermediate degrees of efficiency of the communication or transportation technology*.

Since T is U-shaped in β, the process of urbanization associated with a more efficient communication or transportation technology involves two opposite phases. In the first one, cities get more densely packed; individuals use less land, but interact more with people residing in the other cities. In the second phase, cities are more sprawled; individuals use more land and keep interacting more with the other cities because the spatial impedance is now sufficiently weak for the value of the externality to be high everywhere in a city. This concurs with the bell-shaped curve of spatial development of economic geography according to which the concentration of activities would go hand in hand with the first stage of economic integration; as transportation cost keeps decreasing, there

[6] Without loss of generality, we may normalize to 0 the phase of the function $\sin(\cdot)$ because the choice of the origin is arbitrary.

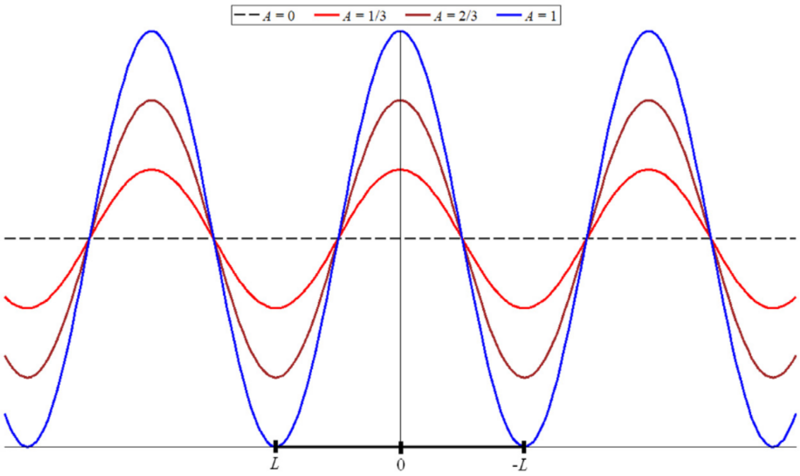

Fig. 1. Population densities in a homogeneous world.

would be a spatial redeployment of activities, as shown by Krugman and Venables (1995); Puga (1999), and Ottaviano et al. (2002) in various settings richer than Krugman's core-periphery model.

How does welfare vary with the average population size and the nature of the equilibrium? The answer is given by the next proposition proven in Appendix D.

Proposition 4. *Assume the utility function (1) and the spatial externality (2). For all values of ϕ, the equilibrium utility level is given by*

$$u^* = \bar{n}\left(\frac{2}{\beta} - \alpha\right) = \bar{n}\alpha(\phi - 1). \tag{8}$$

Thus, the equilibrium utility level u^* rises with ϕ. In other words, improving upon the efficiency of the communication or transportation technology always makes people better-off. Furthermore, when $\phi < 1$, (8) implies that the utility level decreases when the average population grows. Before the onset of permanent settlements people remained dispersed in small hunter–gatherer bands because land could not support their way of life in large numbers. In this context, demographic growth is detrimental to people because the congestion effect is strong and/or the externality effect is weak.

With the passage of time, people gradually abandoned their nomadic life because food production generated higher returns per unit of land. Growth of social awareness, culture, kinship, and religion in early humans increased the need for spatial interaction (extended the range of the spatial externality) and reduced the local effects of proximity until $\phi > 1$. Then, spatial uniformity was replaced by regular agglomerations representing the onset of early human settlements (the uniform distribution is no longer stable). Such settlements were forming by homogeneous populations well before 10,000 BC. In other words, when β became smaller than $2/\alpha$, from the historical point of view, *a seamless world organizes itself spontaneously as a system of identical central places*.

Moreover, since u^* decreases with α, a shock that makes individuals more prone to live together (α decreases) leads to a higher welfare level. Similarly, a shock that facilitates spatial interaction (β decreases) also triggers a higher equilibrium utility level. Furthermore, when $\phi - 1$ becomes positive, urbanization arises and goes together with a rise in welfare. In other words, *any shock that leads individuals to agglomerate is also welfare-enhancing*. In this case, an increase in population is beneficial to all individuals; otherwise demographic growth is harmful to people.

Food production in prehistoric times gradually accumulated food surpluses which, in turn, gradually freed some individuals from food production, created occupational specialties and accelerated social organization away from the roughly egalitarian norm of hunter–gatherer groups toward more complex schemes. Representing *specialization*, the mother of cities, requires at least two different populations. This is what we do in the next section, where we discuss how various urban hierarchies can emerge if we expand the current framework to account for two, rather than one, populations.

3. Hierarchy in a heterogeneous world

In this section, we show that hierarchy may emerge in the above setting when we consider two, rather than one, populations. In addition, we will see that alternative configurations may also arise at the equilibrium.

3.1. The model

Consider two populations specialized in different activities, which are distributed over $\mathbb{R}$ with densities $n_1(x)$ and $n_2(x)$, respectively. Both means $\bar{n}_1$ and $\bar{n}_2$ are finite. In this economy, the spatial distribution of both populations concerns everyone. This is reflected in the structure of the utility function of a j-type individual, $j = 1, 2$, which is given by:

$$u_j(x) = \gamma_{jj}E_{jj}(x) + \gamma_{jk}E_{jk}(x) - \alpha_{jj}n_j(x) - \alpha_{jk}n_k(x). \tag{9}$$

In (9), $E_{jk}(x)$ is the spatial externality a j-type individual receives from k-type agents:

$$E_{jk}(x) \equiv \int_{-\infty}^{\infty} \exp\{-\beta_{jk}|x-y|\}n_k(y)\mathrm{d}y, \qquad j,k = 1,2, \tag{10}$$

where $\beta_{jk} > 0$ are the spatial decay factors specific to j-type agents when they interact with k-type agents, where j may be equal to or different from k. In other words, the preferences (9) account for interactions within and between populations.

One may think of different types of interaction between the two populations. For example, $\alpha_{jk} > 0$ means that a j-type agent suffers from the congestion associated with the fact of sharing the same local environment with the k-type agents. However, since α_{jk} is specific to each population, the two populations can generate various crowding effects. To illustrate, consider the following two examples in which the parameters α_{jk} have different signs. First, let us refer to the agents as *consumers* (type 1) and *firms* (type 2). In this case, consumers are attracted to places where the number of firms is high because there are more and better opportunities ($\alpha_{12} < 0$), while firms are attracted by places where consumers are numerous because there the expected volume of business is large ($\alpha_{21} < 0$); consumers are repulsed by places where the number of consumers is high because they dislike congestion ($\alpha_{11} > 0$), while firms dislike places where there are many firms because competition in space is very localized ($\alpha_{22} > 0$). Second, we may also capture segregation behaviors between two groups of individuals, based on race or religion, by assuming $\alpha_{jj} < 0$ and $\alpha_{jk} > 0$, that is, j-type agents are attracted by their peers, but repelled by the members of the other population. Thus, depending on the sign of the parameters α_{jk}, (9) is able to account for a rich set of local interaction patterns.

Regarding the global interaction parameters, (Jacobs, 1969) suggests that members of a population learn from the other population, which implies $\gamma_{jk} > 0$ for $j \neq k$. When population members are firms belonging to different industries, assuming $\gamma_{jk} > 0$ means that "urbanization economies" are at work. When $\gamma_{jj} > 0$, we have "localization economies". Both concepts have attracted a lot of attention in modern studies of agglomeration economies.

Let $\mathbf{A}$ be the (2×2)-matrix

$$\mathbf{A} \equiv \begin{pmatrix} \alpha_{11} & \alpha_{12} \\ \alpha_{21} & \alpha_{22} \end{pmatrix} \tag{11}$$

of congestion disutilities. Without much loss of generality, we assume throughout the paper that this matrix is invertible, i.e., $\det(\mathbf{A}) \equiv \alpha_{11}\alpha_{22} - \alpha_{12}\alpha_{21} \neq 0$. On the other hand, it is reasonable to assume that each agent (weakly) benefits from interacting with the others, thus meaning $\gamma_{jk} \geq 0$ and $\gamma_{jj} \geq 0$ for $j, k = 1, 2$. If j-agents do not value interacting with k-agents, we have $\gamma_{jk} = 0$.

Very much like in the case of one type of agents, the equilibrium densities of agents must satisfy the following equilibrium conditions:

$$u_j(x) = u_j^*, \qquad j = 1, 2, \tag{12}$$

where the equilibrium utility levels u_j^*, $j = 1, 2$, are two unknown constants.

