Papers in Regional Science 103 (2024) 100022

Contents lists available at ScienceDirect

Papers in Regional Science

journal homepage: www.sciencedirect.com/journal/papers-in-regional-science

The rise (and fall) of tech clusters☆

Sergei Kichko[a,*], Wen-Jung Liang[b], Chao-Cheng Mai[c], Jacques-Francois Thisse[d], Ping Wang[e]

[a] Department of Economics and Management, University of Trento, Italy
[b] National Dong Hwa University, Taiwan, Republic of China
[c] Academia Sinica and Tamkang University, Taiwan, Republic of China
[d] CORE-Universite catholique de Louvain, Belgium and CEPR, UK
[e] Washington University in St. Louis and NBER, USA

ARTICLE INFO

JEL classification:
D51
L22
O33
R13

Keywords:
High-tech city
Knowledge spillovers
Intermediate firm clustering
Land use
Commuting
R&D

ABSTRACT

Tech clusters play a growing role in knowledge-based economies by accommodating high-tech firms and providing an environment that fosters location-dependent knowledge spillovers and promote R&D investments by firms. Yet, not much is known about the economic conditions under which such entities may form in equilibrium *without* government interventions. This paper develops a spatial equilibrium model with a competitive final sector and a monopolistically competitive intermediate sector, which allows us to determine necessary and sufficient conditions for a tech cluster to emerge as an equilibrium outcome. We show that strongly localized knowledge spillovers, skilled labor abundance, and low commuting costs are key drivers for a tech cluster to form. With continual improvements in infrastructure and communication technology that lowers distance decay in knowledge spillover or coordination costs, tech clusters will eventually be fragmented.

1. Introduction

The concept of tech clusters has gained the favor of many analysts and policy-makers (see Kerr and Robert-Nicoud, 2020, for a detailed review and an exhaustive list of references). Even though the idea of industrial district has been around for a long time (Marshall, 1890, ch. X), it was not until the 1990 s that the related concept of tech cluster, or science park, has been developed (Castells and Hall, 1994; Saxenian, 1994; Porter, 2000). Although there is a rich variety of tech clusters (Klepper, 2010; McCarthy et al., 2018), they do share some common features, which help us distinguishing them from other types of clusters. We then define a *tech cluster* as a settlement that (i) accommodates *knowledge-intensive firms* and (ii) encourages firms to undertake *R&D investments* because they benefit from *knowledge spillovers across the cluster* (Kerr and Robert-Nicoud, 2020). Hence, a firm's productivity depends on how much it invests in R&D, but also on the total amount of knowledge available in the cluster, whose value is determined by all firms' R&D expenditure weighted by the distances that separate them.

What characterizes skilled workers here is their ability to undertake different tasks in the intermediate and final sectors. In particular, when new and specialized intermediate inputs contain a large amount of knowledge, combining them to produce the final good requires specific coordination and learning activities (Becker and Murphy, 1992). This amounts to assuming that the final sector needs a number of skilled labor units when it uses newly designed intermediate inputs. What is more, a city hosting a tech cluster provides office and lab space, as well as housing, retail, restaurants and other leisure facilities within a compact geographic area (Katz and Bradley, 2013; McCarthy et al., 2018). To be precise, we consider a high-tech cluster as a special type of city whose spatial structure is determined endogenously by the interaction between multiple stakeholders through the above-mentioned channels.

Despite the importance of tech clusters in the real world, it is fair to say that this concept has attracted little attention in economic theory.

☆ We thank two anonymous reviewers, the editor R. Nicolini, M. Berliant, M. Fujita, S. Peng, E. Rossi-Hansberg, and participants at the Midwest Economic Theory Meeting, North American Regional Science Association Meeting, and the Society for Advanced Economic Theory Conference for useful comments. Financial support from Academia Sinica, the National Science Council, Tamkang University, and the Weidenbaum Center on the Economy, Government, and Public Policy to enable this international collaboration is gratefully acknowledged for funding. We are saddened that the second and the third authors passed away during the research collaboration process. The usual disclaimer applies.
* Corresponding author.

https://doi.org/10.1016/j.pirs.2024.100022
Received 11 December 2023; Received in revised form 16 April 2024; Accepted 16 April 2024

This is where we hope to contribute by providing a full-fledged general equilibrium model that allows us to determine under which conditions a high-tech city emerges as a *decentralized equilibrium outcome*. The main tenet of this paper is that *the emergence and efficiency of a tech cluster is intimately related to the spatial structure of the area that hosts it*. To show this, we view a tech cluster as a city formed by firms involved in R&D activities and which interact to determine endogenously knowledge spillovers. We echo Kerr and Robert-Nicoud (2020) by focusing on knowledge spillovers and R&D activities. In line with the empirical literature, we recognize that knowledge spillovers both depend on localization via a distance-decay function and raise the marginal product of R&D and production technologies.[1] These specific features of a high-tech city differentiate our model from typical city formation models where cities are marketplaces, transportation nodes, or factory towns. Endogenous knowledge spillovers give rise to varying strength of spatial externality which depend on the endogenous choice of firms' location within city. By its nature, a spatial externality takes the highest value under high-tech city structure. This motivates us to focus on this particular city structure in the benchmark framework.

Additionally, our model contrasts with the conventional ones by allowing the size of the high-tech city to be endogenous. This is an important feature enabling us to explain why tech clusters first grow and, then, decline. Despite several simplifications, it is our belief that our setting will contribute to a better understanding of the formation of high-tech cities and will allow for a more precise quantification of their economic consequences and a better evaluation of relevant policies.