Combining the equilibrium conditions (12) with (9) and (10) and taking the deviations $\widetilde{n}_j(x) \equiv n_j(x) - \bar{n}_j$ from the mean on both sides, we obtain *a system of two Fredholm integral equations*:

$$\widetilde{n}_j(x) = \int_{\mathbb{R}} g_{jj}(x, y)\widetilde{n}_j(y)\mathrm{d}y + \int_{\mathbb{R}} g_{jk}(x, y)\widetilde{n}_k(y)\mathrm{d}y, \qquad j, k = 1, 2, \quad j \neq k, \tag{13}$$

where the kernels $g_{jk}(x, y)$ are defined as follows:

$$\begin{pmatrix} g_{11}(x, y) & g_{12}(x, y) \\ g_{21}(x, y) & g_{22}(x, y) \end{pmatrix} = \mathbf{A}^{-1}\begin{pmatrix} \gamma_{11} & \gamma_{12} \\ \gamma_{21} & \gamma_{22} \end{pmatrix} \times \begin{pmatrix} \exp\{-\beta_{11}|x-y|\} & \exp\{-\beta_{12}|x-y|\} \\ \exp\{-\beta_{21}|x-y|\} & \exp\{-\beta_{21}|x-y|\} \end{pmatrix}.$$

Note the difference between (13) and (5), as the latter only involves $n(x)$. By contrast, in (13), $n_j(x)$ also depends on $n_k(x)$. Hence, the distributions of the two populations are fully interdependent. The above expressions show that the spatial equilibrium depends on the nature and intensity of the *local* and *global* interactions *within* and *between* populations. We aim to determine what the economic landscape becomes when we superimpose the two distributions to form the distribution of the total population $n_1(x) + n_2(x)$. Note also there are two-way interactions between any two central places. Horizontal relations among central places are thus superimposed on the pyramidal structure of the urban system, like in Fujita et al. (1999b).

Repeating the argument of Appendix D, it is readily verified that

$$\bar{E}_{jk} = \frac{2}{\beta_{jk}}\bar{n}_k, \qquad j, k = 1, 2, \ j \neq k. \tag{14}$$

Computing the mean of both sides of (9), we obtain:

$$u_j^* = \gamma_{jj}\bar{E}_{jj} + \gamma_{jk}\bar{E}_{jk} - \alpha_{jj}\bar{n}_j - \alpha_{jk}\bar{n}_k, \qquad j, k = 1, 2, \ j \neq k. \tag{15}$$

Combining (14) and (15) yields *the equilibrium welfare level of j-type individuals*:

$$u_j^* = \left(2\frac{\gamma_{jj}}{\beta_{jj}} - \alpha_{jj}\right)\bar{n}_j + \left(2\frac{\gamma_{jk}}{\beta_{jk}} - \alpha_{jk}\right)\bar{n}_k, \qquad j, k = 1, 2, \ j \neq k. \tag{16}$$

As in Section 2, the equilibrium welfare level u_j^* of a j-type agent is the same at all spatial equilibria. The expression (16), which is an extension of (8) to the case of two populations, has the following implication: depending on the values of the various parameters, the two population means $\bar{n}_1$ and $\bar{n}_2$ have a positive or a negative impact on u_1^* and u_2^*. Indeed, (16) implies that the impact of an increase in $\bar{n}_j$ on u_k^* is positive if and only if the inequality

$$2\frac{\gamma_{jk}}{\beta_{jk}} > \alpha_{jk}$$

holds. This is so when the j-type individuals value much their interactions with the k-type individuals (γ_{jk} is large), when they have developed an efficient communication or transportation technology (β_{jk} is low), when the crossed congestion effect between the two populations is weak (α_{jk} is low), or when any combination of these conditions holds. On the other hand, if the opposite inequality holds, we have

$$\frac{\partial u_j^*}{\partial \bar{n}_k} < 0$$

because the congestion effect generated by the k-type individuals overcomes the benefit made from interacting with these individuals.

As in the one-population case, the uniform distributions $n_j^*(x) \equiv \bar{n}_j$, $j = 1, 2$, always satisfy the equilibrium conditions (12) because the spatial externalities E_{jj} and E_{jk} are constant. However, the uniform equilibrium is not necessarily unique and stable.

3.2. Equilibrium in an asymmetric world

The general pattern of interactions associated with preferences (9) and externalities (10) is too complex to allow for an intuitively appealing characterization such as the one obtained in Section 2. Indeed, in the asymmetric case, the number of parameters has increased from 2 to 10 (without loss of generality, for each population one parameter may be normalized to 1). In this case, solving the system (13) is a hard task. Nevertheless, the following result provides a classification of possible spatial equilibria. For this, we need the following definitions.

Let $\mathbf{D}$ be a (4×4)-matrix independent of x defined as follows:

$$\mathbf{D} \equiv \mathbf{Q} - 2\mathbf{B}\mathbf{A}^{-1}\mathbf{\Gamma},$$

where

$$\mathbf{Q} \equiv \begin{pmatrix} \beta_{11}^2 & 0 & 0 & 0 \\ 0 & \beta_{12}^2 & 0 & 0 \\ 0 & 0 & \beta_{21}^2 & 0 \\ 0 & 0 & 0 & \beta_{22}^2 \end{pmatrix}, \qquad \mathbf{B} \equiv \begin{pmatrix} \beta_{11} & 0 \\ 0 & \beta_{12} \\ \beta_{21} & 0 \\ 0 & \beta_{22} \end{pmatrix},$$

$$\mathbf{\Gamma} \equiv \begin{pmatrix} \gamma_{11} & \gamma_{12} & 0 & 0 \\ 0 & 0 & \gamma_{21} & \gamma_{22} \end{pmatrix},$$

while $\mathbf{A}$ is given by (11). Therefore, the matrix $\mathbf{D}$ is obtained by combining in a certain way *all* the parameters of (9) and (10).

The following result is proven in Appendix E.

Proposition 5. *Assume the utility functions (9) and the spatial externalities (10).*

(i) If $\mathbf{D}$ has no strictly negative real eigenvalue, then the spatial equilibrium is unique and given by the uniform distributions

$$n_j^*(x) \equiv \bar{n}_j, \quad j = 1, 2;$$

(ii) if $\mathbf{D}$ has one strictly negative real eigenvalue λ, then the non-uniform spatial equilibria are periodic and given by

$$n_i^*(x) = \bar{n}_i + A_i \sin(\sqrt{-\lambda}\, x), \quad i = 1, 2$$

where A_i is a positive constant such that $A_i \leq \bar{n}_i$;

(iii) if $\mathbf{D}$ has at least two strictly negative real eigenvalues, $\lambda_1 < \lambda_2 < 0$, then the spatial equilibria are given by:

$$n_i^*(x) = \bar{n}_i + A_{i1}\sin(\sqrt{-\lambda_1}\, x) + A_{i2}\sin(\sqrt{-\lambda_2}\, x + \varphi), \quad i = 1, 2$$

where A_{11}, ..., A_{22} are constants such that $0 \leq A_{i1} + A_{i2} \leq \bar{n}_i$ for $i = 1, 2$, while $\varphi \in [0, 2\pi)$.[7]

[7] Like in Proposition 2, we have normalized the first phase to zero. We cannot do the same for the second one. However, the value of φ is irrelevant for our interpretation of the proposition.

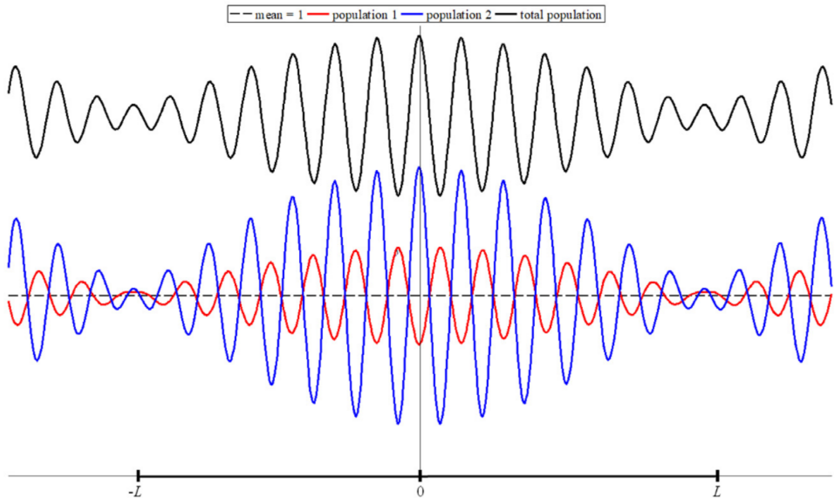

Fig. 2. Christaller-like population distributions.

In case (i), proving the stability of the uniform distributions $n_j^*(x) \equiv \bar{n}_j$ is a rather complicated task. However, this equilibrium is stable under the following non-pathological restrictions. Assume that $\alpha_{11} = \alpha_{22} = \alpha_o$ (subscript o for "own") and $\alpha_{12} = \alpha_{21} = \alpha_c$ (subscript c for "cross"). We capture the idea that likes are less bothered by likes by assuming that $\alpha_c \geq \alpha_o \geq 0$. Furthermore, we assume that individuals learn equally from all neighbors, that is, β_{jk} and γ_{jk} are the same for all combinations of $j, k = 1, 2$: $\beta_{jk} = \beta > 0$ and $\gamma_{jk} = \gamma > 0$. We show in the Supplementary Material (Result 2) that *the uniform equilibrium is stable if and only if no other equilibria exist.*[8]

In case (ii), the total population is distributed according to $n_1^*(x) + n_2^*(x) = \bar{n}_1 + \bar{n}_2 + (A_1 + A_2)\sin(\sqrt{-\lambda}\, x)$. This case is comparable to Proposition 2 where here the distributions $n_1^*(x)$, $n_2^*(x)$ and $n_1^*(x) + n_2^*(x)$ reach their local maxima at the same locations. By contrast, in case (iii), the two population distributions do not reach their local maxima at the same locations. As illustrated by Figs. 2 and 3, this implies that *the central places have different sizes and form a hierarchy within the interval* $[-L, L]$ whose length is endogenous. Furthermore, the population shares, $n_i^*(x)/(n_1^*(x) + n_2^*(x))$, vary with x, meaning that the composition of central places changes with their locations.

Though the matrix **D** is difficult to study, Proposition 5 provides a useful guide to obtain numerically the spatial equilibrium when the values of the parameters are specified. Indeed, this proposition shows that characterizing the eigenvalues of the matrix **D** is sufficient to determine the nature of spatial equilibria. To illustrate, we consider the following two examples.

Fig. 2 is drawn for a set of parameter values which describe two totally asymmetric populations: $\alpha_{11} = 1.7$, $\alpha_{12} = 1.6$, $\alpha_{21} = 1.2$, $\alpha_{22} = 1.4$; $\beta_{11} = 0.25$, $\beta_{12} = 0.75$, $\beta_{21} = 0.4$, $\beta_{22} = 0.3$; $\gamma_{11} = 1$, $\gamma_{12} = 0.85$, $\gamma_{21} = 1.15$, and $\gamma_{22} = 1$.[9] The spatial equilibrium involves *a primate city* located at $x = 0$ over the interval $[-L, L]$ whose length is endogenous. As the distance to this city increases, the distribution $n_k^*(x)$ swings around the mean $\bar{n}_k = 1$, for $k = 1, 2$. More specifically, the oscillations of $n_k^*(x)$ dampen and converge to 0 as $|x|$ increases. As for the least populated places, their size rises with the distance to $x = 0$. The pattern displayed in Fig. 2 resembles the regional section of a megalopolis, where peak densities decline away from a metropolitan area until they begin to rise toward the next. The population level is below the mean for locations close to the middle point between two adjacent central places because the corresponding places are in the shadow of the bigger cities.

Another illustration is given by Fig. 3 where we borrow the values used to build Fig. 2, apart from $\beta_{11} = \beta_{22} = 0.5$ and $\beta_{12} = \beta_{21} = 1.5$. In other words, the spatial decay effect is now stronger between, rather than within the two populations, which seems plausible. In this case, we obtain a pattern that vastly differs from that of Fig. 2 since five cities have almost the same largest size, e.g., the five most important German metropolitan areas. There are also several smaller cities having different sizes. The global pattern, defined over the interval $[-L, L]$, involves 11 central places. This pattern keeps being repeated along the sequence of intervals $[-L + 2kL, L + 2kL]$ for $k = \pm 1, 2, \ldots$.

These examples are sufficient to show that our setting, despite its disarming simplicity, may generate a broad gallery of hierarchical urban systems. Depending on the intensities of the forces at work, diverse spatial patterns may emerge. This reflects the variety of urban systems we observe in the real world. In particular, our model shows that the possible equilibrium hierarchies of central places involve much richer structures than the one hypothesized by Christaller (1933) and Lösch (1940), which arises only for specific values of the parameters. Furthermore, the urban system defined by $n_1^*(x) + n_2^*(x)$ is symmetric about $x = 0$ over the interval. That a large number of urban systems exist over the real line should not be treated as a surprise since our space is unbounded. Last, the size of lower order central places does not decrease monotonically with the distance to the primate cities: small cities may arise between two large cities.

We close this section by interpreting what we have in a way that corresponds to classical central place theory. Hierarchy here refers to a spatial system of different city sizes. By contrast, hierarchy in central place theory refers to a strict ordering of central places such that all the goods available in a central place of a given order are also available in all central places of higher order. Our results lead us to predict that spatial arrangements of higher order in our framework will exhibit a complex variety of hierarchical structures that come closer to the flexibility we experience in the real world. In this sense, our paper may be viewed as a primer to central place theory.

3.3. Equilibrium in a symmetric world

Since the assumptions of Proposition 5 are difficult to interpret, we consider the special case of a symmetric setting: (i) the preferences for local interactions are the same between populations, i.e., $\alpha \equiv \alpha_{11} = \alpha_{22} > 0$ and $\alpha_{12} = \alpha_{21} = 0$; (ii) the intensity of the spatial externality both within and between populations is the same, i.e., $\beta \equiv \beta_{jk} > 0$, $j, k = 1, 2$; and (iii) the preferences for global interactions are the same within each population, i.e., $\gamma_{11} = \gamma_{22} = 1$, and the same between the two populations, i.e., $\gamma_{12} = \gamma_{21} = \gamma \in (0, 1)$. Thus, our symmetric setting involves only three parameters, i.e., α, β, and γ.

As in Section 2, we set $\phi \equiv 2/(\alpha\beta)$. The following result is proven in Appendix F.

[8] Note that this result is consistent with Proposition 1 obtained in the case of a single population.

[9] Note that γ_{11} and γ_{22} have been normalized to 1 without loss of generality.

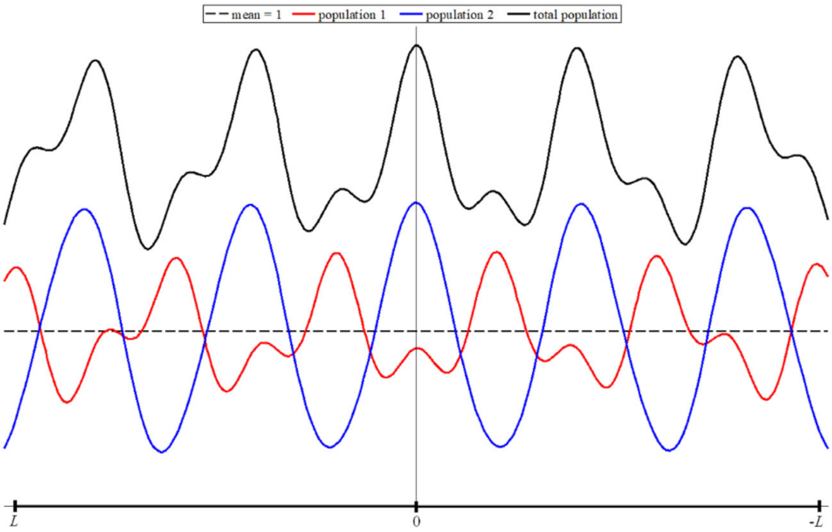

Fig. 3. The "Big Five" in a heterogeneous world.

Proposition 6. *Assume the utility functions (9) and the spatial externalities (10). Then, we have:*

(i) if $\phi \leq 1/(1+\gamma)$, there exists a unique spatial equilibrium. This equilibrium is such that both populations are uniformly distributed across space;

(ii) if $1/(1+\gamma) < \phi \leq 1/(1-\gamma)$, the non-uniform equilibria are periodic, and the central places have the same size;

(iii) if $\phi > 1/(1-\gamma)$, the non-uniform equilibria involve central places of different sizes.

Thus, we need two distinct and interacting populations for a hierarchy of central places to emerge. When there is no interaction, i.e., $\gamma = 0$, there is only one strictly negative eigenvalue (see the proof of Proposition 6). In this case, as shown by Proposition 5, the configuration (iii) does not arise. Proposition 6 also shows that a hierarchy across central places emerges under either weak spatial impedance, that is, low transportation and/or communication costs, or low congestion costs, or both. These conditions are reminiscent of those obtained in Proposition 2 for the equilibrium distribution to be non-uniform in the case of a single population.

4. Concluding remarks

We proposed a setting that captures local and global interactions across individuals. The main thrust of the paper was to study (i) when the interplay between agglomeration and dispersion forces leads to the emergence of central places and (ii) how the corresponding trade-off determines the sizes and locations of central places. Although we used simple explicit functional forms, we have seen that our model displays enough versatility to generate a wide range of equilibrium patterns, which we can characterize analytically. Figs. 2 and 3 show that, even in a simple setting, a marginal change in a key parameter may lead to much contrasted results. Therefore, the modeling strategy of cities and the choice of urban parameters require caution in quantitative models of urban systems.