To achieve our goal, we develop a model that captures the following basic features: (i) a composite consumption good is produced by using an endogenous range of specialized inputs provided by intermediate firms whose combination generates coordination problems that require the hiring of production-line designers; (ii) high-skilled workers are hired to produce intermediate goods or to conduct R&D in the intermediate sector; (iii) the productivity of an intermediate firm depends on its level of R&D investments and inter-firm spillovers, the intensity of which depends on how intermediate firms are distributed across space; and, (iv) both workers and intermediate firms are spatially mobile and use land. The novelty of our approach lies in an agglomeration force that combines firms' R&D investments and the existence of localized knowledge spillovers. On the other hand, the dispersion force, which is generated by both intermediate firms' and workers' demand for land and costly commuting, is common to most models of city formation. As usual, the equilibrium distribution of firms and workers is determined as the balance between these two opposite forces.

Our main results may be summarized as follows. We begin by establishing necessary and sufficient conditions for a high-tech cluster to emerge as a spatial equilibrium outcome. This in turn will allow us to uncover the reasons explaining why high-tech clusters may or may not be formed. More specifically, we identify three key rationales for firms to have incentives for clustering in a high-tech cluster: (i) *highly localized knowledge spillovers*, (ii) *relatively inexpensive commuting costs*, and (iii) *abundance of workers for research and tech production*. These are all typical features of new high-tech industries that make a cluster, which accommodates more intermediate firms and fosters research activities, more likely to emerge. This may explain why knowledge-intensive firms form a high-tech cluster such as the Silicon Valley, the Hsinchu Science-Based Industrial Park in Taiwan, the Cambridge Science Park in the UK, or the Nanjing high-tech cluster in the Yangtze River Delta Region (Saxenian, 1994; Chen, 2008; Arzaghi and Henderson, 2008; Li and Zhu, 2017) and why in the absence of localized knowledge spillovers, simple cluster policies are not sufficient for a high-tech city or a local innovative system to develop (see Duranton et al., 2010, for a critical appraisal of such policies). Although our analysis of spatial externalities remains relatively rudimentary, our paper may be viewed as an attempt to open the black box of spatial externalities by means of a setting that take firms' behavior as the driving force. For instance, we show that centrally located firms invest more in R&D while their output is higher. This result is driven by stronger spatial externality at central locations which leads to higher productivity of firms settled down there.

In our framework, a tech cluster hosts an intermediate sector in which firms' TFP is mainly driven by the aggregate R&D expenditure. One may inquire what drives R&D investments in a tech cluster. We find such R&D activities rise with stronger distance decay, weaker knowledge spillover externality, a lower coordination cost, or a larger skilled labor pool.

We further show that abundant availability of workers fosters a larger tech cluster if one emerges, whereas highly localized spillovers leads to a smaller tech cluster but makes it more likely to arise in equilibrium. By contrast, continual improvements in infrastructure and communication technology that lowers distance decay in knowledge spillover and coordination costs may lead to the fragmentation of tech clusters as observed in the case of Silicon Valley (Saxenian, 1994). In addition, tech clusters should not be viewed as the panacea for local development. Since they offer high wages, they attract workers. However, too large an inflow of migrants may lead to their geographical fragmentation when high land rents and commuting costs become a concern to potential entry of small specialized firms. This may explain why most successful tech clusters or science parks do not emerge within big cities but in their vicinity.[2]

While Buzard et al. (2017) document the strong geographical concentration of high-tech sectors, the meta-analysis of 168 empirical papers undertaken by Grashof (2020) suggests that high-tech firms tend to benefit more than low- or medium-tech firms from positive cluster effects (see also Gornig and Schiersch, 2024). Tech clusters are obvious examples of such groups of firms that should benefit from geographical proximity. Interestingly, no single variable is able to explain why clusters exist. As observed by Grashof (2020), clusters emerge when several distinct factors are combined. Our analysis confirms this observation in that the combination of the above-mentioned three factors is likely to promote the emergence of a tech cluster. Another interesting takeaway is that factors fostering the emergence of a tech cluster may not boost a larger cluster with more high-tech firms and may not sustain the cluster to avoid its eventual fragmentation. This may add insight toward understanding the large variation observed in the patterns of tech clusters or science parks across time and space, which concurs with Grashof (2020) who points out a great variety of real-world situations.

1.1. Related literature

Clusters have attracted the attention of researchers and policy makers alike, but the analytical literature is meager (Belleflamme et al., 2000; Duranton et al., 2010). We summarize the few related models that have been developed in urban economics and highlight how our paper differs from them. Their common aim is to explain the emergence of business centers within cities. They include O'Hara (1977), Ogawa and Fujita (1980), Fujita and Ogawa (1982), Berliant et al. (2002), and Lucas and Rossi-Hansberg (2002). These models focus on gravity-like reduced forms in which knowledge spillovers and/or spatial externalities are mechanically presumed. An exception is Berliant et al. (2002) in which aggregate capital affects spillovers, following the pioneering work by Romer (1986). Because tech clusters are knowledge-intensive, our primary driver is firms' R&D activities, so that spillovers are in terms of aggregate R&D activities. Importantly, R&D

[1] See Arzaghi and Henderson (2008), Greenstone et al. (2010), Lychagin et al. (2016), and Buzart et al. (2017), as well as critical reviews by Audretsch and Feldman (2004) and Carlino and Kerr (2015).

[2] For example, Silicon Valley is over 30 miles south of San Francisco, the Cambridge Science Park in the UK more than 60 miles north of London, and Hsinchu Science-Based Industrial Park in Taiwan 50 plus miles away from Taipei.

behaves very differently from capital in spatial settings: while capital is fully mobile and paid at a capital rental independent of locations, R&D labor exerts costly commuting and is paid by firms at a location-dependent wage. As a result, R&D labor influences the spatial pattern differently from capital. Further, we explicitly model the supply of intermediate inputs and business services such as production line design and coordination, which affect the overall productivity of firms (see Peter and Ruane, 2023, who highlight the importance of intermediate inputs). In this respect, we recognize that combining a wider range of intermediates is a more complex task at the firm level, which requires additional resources (Becker and Murphy, 1992).