Admittedly, the flip side of our approach is the use of reduced forms, such as (1) and (9), to describe agents' preferences. We acknowledge that a model with detailed micro-foundations is preferable to ours. Yet, the various attempts surveyed in the introduction often rely on ad hoc assumptions or treat cities as unrelated entities. Therefore, we find it fair to say that a full theory of urban systems is still lacking. This is why we believe that our approach may contribute to a better understanding of the reasons that lie behind the existence of striking regularities across urban systems by offering a framework against which to test more specific models.[10] Indeed, our setting displays enough versatility to replicate various urban structures. This can be achieved by identifying the conditions to be imposed on the model parameters for the spatial equilibrium to exhibit the desired features.

[10] One candidate that comes immediately to mind is the Zipf Law.

Appendix A. Proof of Proposition 1

We may rewrite the equilibrium condition (5) as follows:

$$\mathbf{n} = \mathbf{G}\mathbf{n} + u^*\mathbf{1}, \tag{A.1}$$

where $\mathbf{n}$ is the population distribution $n(x)$ viewed as an element of the Banach space $C(\mathbb{R})$ of bounded piecewise continuous real-valued functions, $\mathbf{1}$ is the function which is equal to 1 for all $x \in \mathbb{R}$, while $\mathbf{G} : C(\mathbb{R}) \to C(\mathbb{R})$ is the Fredholm-type integral operator associated with the negative-exponential convolution kernel:

$$(\mathbf{G}\mathbf{n})(x) \equiv \int_{-\infty}^{\infty} g(x,y)n(y)\mathrm{d}y, \qquad g(x,y) \equiv \frac{1}{\alpha}\exp\{-\beta|x-y|\}. \tag{A.2}$$

We first show the following intermediate result.

Lemma 1. *The dominant eigenvalue (i.e., the largest one in the absolute value) of* $\mathbf{G}$ *is given by* $\lambda_{\max}(\mathbf{G}) = \phi$.

Proof. It is readily verified that $\mathbf{1}$ is the eigenfunction of $\mathbf{G}$, while the corresponding eigenvalue is $1/\phi$. Furthermore, as the kernel $g(x, y)$ is symmetric, the operator $\mathbf{G}$ is self-adjoint. Hence, any two distinct eigenfunctions $\mathbf{n_1}$ and $\mathbf{n_2}$ must be pairwise orthogonal:

$$\int_{-\infty}^{\infty} n_1(x)n_2(x)\mathrm{d}x = 0.$$

This implies that $\mathbf{1}$ is the only eigenfunction of $\mathbf{G}$ (up to a scalar multiplier), which is strictly positive. Since the kernel $g(x, y)$ is strictly positive, the Frobenius-Perron theorem implies that both its dominant eigenvalue and the corresponding eigenfunction must be strictly positive. Since $\mathbf{1}$ is the only positive eigenfunction of $\mathbf{G}$, the corresponding eigenvalue must be the dominant one. Q.E.D.

We now proceed with the proof of Proposition 1.

By Lemma 1, $\phi < 1$ implies $\lambda_{\max}(\mathbf{G}) < 1$, which means that the operator in the right-hand side of (A.1) is a contraction mapping. The unique solution to (A.1) is then given by the Neumann series:

$$\mathbf{n} = u^*\left(\mathbf{I_d} - \mathbf{G}\right)^{-1}\mathbf{1} = u^* \sum_{k=0}^{\infty} \mathbf{G}^k\mathbf{1}, \tag{A.3}$$

where $\mathbf{I_d}$ is the identity operator.

Furthermore, for each $k = 0, 1, 2, \ldots$, we have:

$$\mathbf{G}^k\mathbf{1} = \phi^k\mathbf{1}.$$

Plugging this expression into (A.3) and summing the resulting geometric series, we obtain:

$$\mathbf{n} = \frac{u^*}{1-\phi}\,\mathbf{1}.$$

Consequently, the equilibrium distribution can only be uniform: $n(x) = \bar{n}$.

We now prove the stability of the uniform steady-state. Since the equilibrium utility level u^* is the same for all equilibria (see (iii) below), we may consider the following continuous-time dynamic adjustment process:

$$\frac{\mathrm{d}n_t(x)}{\mathrm{d}t} = \delta\left[u_t(x) - u^*\right], \tag{A.4}$$

where $\delta > 0$ is a constant, $n_t(x)$ is the population distribution at location $x \in \mathbb{R}$ and time $t \in \mathbb{R}_+$, while $u_t(x)$ is the utility level of an agent located at x when the population distribution is $\mathbf{n}_t$. Clearly, a distribution $\mathbf{n} \in C(\mathbb{R})$ is a spatial equilibrium if and only if it is a steady state of the differential Eq. (A.4): for all $x \in \mathbb{R}$,

$$n_t(x) = n(x) \Longrightarrow \mathrm{d}n_t(x)/\mathrm{d}t = 0.$$

Using the operator $\mathbf{G}$ defined by (A.2), we may rewrite (A.4) as follows:

$$\frac{\mathrm{d}\mathbf{n}_t}{\mathrm{d}t} = \delta(\mathbf{u}_t - u^*\mathbf{1}), \tag{A.5}$$

where $\mathbf{u}_t \in C(\mathbb{R})$ is defined by

$$\mathbf{u}_t \equiv \alpha(\mathbf{G} - \mathbf{I_d})\mathbf{n}_t.$$

By (8), we also have:

$$u^*\mathbf{1} = \alpha(\phi - 1)\bar{n}\mathbf{1} = \alpha\bar{n}(\mathbf{G} - \mathbf{I_d})\mathbf{1},$$

where the second equality holds because $\mathbf{1}$ is an eigenfunction of $\mathbf{G}$ corresponding to its principal eigenvalue, $\lambda_{\max}(\mathbf{G}) \equiv \phi$ (see Lemma 1).

Plugging the expressions for $\mathbf{u}_t$ and $u^*\mathbf{1}$ into (A.5), we obtain the following linear differential equation:

$$\frac{\mathrm{d}\tilde{\mathbf{n}}_t}{\mathrm{d}t} = \alpha\delta(\mathbf{G} - \mathbf{I_d})\tilde{\mathbf{n}}_t, \tag{A.6}$$

where $\tilde{\mathbf{n}}_t \equiv \mathbf{n}_t - \bar{n}\mathbf{1}$ is the deviation of the population distribution from the mean.

Studying the stability of the uniform distribution under (A.5) is equivalent to studying the stability of the zero steady state of (A.6). Since the linear operator $\mathbf{G} - \mathbf{I_d}$ in the right-hand side of (A.6) is a bounded operator, we may use the standard stability methods of linear differential equations to pin down the condition for all the eigenvalues of $\mathbf{G} - \mathbf{I_d}$ to be negative (Pontryagin, 1962). It follows from Lemma 1 that this condition is given by $\phi < 1$. Q.E.D. □

Appendix B. Proof of Proposition 2

Assume now that $\phi \geq 1$. In this case, the solution to (5) need not be unique.

The following lemma is proven in the Supplementary Material.

Lemma 2. *Let $n^*(x)$ be a spatial equilibrium population distribution and (x_1, x_2) an open interval such that $n^*(x) > 0$ over (x_1, x_2). Then, $n^*(x)$ is twice continuously differentiable over (x_1, x_2).*

We now proceed with the proof of Proposition 2.

Restate (5) as follows:

$$\tilde{n}(x) = \frac{1}{\alpha}\int_{-\infty}^{x} \exp\{-\beta(x-y)\}\tilde{n}(y)\mathrm{d}y + \frac{1}{\alpha}\int_{x}^{\infty} \exp\{\beta(x-y)\}\tilde{n}(y)\mathrm{d}y, \tag{B.1}$$

where $\tilde{n}(x)$ is the deviation of the population distribution from the mean: $\tilde{n}(x) \equiv n^*(x) - \bar{n}$. Differentiating twice both sides of (B.1) with respect to x yields:

$$\tilde{n}''(x) = -\beta^2(\phi - 1)\tilde{n}(x). \tag{B.2}$$

We seek a non-trivial solution to (B.2) which satisfies the following condition: $\tilde{n}(0) = \tilde{n}(l) = 0$ for some $l > 0$. This is the simplest case of the Sturm–Liouville problem. It is well known that a non-trivial solution to (B.2) exists only if α and β satisfy the following condition:

$$\beta^2(\phi - 1) \in \left\{\lambda_k \mid k = 1, 2, \ldots,\right\}, \qquad \lambda_k \equiv \left(\frac{\pi k}{l}\right)^2. \tag{B.3}$$

The numbers λ_k in (B.3) are the eigenvalues of the Sturm–Liouville operator $-\mathrm{d}^2/\mathrm{d}x^2$. The corresponding non-trivial solutions to (B.2), that is, the eigenfunctions of $-\mathrm{d}^2/\mathrm{d}x^2$, are given by

$$\tilde{n}_k(x) = A\sin\left(\sqrt{\lambda_k}\,x\right), \qquad k = 1, 2, \ldots, \tag{B.4}$$

where A is an arbitrary non-zero constant. Combining (B.3) with (B.4) yields:

$$\tilde{n}(x) = A\sin\left(\beta\sqrt{\phi - 1}\,x\right),$$

which is equal to (6) after adding $\bar{n}$ to both sides and taking (4) into account. Since the population size at x cannot be negative, it must be that $0 \leq A \leq \bar{n}$. Q.E.D.