Although spatial externalities have attracted a lot of attention in urban economics (see Carlino and Kerr, 2015, for a survey), we are not aware of a microeconomic model that investigates how spatial externalities emerge from *intentional* firms' R&D and location decisions. Here it is worth mentioning Desmet and Rossi-Hansberg (2014) who study how firms choose to innovate in a dynamic technology diffusion framework characterized by spatial frictions. However, unlike us, they do not consider workers' commuting costs, which play an important role here in pinning down the pattern of a tech cluster and explain why most successful tech-clusters or science parks do not emerge in mega-cities. Another key difference is that they cannot characterize equilibrium analytically, so they have to source to quantitative analysis.

Different from all papers above, our model allows the size of the tech cluster to be endogenously determined, which yields new comparative statics and serves to explain the rise and the fall of tech clusters. Furthermore, most of our results are determined analytically while most existing papers appeal to numerical analyses.

The remainder of the paper is organized as follows. The model is presented in Section 2, while Section 3 describes the necessary and sufficient conditions for a high-tech city to emerge, Section 4 examines its properties and the conditions for fragmentation of the high-tech city. Section 5 concludes.

2. The model

The economy consists of a featureless one-dimensional space Z and a continuum H of skilled workers. Land density and the opportunity cost of land are normalized to one. The intermediate sector produces a differentiated intermediate good under monopolistic competition using land and skilled labor with endogenous R&D that enhances firm productivity. Each intermediate is provided by a single firm and each firm supplies a single intermediate good. Intermediate firms and workers choose their location within the urban area and consume a fixed amount of land normalized to one. Since a firm located at z makes the same choices in equilibrium regardless of the variety it produces, we index intermediate inputs by their place of production z.

A unit mass of firms produce the final good under perfect competition by using the basket of intermediate goods. The final and intermediate goods are shipped within the city at no costs, thus implying that the prices of intermediate goods are independent of where intermediate and final producers are located.[3] The final producers locate at city center. This assumption allows us to focus on symmetric distributions of intermediate firms and workers.[4]

Apart from land, skilled workers consume the final good. Each worker is endowed with one unit of labor that she supplies inelastically. Commuting between the residence $y \in Z$ and workplace $z \in Z$ requires $t|y - z|$ units of the numeraire, where $t > 0$ is the commuting rate.

Since each worker and each firm consumes one unit of land, the total demand for land is equal to $N + H$ where N is the mass of intermediate firms that will be endogenously determined in equilibrium. Therefore, the city is given by the interval

$$Z = \left[-\frac{N+H}{2}, \frac{N+H}{2}\right],$$

which implies that the city size varies with the mass of intermediate firms (N) and the population size (H). Following the literature, we assume that the city is geographically symmetric and centered at $z = 0$.

We focus on a high-tech city where all firms are clustered around city center, which is flanked by two residential areas.[5] In this case, the intermediate producers are uniformly distributed within a tech cluster whose spatial extent is given by,

$$Z^F = [-N/2, N/2],$$

with density $h^F = 1$. Therefore, workers are uniformly distributed over the residential area

$$Z^W = \left[-\frac{N+H}{2}, -\frac{N}{2}\right] \cup \left[\frac{N}{2}, \frac{N+H}{2}\right]$$

with density $h^W = 1$.[6]

2.1. The final sector

The final sector produces the numeraire according to the following production function:

$$Y = \int_{Z^F} x(z)^{\frac{\sigma-1}{\sigma}} \mathrm{d}z, \tag{1}$$

where $\sigma > 1$, and maximizes profits given by

$$\Pi = Y - \int_{Z^F} p(z)x(z)\mathrm{d}z - w(0)\phi N, \tag{2}$$

subject to (1). In (2), $w(0)\phi N$ stands for the cost of designing the production line where $\phi > 0$ units of skilled labor is needed to use a new intermediate input. This cost has the nature of an endogenous fixed cost for the final sector. Final production exhibits decreasing returns to scale so that coordination and learning costs can be covered.

The final sector chooses the mass N of intermediate goods and the quantity x of each variety. Since shipping the intermediate goods to the final sector is costless, the price of an intermediate good i is the same within the city. Therefore, plugging (1) in (2) and differentiating Π yields the profit-maximization conditions:

$$\begin{aligned} \frac{\mathrm{d}\Pi}{\mathrm{d}x} &= \frac{\sigma-1}{\sigma}x(z)^{-\frac{1}{\sigma}} - p(z) = 0, \\ \frac{\mathrm{d}\Pi}{\mathrm{d}N} &= x(N/2)^{\frac{\sigma-1}{\sigma}} - w(0)\phi - p(N/2)x(N/2) = 0. \end{aligned} \tag{3}$$

Thus, the final sector's inverse demand for intermediate variety produced at z is location-specific and given by,

$$p(z) = \frac{\sigma-1}{\sigma}x(z)^{-\frac{1}{\sigma}}. \tag{4}$$

Plugging (4) into (3), we obtain the output of intermediate firm located farthest away from the city center:

$$x(N/2)^{\frac{\sigma-1}{\sigma}} = \sigma w(0)\phi. \tag{5}$$

2.2. The intermediate sector

A firm located at $z \in Z^F$ produces the intermediate good $z \in [-N/2, N/2]$ and uses one unit of land, $R(z)$ units of skilled labor

[3] Dealing with shipping costs is fairly straightforward and does not add much to our results.

[4] Using another location for the final sector does not affect the nature of our main results due to free trade of the final good.

[5] In Section 4.2 we further discuss conditions under which high-tech city is fragmented into multiple clusters.