Appendix C. Proof of Proposition 3

The proof involves two steps. In the first one, we show that (6) is unstable when $A < \bar{n}$. In the second step, we show that (7) is stable.

Step 1. Let us first show that, for any A such that $0 \leq A < \bar{n}$, the function $\tilde{\mathbf{n}}_A \in C(\mathbb{R})$ defined by $\tilde{n}_A(x) \equiv A\sin\left(\beta\sqrt{\phi - 1}\,x\right)$ is an unstable steady state of (A.6). Instability means that a small perturbation in the initial conditions of (A.6) does not decay over time. Formally, there exists a function $\mathbf{v} \in C(\mathbb{R})$, such that, if we set $\tilde{\mathbf{n}}_0 = \tilde{\mathbf{n}}_A + \varepsilon\mathbf{v}$ where $\varepsilon > 0$ is arbitrarily small, then the solution $\tilde{\mathbf{n}}_t$ to (A.6) starting from $\tilde{\mathbf{n}}_0$ at $t = 0$ does not converge back to the equilibrium $\tilde{\mathbf{n}}_A$ as $t \to \infty$.[11] To show this, we need the following Lemma.

Lemma 3. *Each $\lambda \in (0, \phi)$ is an eigenvalue of the integral operator $\mathbf{G}$, while the corresponding eigenfunctions are given by*

$$v_\lambda(x) \equiv A\sin\left(\beta\sqrt{\frac{\phi - \lambda}{\lambda}}\,x\right) + B\cos\left(\beta\sqrt{\frac{\phi - \lambda}{\lambda}}\,x\right), \tag{C.1}$$

where A and B are arbitrary constants.

Proof. By definition, $\mathbf{v}$ is an eigenfunction of $\mathbf{G}$ if and only if it is a solution to the equation $\mathbf{G}\mathbf{v} = \lambda\mathbf{v}$ or, equivalently,

$$\lambda v(x) = \frac{1}{\alpha}\int_{-\infty}^{\infty} \exp\{-\beta|x - y|\}v(y)\mathrm{d}y. \tag{C.2}$$

Eq. (C.2) is known as the Lalesco–Picard equation. Our result then follows directly from the analysis of that equation provided in Polyanin and Manzhirov (2008, Ch.4). Q.E.D.

As implied by Lemma 1, the operator $\mathbf{G} - \mathbf{I_d}$ is bounded. Hence, the solution to (A.6) can be rewritten as follows:

$$\tilde{\mathbf{n}}_t = \exp\left\{\alpha\delta t\left(\mathbf{G} - \mathbf{I_d}\right)\right\}\tilde{\mathbf{n}}_0, \tag{C.3}$$

where $\exp\left\{\alpha\delta t\left(\mathbf{G} - \mathbf{I_d}\right)\right\}$ is a bounded operator defined by the following series:

$$\exp\left\{\alpha\delta t\left(\mathbf{G} - \mathbf{I_d}\right)\right\} \equiv \sum_{k=0}^{\infty}\frac{(\alpha\delta t)^k}{k!}\left(\mathbf{G} - \mathbf{I_d}\right)^k. \tag{C.4}$$

Since $\phi > 1$, we may choose $\lambda \in (1, \phi)$. Furthermore, we set $\tilde{\mathbf{n}}_0 = \tilde{\mathbf{n}}_A + \varepsilon\mathbf{v}_\lambda$, where $0 < \varepsilon < \bar{n} - A$, which is possible because $A < \bar{n}$. Therefore, $\tilde{\mathbf{n}}_0$ is always strictly positive. As implied by Lemma 3, any $\mathbf{v}_\lambda$ given by (C.1) is an eigenfunction of $\mathbf{G} - \mathbf{I_d}$, the corresponding eigenvalue being

[11] Since we work in the space $C(\mathbb{R})$ of bounded continuous functions, the relevant type of convergence is the *uniform consergence*, i.e., convergence in the sup-norm in $C(\mathbb{R})$.

$\lambda - 1$. Combining this with $(\mathbf{G} - \mathbf{I_d})\tilde{\mathbf{n}}_A = \mathbf{0}$ and using (C.3) – (C.4), we get:

$$\tilde{\mathbf{n}}_t - \tilde{\mathbf{n}}_A = \sum_{k=0}^{\infty} \frac{(\alpha\delta t)^k}{k!} \left(\mathbf{G} - \mathbf{I_d}\right)^k (\tilde{\mathbf{n}}_0 - \tilde{\mathbf{n}}_A) = \varepsilon \sum_{k=0}^{\infty} \frac{[(\lambda-1)\alpha\delta t]^k}{k!} \mathbf{v}_\lambda = \varepsilon \exp\{(\lambda - 1)\alpha\delta t\}\mathbf{v}_\lambda.$$

Hence, $\|\tilde{\mathbf{n}}_t - \tilde{\mathbf{n}}_A\| = \varepsilon \exp\{(\lambda-1)\alpha\delta t\}$, where $\|\cdot\|$ stands for the uniform norm in $C(\mathbb{R})$. Since $\varepsilon > 0$ and $\lambda > 1$, the distance between the perturbed solution $\tilde{\mathbf{n}}_t$ and the steady state $\tilde{\mathbf{n}}_A$ increases with time. As a result, whenever $0 \le A < \bar{n}$, the equilibrium distribution given by $n_A(x) \equiv \bar{n} + \tilde{n}_A(x)$ is unstable.

Step 2. Proving the stability of the equilibrium (7), denoted $\mathbf{n}^*$, for the whole space of admissible perturbations is a formidable task. However, we can show that $\mathbf{n}^*$ is stable with respect to any perturbation $\mathbf{v}$ which can be represented by a uniformly convergent Fourier series. The set of such functions is known as the space $\Pi(\mathbb{R})$ of *almost-periodic functions* (Bohr, 2018). We find this level of generality sufficient for our purpose because $\Pi(\mathbb{R})$ is an infinite-dimensional closed subspace of $C(\mathbb{R})$ which contains all periodic functions, as well as all the linear combinations of those.[12]

The proof goes as follows. For $0 \le A < \bar{n}$, we have $n_A(x) > 0$ for all $x \in \mathbb{R}$. Hence, for any $v \in C(\mathbb{R})$, the perturbation εv is admissible for sufficiently small values of ε since we have $n_A(x) + \varepsilon v(x) \ge 0$ for all $x \in \mathbb{R}$. This is no longer true for $\mathbf{n}^*$. Indeed, let us define:

$$x_j \equiv \frac{(4j-1)\pi}{2\beta\sqrt{\phi - 1}}, \qquad j = 0, \pm 1, \pm 2, \dots, \tag{C.5}$$

Clearly, for each $j = 0, \pm 1, \pm 2, \dots$, we have $n^*(x_j) = 0$. Hence, to guarantee that $\varepsilon v \in C(\mathbb{R})$ is an admissible perturbation of the steady state $\tilde{\mathbf{n}}^* \equiv \mathbf{n}^* - \bar{n}\mathbf{1}$ of (A.6), we need to impose the following conditions:

$$v(x_j) \ge 0, \qquad \text{for all } j = 0, \pm 1, \pm 2, \dots,. \tag{C.6}$$

Hence, to obtain the desired result, we show that, for any $\mathbf{v} \in \Pi(\mathbb{R})$, divergence of the solution $\tilde{\mathbf{n}}_t$ to (A.6) such that $\tilde{\mathbf{n}}_0 = \tilde{\mathbf{n}}^* + \varepsilon\mathbf{v}$ from $\tilde{\mathbf{n}}^*$ implies that the conditions (C.6) are violated.