[6] We could allow a mass of unskilled workers to reside in outskirts and be employed in the final good sector. Such an extension would not change any of our main findings.

for R&D, and $L(z)$ units of skilled labor for production. Then, skilled workers conduct R&D and produce (high-tech) intermediate good. This firm produces the quantity $x(z)$ according to a Cobb-Douglas technology:

$$x(z) = A(z) \cdot L(z)^{\beta} l^{1-\beta}, \tag{6}$$

where $\beta \in (0, 1)$ and $l = 1$ is a firm's land requirement. Land is then a fixed factor of production, which implies that firms face decreasing returns to scale.

The total factor productivity (TFP) of an intermediate firm $A(z)$ is given by a Cobb-Douglas aggregator of the firm's R&D employment $R(z)$ and a *spatial externality* that combines a Romer-Lucas external effect weighted by a distance-decay function $S(z)$:

$$A(z) = R(z)^{\theta} \cdot (\bar{R}S(z))^{\alpha}, \tag{7}$$

where $\bar{R}$ is the total mass of workers involved in R&D activities, $S(z)$ is a measure of firms' spatial concentration around z with $\theta \in (0, 1)$, and $\alpha \in (0, 1)$ stands for the strength of knowledge spillovers. Our setup of the endogenous TFP captures Romer's idea that externalities enter the system through an aggregator of individual firms' decisions, but we recognize that a firm's TFP is also influenced by this firm's R&D policy. Firms' endogenous decisions of R&D interacting with location-dependent knowledge spillovers via $\bar{R}S(z)$, which differentiates our paper from all previous studies cited in the introduction.

One of the distinctive features of our setting is that a firm's TFP is endogenized through the following three channels: (i) the number of researchers hired to conduct R&D in each firm, $R(z)$, (ii) the total number of researchers working the city, $\bar{R}$, and (iii) the spatial distribution of firms/researchers captured by $S(z)$. This modeling strategy is consistent with the idea that spillovers increase the level and productivity of R&D investments (Aghion and Jaravel, 2015; Helmers, 2019). It is also in line with Chyi et al. (2012) who find that knowledge spillovers are especially strong across intermediate firms located in science parks. That said, it is worth stressing that *one feature of a high-tech city is the presence of strong endogenous R&D shifter $R^{\theta}\bar{R}^{\alpha}$ in* (7). In this context, the efficiency of an intermediate firm depends on its level of R&D activities, as well as on the total mass of researchers whose productivity rises with their number (this was highlighted by Romer, 1986, and Lucas, 1988). While the urban economics literature highlights the existence of various types of agglomeration economies, the impact of the spatial structure on the endogenous R&D shifter $R^{\theta}\bar{R}^{\alpha}$ is often overlooked although its importance is stressed by empirical evidence (e.g., Siegel et al., 2003).

In (7), the spatial externality at z is generated by both the size of the research pool in the intermediate sector, $\bar{R}$, and the spatial distribution of intermediate firms, $S(z)$. In line with most of the literature, we assume that the distance-decay effect is given by a negative exponential (Fujita and Ogawa, 1982; Lucas and Rossi-Hansberg, 2002; Desmet and Rossi-Hansberg, 2014; de Palma et al., 2019):

$$S(z) \equiv \int_{Z^F} \mathrm{e}^{-\gamma|z-y|} h^F(z) \mathrm{d}y = \frac{1}{\gamma}[2 - \mathrm{e}^{-\gamma N/2}(\mathrm{e}^{-\gamma z} + \mathrm{e}^{\gamma z})], \tag{8}$$

since $h^F = 1$. Admittedly, $S(z)$ has the nature of a reduced-form that aims to capture a rich set of interactions. In this respect, we want to stress that Smith (1978) has provided micro-foundations for (8) that went unnoticed. He shows that, when knowledge flows are drawn randomly, the distance-decay function that describes the spatial diffusion of knowledge across space is given by a negative exponential if and only if a low-cost knowledge flow between two locations is more likely to be observed than a high-cost flow. Whereas $S(z)$ decreases with the spatial impedance parameter γ that measures the severity of distance-decays of positive spatial externality, it increases at a decreasing rate with the mass of intermediate firms N that gives rise to an agglomeration force encouraging firm clustering. Note also that $S(z)$ is strictly decreasing and concave in z.

Since workers bear commuting costs, the equilibrium wage rate $w(z)$ is location-specific. The immobility of land implies that the land rent $r(z)$ is also location-specific. Taking wages and land rents as given, each firm chooses its location z, output price $p(z)$, and the numbers of workers allocated to its production and R&D activities to maximize profits:

$$\pi(z) = p(z)x(z) - w(z)[R(z) + L(z)] - r(z). \tag{9}$$

The land rent $r(z)$ plays the role of an endogenous and location-specific fixed cost for an intermediate firm located at z. If land were not an input of the intermediate sector, firms would have a zero size under free entry and would not invest in R&D activity. This is reminiscent of Desmet and Rossi-Hansberg (2012).

Plugging (4) and (6) into (9) yields the following profit function:

$$\pi(z) = \frac{\sigma-1}{\sigma}[R(z)^{\theta} \cdot (\bar{R}S(z))^{\alpha} \cdot L(z)^{\beta}]^{\frac{\sigma-1}{\sigma}} - w(z)[R(z) + L(z)] - r(z). \tag{10}$$

Applying the first-order conditions with respect to $R(z)$, $L(z)$, and z leads to the following equations:

$$w(z) = \theta\left(\frac{\sigma-1}{\sigma}\right)^2 \frac{1}{R(z)} x(z)^{\frac{\sigma-1}{\sigma}}, \tag{11}$$

$$w(z) = \beta\left(\frac{\sigma-1}{\sigma}\right)^2 \frac{1}{L(z)} x(z)^{\frac{\sigma-1}{\sigma}}, \tag{12}$$

$$\alpha\left(\frac{\sigma-1}{\sigma}\right)^2 \frac{\mathrm{d}S(z)}{S(z)\mathrm{d}z} x^{\frac{\sigma-1}{\sigma}}(z) = (R(z) + L(z))\frac{\mathrm{d}w(z)}{\mathrm{d}z} + \frac{\mathrm{d}r(z)}{\mathrm{d}z}.$$

The first two expressions are standard. The last one means that, by moving away from the center, a firm incurs a decrease in the benefit generated by spillovers, which is exactly compensated by a decrease in wage and land rent. This condition is the counterpart of the Alonso-Muth equation obtained in the monocentric city model of urban economics (Fujita, 1989).

We can combine these three conditions to obtain the following equilibrium conditions:

$$\frac{R(z)}{L(z)} = \frac{\theta}{\beta}, \tag{13}$$

$$w(z) = \beta\left(\frac{\sigma-1}{\sigma}\right)^2 \frac{1}{L(z)} x(z)^{\frac{\sigma-1}{\sigma}}, \tag{14}$$

$$\frac{\mathrm{d}r(z)}{\mathrm{d}z} = \frac{L(z)}{\beta}\left[\alpha\frac{w(z)}{S(z)}\frac{\mathrm{d}S(z)}{\mathrm{d}z} - (\theta+\beta)\frac{\mathrm{d}w(z)}{\mathrm{d}z}\right]. \tag{15}$$

The first condition means that the research-production labor ratio is equal to the ratio of their output elasticities, while the second states that the wage is proportional to the output per capita of a variety. The third condition implies that the land rent gradient is flatter when the spillover S is stronger because the benefit earned from clustering is higher.

It remains to show that the first-order conditions are also sufficient. We show in Appendix A that firms' profit functions are strictly concave if and only if the inequality

$$\frac{\sigma}{\sigma-1} > \theta + \beta \tag{16}$$

holds. In what follows, we assume that (16) holds true.[7]

Using (11)–(12), the profit earned by a firm located at z is given by,

$$\pi(z) = \left(\frac{\sigma}{\sigma-1} - \theta - \beta\right)\frac{L(z)w(z)}{\beta} - r(z),$$

which is nonnegative due to (16).

[7] Since the elasticity of production with respect to R&D θ is about 0.12 (Hall et al., 2010), while the land share is approximately 0.06 for manufacturing (Caselli and Coleman, 2001) and 0.15 for services (Brinkman et al., 2015), we may safely conclude that the inequality $\sigma/(\sigma-1) > \theta + \beta$ holds for any admissible value of $\sigma > 1$.

There is free-entry to the market, which leads to zero profits in equilibrium. Therefore, the maximum bid that a firm can offer to set up at z, denoted by $r^F(z)$, is obtained from the zero-profit condition $\pi(z) = 0$:

$$r^F(z) = \left(\frac{\sigma}{\sigma - 1} - \theta - \beta\right)\frac{L(z)w(z)}{\beta}. \tag{17}$$

This defines the bid rent by an intermediate firm at production site z. Differentiating (17) with respect to z and using the envelop theorem yield:

$$\frac{\mathrm{d}r^F(z)}{\mathrm{d}z} = \left(\frac{\sigma}{\sigma - 1} - \theta - \beta\right)\frac{L(z)}{\beta}\frac{\mathrm{d}w(z)}{\mathrm{d}z}.$$

Plugging (15) in this expression, we obtain the following differential equation in w:

$$\frac{\mathrm{d}w(z)}{\mathrm{d}z} = \frac{\alpha(\sigma - 1)}{\sigma}\frac{w(z)}{S(z)}\frac{\mathrm{d}S(z)}{\mathrm{d}z}.$$

Solving this differential equation yields:

$$w(z) = C\cdot[S(z)]^{\frac{\alpha(\sigma-1)}{\sigma}}, \tag{18}$$

where $C > 0$ is the constant of integration given by

$$C = \left[\frac{\gamma}{2(1 - \mathrm{e}^{-\gamma N/2})}\right]^{\frac{\alpha(\sigma-1)}{\sigma}} w(0),$$

where we used (5). Since $S(z)$ is strictly decreasing and strictly concave in z and $\alpha(\sigma - 1)/\sigma < 1$, *the equilibrium wage schedule is strictly decreasing and strictly concave in z*. In other words, the wage falls at an increasing rate as the distance to the center of the cluster increases.

Equalizing (14) and (18) and using (6), (7), and (13) yields

Using (13) and (18), we obtain after simplifications the equilibrium mass of production workers hired by a firm located at z:

$$L(z) = \left[\left(\frac{\beta}{\theta}\right)^{(\alpha+\theta)(\sigma-1)}\left(\frac{C}{\beta}\right)^{\sigma}\left(\frac{\sigma}{\sigma - 1}\right)^{2\sigma}\cdot N^{-\alpha(\sigma-1)}\right]^{\frac{1}{(\alpha+\theta+\beta)(\sigma-1)-\sigma}} \equiv L^*, \tag{19}$$

which is independent of its location z.

Labor is allocated to the following three activities: (i) the mass $\bar{R}$ of workers involved in R&D activities, (ii) the mass ϕN of workers involved in designing the supply chain of the final sector, and (iii) the mass $\bar{L}$ of workers producing the intermediate goods. Since $\bar{L} = NL^*$, and $\bar{R} = NR^* = (\theta/\beta)NL^*$, the labor market clearing, $H = \phi N + \bar{R} + \bar{L}$, implies

$$L^* = \frac{\beta}{\beta + \theta}\left(\frac{H}{N} - \phi\right), \qquad R^* \equiv \frac{\theta}{\theta + \beta}\left(\frac{H}{N} - \phi\right). \tag{20}$$

Although firms hire the same number of researchers, $R^*(N)$, regardless of their locations, a firm's investment in R&D, which is equal to $w^*(z)R^*$, decreases with the distance to the cluster center because $w^*(z)$ decreases with z. Furthermore, using (6) and (7), it is readily verified that the equilibrium output of a firm located at z is equal to

$$x^*(z) = \left(\frac{\theta}{\beta}\right)^{\alpha+\theta}\cdot(NS(z))^{\alpha}\cdot(L^*)^{\alpha+\beta+\theta}. \tag{21}$$

In other words, *firm size decreases with the distance to the cluster center*. Thus, firms located closer to the cluster center invest more in R&D which, together with a stronger spillover effect, makes them more productive and leads to higher output. In addition, the higher the elasticity of production with respect to R&D (θ), the more researchers are employed by each firm, which produces more output. To put it differently, when R&D matters more (higher θ), each firm invests more in R&D and also produce more.