Assume, first, that $\mathbf{v}$ is a periodic function with a period equal to $1/\omega$, where $\omega > 0$ is the fundamental frequency of $\mathbf{v}$. Then, by the Weierstrass approximation theorem (see, e.g., Rudin, 1987, Theorem 4.25), $\mathbf{v}$ can be represented by a uniformly convergent trigonometric Fourier series of the following form:

$$\mathbf{v} = \sum_{k=1}^{\infty} \mathbf{v}_k,$$

where $\mathbf{v}_k$, for each $k = 1, 2, \dots$, is defined as follows:

$$v_k(x) \equiv A_k \sin(2\pi\omega k x) + B_k \cos(2\pi\omega k x).$$

As implied by Lemma 3, for each $k = 1, 2, \dots$, $\mathbf{v}_k$ is an eigenfunction of $\mathbf{G}$ corresponding to the eigenvalue λ_k given by:

$$\lambda_k \equiv \frac{\phi}{1 + (2\pi\omega k/\beta)^2}.$$

Applying the operator $\exp\left\{\alpha\delta t\left(\mathbf{G} - \mathbf{I_d}\right)\right\}$ to both parts of (C.5) yields:

$$\exp\left\{\alpha\delta t\left(\mathbf{G} - \mathbf{I_d}\right)\right\}\mathbf{v} = \sum_{k=1}^{\infty} \exp\left\{(\lambda_k - 1)\alpha\delta t\right\}\mathbf{v}_k.$$

Hence, the deviation $\tilde{\mathbf{n}}_t - \tilde{\mathbf{n}}^*$ of the solution to (A.6) starting from $\tilde{\mathbf{n}}_0 = \tilde{\mathbf{n}}^* + \varepsilon\mathbf{v}$ from the equilibrium $\tilde{\mathbf{n}}^*$ can be represented as follows:

$$\tilde{\mathbf{n}}_t - \tilde{\mathbf{n}}^* = \varepsilon\left[\sum_{\lambda_k > 1} \exp\left\{(\lambda_k - 1)\alpha\delta t\right\}\mathbf{v}_k + \sum_{\lambda_k < 1} \exp\left\{(\lambda_k - 1)\alpha\delta t\right\}\mathbf{v}_k\right]. \tag{C.7}$$

[12] By comparison, we prove the instability of $\mathbf{n}_A^*$ for $0 \le A < \bar{n}$ by constructing a periodic perturbation which leads to divergence from $\mathbf{n}_A^*$ over time.

The second term in the right-hand side of (C.7) converges to zero when $t \to \infty$. Therefore, we may ignore it and focus on the first term. In other words, we may replace (C.5) with a trigonometric polynomial given by:

$$v(x) = \sum_{k=1}^{N(\omega)} \left[A_k \sin(2\pi\omega k x) + B_k \cos(2\pi\omega k x)\right], \tag{C.8}$$

where $N(\omega)$ is a finite integer given by $N(\omega) \equiv \max\{k\!: \lambda_k > 1\}$. It can be shown that for any function of the form (C.8) which is not identically zero there exist two integers, m and r such that $v(x_m) > 0 > v(x_r)$, where x_m and x_r are obtained, respectively, by setting $j = m$ and $j = r$ in (C.5).[13] In other words, when $\mathbf{v}$ is periodic and such that the solution $\tilde{\mathbf{n}}_t$ to (A.6) starting from $\tilde{\mathbf{n}}_0 = \tilde{\mathbf{n}}^* + \varepsilon\mathbf{v}$ does not converge back to $\tilde{\mathbf{n}}^*$ as $t \to \infty$, the conditions (C.6) are violated. Hence, no periodic perturbation leads to divergence from $\tilde{\mathbf{n}}^*$.

Second, by definition, an almost-periodic function $\mathbf{v} \in \Pi(\mathbb{R})$ can be uniformly approximated by periodic functions with an arbitrary degree of precision. Combining this with the fact that $\exp\{\alpha\delta t(\mathbf{G} - \mathbf{I_d})\}$ is a continuous operator, we conclude that an almost-periodic perturbation cannot lead to the divergence of $\tilde{\mathbf{n}}_t$ from $\tilde{\mathbf{n}}^*$. Q.E.D. □

Appendix D. Proof of Proposition 4

The equilibrium condition is as follows:

$$E(x) - \alpha n(x) = u^*, \qquad \text{for all } x \in \mathbb{R},$$

where u^* is an unknown constant. To determine u^*, we first integrate both sides of this condition over $[-b, b]$ and multiply both sides by $1/(2b)$, where $b > 0$:

$$\frac{1}{2b}\int_{-b}^{b} E(x)\mathrm{d}x - \frac{\alpha}{2b}\int_{-b}^{b} n(x)\mathrm{d}x = u^*.$$

Since the mean is exogenous and given by

$$\bar{n} \equiv \lim_{b\to\infty} \frac{1}{2b}\int_{-b}^{b} n(x)\mathrm{d}x,$$

it is sufficient to show that

$$\lim_{b\to\infty} \frac{1}{2b}\int_{-b}^{b} E(x)\mathrm{d}x = \frac{2}{\beta}\bar{n}. \tag{D.1}$$

to obtain the desired result.

Let $M \equiv \sup_{x\in\mathbb{R}} n(x) < \infty$. Then, we have:

$$E(x) \equiv \int_{-\infty}^{\infty} \exp\{-\beta|x-y|\}n(y)\mathrm{d}y \le M\int_{-\infty}^{\infty} \exp\{-\beta|x-y|\}\mathrm{d}y = \frac{2}{\beta}M.$$

Hence, the limit in the left-hand side of (D.1) exists and is finite. This limit equals $2\bar{n}/\beta$ if the following equality

$$\lim_{b\to\infty} \frac{1}{2b}\int_{-b}^{b}\left[\int_{-\infty}^{\infty} \exp\{-\beta|x-y|\}n(y)\mathrm{d}y\right]\mathrm{d}x = \lim_{b\to\infty} \frac{1}{2b}\int_{-b}^{b}\left[\int_{-\infty}^{\infty} \exp\{-\beta|x-y|\}\mathrm{d}x\right]n(y)\mathrm{d}y \tag{D.2}$$

holds. The limit of the right-hand side of (D.2) is equal to $2\bar{n}/\beta$. As seen above, the limit of the right-hand side of (D.2) also exists. It remains to show that the two limits are equal. To do so, we set:

$$\mathcal{I}(b) \equiv \frac{1}{2b}\int_{-b}^{b}\int_{-b}^{b} \exp\{-\beta|x-y|\}n(y)\mathrm{d}x\mathrm{d}y,$$

[13] Indeed, let us choose $x^+ \in \mathbb{R}$ and $x^- \in \mathbb{R}$ such that $v(x^+) > 0 > v(x^-)$. Then, if $2\pi\omega/\sqrt{\beta(\phi-1)}$ is an irrational number, for any $\varepsilon > 0$ there exist integers M, R, m and r, such that $|x_m - x^+ + M/\omega| < \varepsilon$ and $|x_r - x^- + R/\omega| < \varepsilon$. By continuity of v, it must be true that $v(x_m) > 0 > v(x_r)$. The zero-measure case where $2\pi\omega/\sqrt{\beta(\phi-1)}$ is a rational number is ruled out by using a standard continuity argument.

$$\mathcal{R}_1(b) \equiv \frac{1}{2b}\int_{-b}^{b}\left[\int_{|y|>b}\exp\{-\beta|x-y|\}n(y)\mathrm{d}y\right]\mathrm{d}x,$$

$$\mathcal{R}_2(b) \equiv \frac{1}{2b}\int_{-b}^{b}\left[\int_{|x|>b}\exp\{-\beta|x-y|\}\mathrm{d}x\right]n(y)\mathrm{d}y.$$

It is readily verified that the following identities hold:

$$\frac{1}{2b}\int_{-b}^{b}\left[\int_{-\infty}^{\infty}\exp\{-\beta|x-y|\}n(y)\mathrm{d}y\right]\mathrm{d}x = \mathcal{I}(b)+\mathcal{R}_1(b), \tag{D.3}$$

$$\frac{1}{2b}\int_{-b}^{b}\left[\int_{-\infty}^{\infty}\exp\{-\beta|x-y|\}\mathrm{d}x\right]n(y)\mathrm{d}y = \mathcal{I}(b)+\mathcal{R}_2(b). \tag{D.4}$$

Since $n(x)$ is bounded above by M, we obtain:

$$0<\mathcal{R}_k(b)<\frac{2M}{\beta^2}\cdot\frac{1}{b}, \qquad k=1,2,$$

which implies:

$$\lim_{b\to\infty}\mathcal{R}_1(b)=\lim_{b\to\infty}\mathcal{R}_2(b)=0.$$

Taking the limit on both sides of (D.3) and (D.4) when $b\to\infty$ shows that (i) $\mathcal{I}(b)$ has a finite limit when $b\to\infty$, and (ii) the following equalities hold:

$$\lim_{b\to\infty}\frac{1}{2b}\int_{-b}^{b}\left[\int_{-\infty}^{\infty}\exp\{-\beta|x-y|\}n(y)\mathrm{d}y\right]\mathrm{d}x = \lim_{b\to\infty}\mathcal{I}(b)$$
$$= \lim_{b\to\infty}\frac{1}{2b}\int_{-b}^{b}\left[\int_{-\infty}^{\infty}\exp\{-\beta|x-y|\}\mathrm{d}x\right]n(y)\mathrm{d}y,$$

which proves (D.2). Q.E.D.