Equation (21) also shows that the mass of firms affects the firm size in two opposite ways. On the one hand, it reduces employment at each firm (lower L^*). On the other hand, it raises the extent of spatial externality. Substituting (20) into this expression, it is readily verified that the employment effect dominates, and thus the equilibrium size of a firm decreases with the mass of firms.

Using (19), we may rewrite (17) as follows:

$$r^F(z) = \left(\frac{\sigma}{\sigma - 1} - \theta - \beta\right)\frac{L^*}{\beta}w(z), \tag{22}$$

which implies that the bid rent of an intermediate firm at z is proportional to the wage that prevails at this location. Since $w(z)$ decreases in z, (22) also implies that firms' bid rent function also decreases in z. As firms closer to cluster center are more productive, their operating profits are higher, which allows them to pay higher land rent at central locations.

From the arguments above, we obtain:

Proposition 1. (Firm's R&D decisions) Firms located closer to the cluster center invest more in R&D, have bigger size and large operating profits.

This proposition shows how the diffusion process of knowledge affects firms' decisions through higher benefits for centrally located firms. In other words, knowledge spillovers benefit more centrally located firms which results in higher output and larger operating profits.

2.3. Workers

A worker chooses her residential site y and workplace z to maximizes her consumption of the final good. This amounts to maximizing her net income given by

$$\max_{y,z} I(y, z) = w(z) - t|y - z| - r(y). \tag{23}$$

At the residential equilibrium, workers reach the same utility level, hence earn the same net income I_0:

$$I(y, z) = I_0, \tag{24}$$

where I_0 is endogenous. This condition implies that a worker has no incentive to change either her residential or working places.

We determine the equilibrium mapping J from Z to Z that associates a (potential) job site $J(y) = z$ with a (potential) residential location y. This mapping describes the commuting pattern of workers. More specifically, a worker residing at y works at the location $J(y) = z$ that maximizes her net income:

$$w[J(y)] - t|y - J(y)| = \max_{z\in Z}[w(z) - t|y - z|], \quad y \in Z.$$

Solving workers' problem shows that the maximum bid a worker can offer to reside at location y is given by,

$$r^W(y) = \max_z\{w[J(y)] - t|y - J(y)| - I(y, J(y))|I(y, J(y)) = I_0\}. \tag{25}$$

This therefore defines the bid rent by a worker residing at location y.

The following result is intuitively obvious: in equilibrium, cross-commuting does not occur. Indeed, if two groups of workers cross-commute, any worker belonging to any of these groups would strictly increase her net income by choosing a job site in the area where the other group works. In other words, $J(y)$ increases in y.

Combining (23) and (24) with (25), we obtain:

$$r^W(y) = w(z) - t|y - J(y)| - I_0. \tag{26}$$

Equation (26) means that *workers' bid rent functions are linearand downward slopping in distance.*

3. Spatial equilibrium

The equilibrium land rent $r^*(z)$ is the upper envelop of the two bid rent functions $r^F(z)$ and $r^W(z)$. In other words, whenever a firm or a worker locates at z, its bid rent must be equal to the equilibrium land rent:

$$r^*(z) = \max\{r^F(z), r^W(z), 1\}$$

$$r^*(z) = \begin{cases} r^F(z) & \text{if } h^F(z) > 0 \\ r^W(z) & \text{if } h^W(z) > 0 \end{cases}$$

$$r^*(-(N+H)/2) = r^*((N+H)/2) = 1,$$

where the bid rents $r^W(z)$ and $r^F(z)$ are given by (26) and (22), respectively. Since both $r^W(z)$ and $r^F(z)$ decrease with z, the equilibrium land rent also decreases as the distance to the center rises.

A *spatial equilibrium* is defined by the quantity vector $\{L^*, R^*, x^*(z), N^*\}$ and the price vector $\{p^*(z), w^*(z), r^*(z)\}$ for $z \in Z$ such that the following conditions are satisfied: (i) profits are zero in the final and intermediate sectors; (ii) land and labor markets clear; (iii) market clearing for intermediate goods pins down the mass of intermediate firms; (iv) population constraint: $\int_{Z^W} h^W(z)\mathrm{d}z = H$. The Walras Law implies that the final good market clears.

3.1. Formation of a high-tech city

We begin by establishing necessary and sufficient conditions under which a high-tech city is a spatial equilibrium.

We have seen that the workers' bid rent schedule is linear and downward sloping while the firms' bid rent schedule is strictly decreasing and concave in z. In this case, the following two conditions are necessary and sufficient for a monocentric structure to be sustained as a spatial equilibrium:

(a) firms' and workers' bid rent schedules intersect at $z = N^*/2$;
(b) firms outbid workers at $z = 0$.

Indeed, condition (a) is necessary and sufficient to guarantee that workers outbid firms outside the tech cluster, while condition (b) means that firms outbid workers within the tech cluster. It remains to check what conditions (a) and (b) are for a high-tech city to emerge as an equilibrium outcome (see Fig. 1).

We next determine conditions for (b),

$$r^F(0) > r^W(0) = 1 + t\frac{H+N}{2}, \tag{27}$$

to hold. That is, intermediate firms outbid workers toward the center of the high-tech city – a necessary condition for intermediate firms to cluster around city center. Using condition (a), we show in Appendix B that the firms' bid rent at $z = 0$ is given by

$$r^F(0) = \left(1 + \frac{tH}{2}\right)\left(2\frac{1 - \mathrm{e}^{-\gamma N/2}}{1 - \mathrm{e}^{-\gamma N}}\right)^{\frac{\alpha(\sigma-1)}{\sigma}}. \tag{28}$$

Using this expression, (27) may be rewritten as follows:

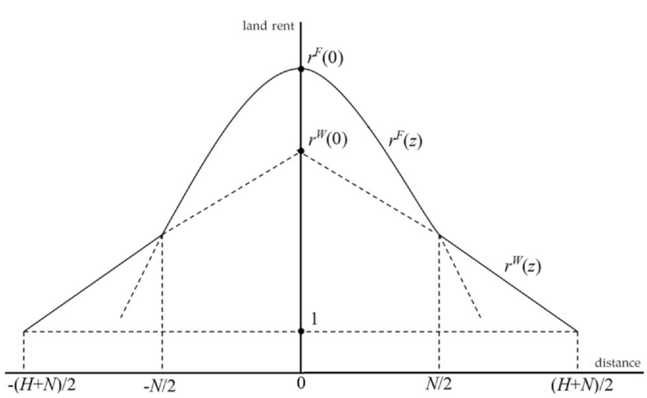

Fig. 1. Bid rent functions.

$$\left(2\frac{1 - \mathrm{e}^{-\gamma N/2}}{1 - \mathrm{e}^{-\gamma N}}\right)^{\frac{\alpha(\sigma-1)}{\sigma}} > \frac{2 + t(H+N)}{2 + tH}. \tag{29}$$

The right-hand side of (29) is independent of γ, while the left-hand side increases with γ and is larger than 1 for all $N > 0$. Therefore, there exists a threshold value $\bar{\gamma}$ such that for $\gamma > \bar{\gamma}$, (29) holds.

We can further characterize the threshold decay function, $\bar{\gamma}(t, \alpha, H)$, depending on commuting, the strength of knowledge spillovers and the pool of workers for R&D and tech production. The left-hand side of (29) is independent of t and larger than 1 for all $N > 0$, while the right-hand side is increasing in t and equal to 1 at $t = 0$. Therefore, lower commuting costs decreases $\bar{\gamma}$ which leads to formation of high-tech city for less localized spillovers. Since the right-hand side of (29) shifts downward when H increases, the value of $\bar{\gamma}$ decreases as well. Furthermore, as the left-hand side of (29) is larger than 1 for all $N > 0$, stronger knowledge spillovers (higher α) leads to (29) to hold for lower values of $\bar{\gamma}$. The analysis above leads to:

Proposition 2. (Formation of a high-tech city) For any given mass N of intermediate firms, a high-tech city is a spatial equilibrium if and only if spillovers are strongly localized, $\gamma > \bar{\gamma}$. Furthermore, a lower commuting cost t, a bigger pool of workers (higher H), or stronger knowledge spillovers (higher α) fosters the formation of high-tech city.

This proposition is consistent with the empirical evidence that high-tech firms have a higher propensity to cluster (Grashof, 2020). First, the information transmitted across high-tech firms is likely to be complex and, therefore, subject to strong distance-decay effects (high γ). Second, such firms also benefit more from stronger knowledge spillovers (high α), echoing the emphasis in the science park literature by Saxenian (1994) and Chyi et al. (2012).

Proposition 2 also highlights the need for modern and efficient transportation and communication infrastructures in a tech cluster.[8] Furthermore, a bigger pool of workers for research and tech production is more likely foster a tech cluster, which explains why tech clusters usually locate near major universities.

3.2. Endogenous size of tech clusters

We turn next to determining the equilibrium number of intermediate firms and hence the endogenous size of tech clusters.

We show in Appendix C that the equilibrium number of intermediate firms that sustains a high-tech city solves the following equation:

$$F(N) \equiv \frac{\sigma\phi}{\beta}\left(\frac{\sigma}{\sigma-1}\right)^2\left(2\frac{1 - \mathrm{e}^{-\gamma N/2}}{1 - \mathrm{e}^{-\gamma N}}\right)^{\frac{\alpha(\sigma-1)}{\sigma}} = \frac{\beta}{\beta+\theta}\left(\frac{H}{N} - \phi\right) \equiv G(N). \tag{30}$$

These lead to:

Proposition 3. (Endogenous size of tech clusters) The equilibrium mass of intermediate firms is unique, satisfying $\Gamma(N) \equiv F(N) - G(N) = 0$.

Proof. The function $F(N)$ is increasing in N with $F(0) > 0$, while $G(N)$ decreases in N with $G(0) \to \infty$ and $\lim_{N\to\infty} G(N) < 0$. Therefore, (30) has a unique solution, which yields the equilibrium mass of firms N^* in the intermediate sector. □

4. Characterization of the equilibrium

We are ready for characterizing the equilibrium featuring a high-tech city. We start with the size of tech clusters and its R&D activities. We establish the following proposition.

[8] This is in accordance with Ogawa and Fujita (1980) who show in a different setting that a monocentric city emerges when commuting costs are sufficiently small.

Proposition 4. (The size of tech clusters) In a high-tech city, a bigger labor pool (higher H), a weaker distance-decay effect (lower γ), weaker knowledge spillover externality (lower α), or a lower coordination cost (lower ϕ) leads to a higher mass of intermediate firms and a larger size of tech clusters.

Proof. It is straightforward to show that $\Gamma(N)$ defined in Proposition 3, an increasing function of N, is shifted upward with γ, α, and ϕ but downward with H. □

We summarize our findings on R&D activities in a high-tech city in the following proposition and then discuss intuition behind both propositions.