Appendix E. Proof of Proposition 5

The expression (10) may be rewritten as follows:

$$E_{jk}(x)=\int_{-\infty}^{x}\exp\{-\beta_{jk}(x-y)\}n_k(y)\mathrm{d}y+\int_{x}^{\infty}\exp\{\beta_{jk}(x-y)\}n_k(y)\mathrm{d}y.$$

Differentiating twice $E_{jk}(x)$ with respect to x yields:

$$E''_{jk}(x)=\beta_{jk}^2E_{jk}(x)-2\beta_{jk}n_k(x), \qquad j,k=1,2.$$

or, in vector-matrix form,

$$\mathbf{E}''(x)=\mathbf{QE}(x)-2\mathbf{Bn}(x),$$

where

$$\mathbf{n}(x)\equiv\left(n_1(x),n_2(x)\right)^{\mathrm{T}}, \qquad \mathbf{E}(x)\equiv\left(E_{11}(x),E_{12}(x),E_{21}(x),E_{22}(x)\right)^{\mathrm{T}}.$$

The equilibrium conditions in vector-matrix form are given by

$$\mathbf{u}^*=\mathbf{\Gamma E}(x)-\mathbf{An}(x), \tag{E.1}$$

where $\boldsymbol{u}^*\equiv\left(u_1^*,u_2^*\right)^{\mathrm{T}}$. Computing the mean of (E.1) yields $\boldsymbol{u}^*=\mathbf{\Gamma}\overline{\mathbf{E}}-\mathbf{A}\overline{\mathbf{n}}$. Subtracting this expression from (E.1) and multiplying both sides by $\mathbf{A}^{-1}$, we obtain:[14]

$$\widetilde{\mathbf{n}}(x)=\mathbf{A}^{-1}\mathbf{\Gamma}\widetilde{\mathbf{E}}(x), \tag{E.2}$$

where $\widetilde{\mathbf{E}}(x)\equiv\mathbf{E}(x)-\overline{\mathbf{E}}$.

Plugging (E.2) into (E.1) and subtracting $\overline{\mathbf{E}}$ from both sides, we come to the following system of linear second-order differential equations:

$$\widetilde{\mathbf{E}}''(x)=\mathbf{D}\widetilde{\mathbf{E}}(x). \tag{E.3}$$

With almost no loss of generality, we may focus on the case where $\mathbf{D}$ has four linearly independent eigenvectors.[15] Let λ_j be the jth eigenvalue of $\mathbf{D}$ and let $\mathbf{S}_j=(s_{1j},s_{2j},s_{3j},s_{4j})$ be the corresponding eigenvector, for $j=1,\ldots,4$. Diagonalizing $\mathbf{D}$, we get:

$$\mathbf{D}=\mathbf{S\Lambda S}^{-1},$$

where

$$\mathbf{\Lambda}\equiv\begin{pmatrix}\lambda_1&0&0&0\\0&\lambda_2&0&0\\0&0&\lambda_3&0\\0&0&0&\lambda_4\end{pmatrix}, \qquad \mathbf{S}\equiv\begin{pmatrix}s_{11}&s_{12}&s_{13}&s_{14}\\s_{21}&s_{22}&s_{23}&s_{24}\\s_{31}&s_{32}&s_{33}&s_{34}\\s_{41}&s_{42}&s_{43}&s_{44}\end{pmatrix}.$$

Consider now the following change of variables

$$\mathbf{Z}(x)\equiv\mathbf{S}^{-1}\widetilde{\mathbf{E}}(x). \tag{E.4}$$

and rewrite the system (E.3) in terms of $\mathbf{Z}(x)$:

$$\mathbf{Z}''(x)=\mathbf{\Lambda Z}(x),$$

or, equivalently,

$$z''_j(x)=\lambda_j z_j(x), \qquad j=1,\ldots,4, \tag{E.5}$$

where $z_j(z)$ is the jth component of $\mathbf{Z}(x)$.

Since the system (E.5) is formed by four independent equations, we can solve each equation separately. Using (E.2) and (E.4), the equilibrium distribution $\mathbf{n}^*(x)$ is related to the solutions $\mathbf{Z}^*(x)$ through the following relationship:

$$\mathbf{n}^*(x)=\mathbf{A}^{-1}\mathbf{\Gamma SZ}^*(x)+\overline{\mathbf{n}}. \tag{E.6}$$

Since $\mathbf{n}^*(x)$ is bounded, the same must hold for $\mathbf{Z}^*(x)$.

The solution to (E.3) is given by

$$z_j^*(x)=A_j\exp\left\{\sqrt{\lambda_j}\,x\right\}+B_j\exp\left\{-\sqrt{\lambda_j}\,x\right\}, \qquad j=1,\ldots,4,$$

where A_j and B_j are arbitrary constants.

Three cases may arise.

(i) If $\mathbf{D}$ has no strictly negative real eigenvalues, $\mathbf{Z}^*(x)$ is bounded over $\mathbb{R}$ if and only if $A_j=B_j=0$ for all j because $x\in\mathbb{R}$. In other words, $\mathbf{Z}^*(x)\equiv\mathbf{0}$ is the only bounded solution to (E.5). It then follows from (E.6) that $\mathbf{n}^*(x)\equiv\overline{\mathbf{n}}$, i.e., the spatial equilibrium is unique and uniform. This proves part (i) of Proposition 5.

(ii) If $\mathbf{D}$ has only one strictly negative real eigenvalue (λ, say), then, the Eq. (E.5) with $j=1$ is such that the non-trivial solutions are the eigenfunctions of the Sturm-Liouville operator $-\mathrm{d}^2/\mathrm{d}x^2$. Following the same logic as in the proof of Proposition 2, we obtain:

$$z_1^*(x)=C_1\sin\left(\sqrt{-\lambda}\,x+\theta_1\right), \tag{E.7}$$

where C_1 is an arbitrary constant, while $\theta_1\in[0,2\pi)$. Without loss of generality we may normalize θ_1 to zero because we do not specify the origin. Clearly, $z_1^*(x)$ is bounded above over $\mathbb{R}$. Combining (E.6) with (E.7) yields:

$$\mathbf{n}^*(x)=\overline{\mathbf{n}}+C_1\sin\left(\sqrt{-\lambda}\,x\right)\mathbf{A}^{-1}\mathbf{\Gamma S}_1,$$

which proves part (ii).

(iii) If $\mathbf{D}$ has at least two distinct real and strictly negative eigenvalues (λ_1 and λ_2, say), then using the same reasoning as in the proof of Proposition 2 yields:

$$z_j^*(x)=C_j\sin\left(\sqrt{-\lambda_j}\,x+\theta_j\right), \qquad j=1,2.$$

Let us normalize θ_1 to zero and set $\theta_2\equiv\theta$. As a result, any distribution of the form

$$\mathbf{n}^*(x)=\overline{\mathbf{n}}+C_1\sin\left(\sqrt{-\lambda_1}\,x\right)\mathbf{A}^{-1}\mathbf{\Gamma S}_1+C_2\sin\left(\sqrt{-\lambda_2}\,x+\theta\right)\mathbf{A}^{-1}\mathbf{\Gamma S}_2$$

is an equilibrium. When $\sqrt{\lambda_1/\lambda_2}$ is not a rational number, $\mathbf{n}^*(x)$ is non-periodic, which implies that the corresponding spatial equilibria involve different extrema. This proves part (iii). Q.E.D.

[14] Since we assume that $\det(\mathbf{A})\neq 0$, the inverse matrix $\mathbf{A}^{-1}$ is well defined.

[15] The case where this condition does not hold has a zero measure.

Appendix F. Proof of Proposition 6

Define the following matrix:

$$\mathbf{M} \equiv \begin{pmatrix} \frac{1}{\alpha} & 0 \\ 0 & \frac{1}{\alpha} \\ \frac{1}{\alpha} & 0 \\ 0 & \frac{1}{\alpha} \end{pmatrix} \begin{pmatrix} 1 & \gamma & 0 & 0 \\ 0 & 0 & \gamma & 1 \end{pmatrix} = \begin{pmatrix} \frac{1}{\alpha} & \frac{\gamma}{\alpha} & 0 & 0 \\ 0 & 0 & \frac{\gamma}{\alpha} & \frac{1}{\alpha} \\ \frac{1}{\alpha} & \frac{\gamma}{\alpha} & 0 & 0 \\ 0 & 0 & \frac{\gamma}{\alpha} & \frac{1}{\alpha} \end{pmatrix}.$$

The **D**-matrix can then be rewritten as follows:

$$\mathbf{D} = \beta^2 \mathbf{I} - 2\beta \mathbf{M}.$$

Let λ_i be the ith eigenvalue of **D**, while μ_i is the ith eigenvalue of **M**, with $i = 1, \ldots, 4$. Then, we have:

$$\lambda_i = \beta^2 - 2\beta\mu_i. \tag{F.1}$$

Since $\mathrm{rank}(\mathbf{M}) = 2$, two eigenvalues are equal to zero, i.e., $\mu_3 = \mu_4 = 0$ (Horn and Johnson, 1985). Combining this with (F.1) implies that $\lambda_3 = \lambda_4 = \beta^2 > 0$. To find λ_1 and λ_2, we need to determine μ_1 and μ_2. To do so, we use the following result: if a matrix is represented as a product of two matrices, the set of non-zero eigenvalues is not sensitive to changing the order of multiplication (Horn and Johnson, 1985, p. 53, Theorem 1.3.20). Changing the order of multiplication in **M**, we find that μ_1 and μ_2 are the eigenvalues of the following (2×2)-matrix:

$$\frac{1}{\alpha}\begin{pmatrix} 1 & \gamma \\ \gamma & 1 \end{pmatrix}.$$

The characteristic equation of this matrix is given by

$$\mu^2 - \frac{2}{\alpha}\mu + \frac{1-\gamma^2}{\alpha^2} = 0.$$

Since $\gamma < 1$, both solutions to this equation are positive and are given by:

$$\mu_{1,2} = \frac{1 \pm \gamma}{\alpha}. \tag{F.2}$$

Using (F.1) and (F.2), we find that λ_1 and λ_2 are given by:

$$\lambda_{1,2} = \beta^2[1 - \phi(1 \pm \gamma)],$$

where $\phi \equiv 2/(\alpha\beta)$. Three cases may arise.

(i) If $\phi \le 1/(1+\gamma)$, using (F.3) shows that λ_1 and λ_2 are both non-negative, which means that **D** has no strictly negative real eigenvalues. By Proposition 5, the spatial equilibrium is unique and uniform, which proves part (i) of Proposition 6.

(ii) If $1/(1+\gamma) < \phi \le 1/(1-\gamma)$, we have: $\lambda_2 < 0 \le \lambda_1$. In this case, Proposition 5 implies that all spatial equilibria are periodic, which proves part (ii) of Proposition 6.

(iii) If $\phi > 1/(1-\gamma)$, we have: $\lambda_2 < \lambda_1 < 0$. Then, Proposition 5 implies part (iii) of Proposition 6. Q.E.D.

Supplementary material

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.jue.2019.01.006.

References

Baldwin, R., Forslid, R., Martin, P., Ottaviano, G.I.P., Robert-Nicoud, F., 2003. Economic Geography and Public Policy. Princeton University Press, Princeton, NJ.
Beckmann, M.J., 1976. Spatial equilibrium in the dispersed city. In: Papageorgiou, Y. (Ed.), Mathematical Land Use Theory, Lexington Books, Lexington, MA, pp. 117–125.
Behrens, K., Robert-Nicoud, F., 2015. Agglomeration theory with heterogeneous agents. In: Duranton, G., Henderson, J.V., Strange, W.C. (Eds.), Handbook of Regional and Urban Economics, Vol. 5. Elsevier, Amsterdam, pp. 171–245.
Berliant, M., Peng, S., Wang, P., 2002. Production externalities and urban configuration. J. Econ. Theory 104, 275–303.
Bohr, H., 2018. Almost Periodic Functions. Courier Dover Publications, New York.
Christaller, W., 1933. Die Zentralen Orte in Süddeutschland, Gustav Fischer Verlag, Jena. Prentice-Hall, Englewood Cliffs, NJ. English translation: The Central Places of Southern Germany 1966
Davis, D.R., Dingel, J., 2019. A spatial knowledge economy. Am. Econ. Rev. 109, 153–170.
Desmet, K., Rossi-Hansberg, E., 2014. Spatial development. Am. Econ. Rev. 104, 1211–1243.
Diamond, R., 2016. The determinants and welfare implications of US workers diverging location choices by skill: 1980–2000. Am. Econ. Rev. 106, 479–524.
Dunford, N., Schwartz, J.T., 1988. Linear Operators. Part II. Wiley, Hoboken, NJ.
Duranton, G., Puga, D., 2004. Micro-foundations of Urban Increasing Returns: Theory. In: Henderson, J.V., Thisse, J.F. (Eds.), Handbook of Regional and Urban Economics, Vol. 4. Elsevier, Amsterdam, pp. 2063–2117.
Eaton, B.C., Lipsey, R.G., 1982. An economic theory of central places. Econ. J. 92, 56–72.
Fujita, M., Mori, T., 1997. Structural stability and evolution of urban systems. Reg. Sci. Urban Econ. 27, 399–442.
Fujita, M., Ogawa, H., 1982. Multiple equilibria and structural transition of non-monocentric urban configurations. Reg. Sci. Urban Econ. 12, 161–196.
Fujita, M., Krugman, P., Venables, A.J., 1999a. The spatial economy. Cities, Regions and International Trade. MIT Press, Cambridge, MA.
Fujita, M., Krugman, P., Mori, T., 1999b. On the evolution of hierarchical urban systems. Eur. Econ. Rev. 43, 209–251.
Fujita, M., Thisse, J.F., 2013. Economics of agglomeration. Cities, Industrial Location and Globalization. Cambridge University Press, Cambridge.
Gabaix, X., Ioannides, Y.M., 2004. The evolution of city size distributions. In: Henderson, J.V., Thisse, J.F. (Eds.), Handbook of Regional and Urban Economics, Vol. 4. Elsevier, Amsterdam, pp. 2341–2378.
Gaubert, C., 2018. Firm sorting and agglomeration. Am. Econ. Rev. 108, 3117–3153.
Head, K., Mayer, T., 2004. The empirics of agglomeration and trade. In: Henderson, J.V., Thisse, J.F. (Eds.), Handbook of Regional and Urban Economics, Vol. 4. Elsevier, Amsterdam, pp. 2609–2669.
Helpman, E., 1998. The size of regions. In: Pines, D., Sadka, E., Zilcha, I. (Eds.), Topics in Public Economics. Theoretical and Applied Analysis. Cambridge University Press, Cambridge, pp. 33–54.
Helsley, R.W., 1990. Knowledge and production in the CBD. J. Urban. Econ. 28, 391–403.
Henderson, J.V., 1974. The sizes and types of cities. Am. Econ. Rev. 64, 640–656.
Henderson, J.V., 1988. Urban Development. Theory, Fact and Illusion. Oxford University Press, Oxford.
Horn, R.A., Johnson, C.R., 1985. Matrix Analysis. Cambridge University Press, Cambridge.
Hsu, W.T., 2012. Central place theory and city size distribution. Econ. J. 122, 903–932.
Jackson, M.O., Rogers, B.W., Zenou, Y., 2017. The economic consequences of social-network structure. J. Econ. Lit. 55, 49–95.
Jacobs, J., 1969. The Economy of Cities. Random House, New York.
Kolmogorov, A.N., Fomin, S.V., 2012. Introductory Real Analysis. Dover Publications, New York.
Krugman, P., 1991. Increasing returns and economic geography. J. Pol. Econ. 99, 483–499.
Krugman, P., Venables, A.J., 1995. Globalization and the inequality of nations. Q. J. Econ. 110, 857–880.
Lösch, A., 1940. Die räumliche ordnung der wirtschaft, gustav fischer, jena. English translation: The Economics of Location. Yale University Press, New Haven, CN. (1954).
Lucas Jr., R.E., Rossi-Hansberg, E., 2002. On the internal structure of cities. Econometrica 70, 1445–1476.
Matsuyama, K., 2008. Symmetry-breaking. In: Blume, L.E., Durlauf, S. (Eds.), The New Palgrave Dictionary of Economics. Palgrave Macmillan, London, pp. 6484–6487.
Mossay, P., Picard, P.M., 2011. On spatial equilibria in a social interaction model. J.Econ. Theory 146, 2455–2477.
Mulligan, G.F., Partridge, M.D., Carruthers, J.I., 2012. Central place theory and its reemergence in regional science. Ann. Reg. Sci. 48, 405–431.
Ogawa, H., Fujita, M., 1980. Equilibrium land use patterns in a nonmonocentric city. J. Reg. Sci. 20, 455–475.
Ottaviano, G.I.P., Tabuchi, T., Thisse, .J.F., 2002. Agglomeration and trade revisited. Int. Econ. Rev. (Philadelphia) 43, 409–436.
Papageorgiou, Y.Y., Smith, T.R., 1983. Agglomeration as local instability of spatially uniform steady–states. Econometrica 51, 1109–1119.
Polyanin, A.D., Manzhirov, A.V., 2008. Handbook of Integral Equations, second edition Chapman & Hall/CRC Press, London.
Pontryagin, L.S., 1962. Ordinary Differential Equations. Elsevier, Amsterdam.
Puga, D., 1999. The rise and fall of regional inequalities. Eur. Econ. Rev. 43, 303–334.
Rossi-Hansberg, E., 2005. A spatial theory of trade. Am. Econ. Rev. 95, 1464–1491.
Rudin, W., 1987. Real and Complex Analysis. McGraw-Hill, New York.
Smith, T., 1978. A cost-efficieny principle of spatial interaction behavior. Reg. Sci. Urban Econ. 8, 313–337.
Tabuchi, T., 1998. Urban agglomeration and dispersion: a synthesis of alonso and krugman. J. Urban Econ. 44, 333–351.
Tabuchi, T., Thisse, J.F., 2011. A new economic geography model of central places. J. Urban Econ. 69, 240–252.