Proposition 5. (R&D investment) In a high-tech city, a bigger labor pool (higher H), a stronger distance-decay effect (higher γ), stronger knowledge spillover externality (higher α), or a lower coordination cost (lower ϕ) induces firms to invest more in R&D and results in high R&D activities in the high-tech city as a whole.

Proof. First, from (20), $R^* = (\theta/(\theta+\beta))(H/N^* - \phi)$ is decreasing in N^* and does not involve γ. Therefore, since N^* decreases with γ and α, R^* increases with γ and α. Second, substituting (20) into (30) shows that $R^* = (\theta/\beta)F(N^*)$, which is an increasing function of N^*. Since this expression does not include H and ϕ, Proposition 4 implies that N^* increases with H and decreases with ϕ, and thus the same holds for R^*. Third, since both R^* and N^* increase with H and lower ϕ, the same holds for aggregate R&D, $\bar{R} = N^*R^*$. Finally, aggregate R&D may be expressed as $\bar{R} = N^*R^* = (\theta/(\theta+\beta))(H - \phi N^*)$, which is decreasing in N^*, therefore, it increases with both γ and α. □

Intuitively, abundant availability of workers or lower coordination cost enables high-tech production with a longer production line, thus inducing more intermediate firms to cluster which hires more R&D employees.

While the effect of a bigger labor pool is straightforward, the intuition behind the impacts of γ and α is worth noting. A weaker distance-decay effect counters the necessary dispersion of intermediate firms as a result of required land usage and the rising wage cost to compensate longer commuting by workers, thereby permitting a larger tech cluster to form in equilibrium. However, it also leads to fragmentation of high-tech cluster if it falls below $\bar{\gamma}$. Then, lower γ yields larger but less sustainable high-tech city. The first reason is that the spatial externality (8) is very localized under low γ. Second, it increases very slowly with N especially at the edge of the cluster, $S(N/2)$, and an increasing number of firms mitigates even more an increase in $S(z)$. Simultaneously, each firm invests less in R&D such that overall pool of R&D employees decreases. All these effects eventually reduce firms' TFP (7) and lead to a fragmentation of high-tech cluster. Contrast to that, higher γ yields geographically compact cluster with smaller number of firms which invest more in R&D such that overall R&D pool is larger. This improves the sustainability of tech cluster because distance-decay effect is strong especially within proximity of city center.

Finally, weaker knowledge spillover externality (lower α) has a similar effect. Indeed, the direct effect of lower α reduces firms' TFP via reduction in knowledge spillovers. Then, it produces the same impact as a more localized distance-decay effect. It fosters firm to invest less in R&D, thus, there are more but smaller firms. As the size of tech cluster expands, it becomes less sustainable due to weaker knowledge spillovers at the edge of the cluster which leads to its fragmentation. To sum up, *stronger distance-decay effect or knowledge spillover externality makes a high-tech city more likely to form, but the equilibrium city size is smaller.*

4.1. The fragmentation of tech clusters

In this section, we investigate conditions under which high-tech cluster decentralizes into multiple clusters. To address this question, it is convenient to define $t^* \equiv T(N^*)$ as the commuting cutoff such that (29) holds for equality, i.e.,

$$\left(2\frac{1-\mathrm{e}^{-\gamma N^*/2}}{1-\mathrm{e}^{-\gamma N^*}}\right)^{\frac{\alpha(\sigma-1)}{\sigma}} = \frac{2+t^*(H+N^*)}{2+t^*H} \tag{31}$$

A high-tech is more likely to sustain if the inequality (29) continues to hold true, or, equivalently, if $t < t^*$. Plugging the equilibrium mass of firms N^* determined by (30) into the above expression yields a relationship between N^* and t^*:

$$\left(\frac{\sigma-1}{\sigma}\right)^2 \frac{\beta^2}{\sigma\phi(\beta+\theta)}\left(\frac{H}{N^*}-\phi\right) = \frac{2+t^*(H+N^*)}{2+t^*H}. \tag{32}$$

Totally differentiating this expression and solving for $\mathrm{d}t^*/\mathrm{d}N^*$, we obtain:

$$\frac{\mathrm{d}t^*}{\mathrm{d}N^*} = -\frac{(2+t^*H)^2}{2N^*}\left[\left(\frac{\sigma-1}{\sigma}\right)^2 \frac{\beta^2}{\sigma\phi(\beta+\theta)}\frac{H}{(N^*)^2} + \frac{t^*}{2+t^*H}\right] < 0.$$

In other words, the commuting cutoff t^* decreases with the equilibrium mass of intermediate firms N^*. Note that labor requirement for production-line designers/coordinators does not affect the commuting cutoff directly. Thus, Proposition 5 implies that a lower coordination cost ϕ would raise the mass of intermediate firms and reduce t^* which eventually yields fragmentation of a high-tech city. More specifically, consider an economy with continual improvements in infrastructure and communication technology that lowers coordination costs ϕ. At an initial equilibrium (N^*, t^*), the condition for the formation of the high-tech city $t < t^*$ is met. With continual reduction in ϕ, the size of the high-tech city continues to grow until N^{**} at which $t = T(N^{**})$ holds, after which it ceases to grow even if ϕ continues to fall. At this stage a single high-tech city will be decentralized into multiple ones. This *fragmentation of tech clusters* may serve to explain the evolution of the Silicon Valley, which began with Palo Alto Industrial Park in 1951, expanded to Mountain View, then to Sunnyvale, Santa Clara and, eventually, San Jose and its adjacent municipals (Saxenian, 1994, pp. 29–30).

Recall from Proposition 4 that a weaker distance-decay effect (lower γ) makes a high-tech city larger (higher N^*). Rearranging (32) yields:

$$\frac{1}{N^*}\left[\left(\frac{\sigma-1}{\sigma}\right)^2 \frac{\beta^2}{\sigma\phi(\beta+\theta)}\left(\frac{H}{N^*}-\phi\right)-1\right] = \frac{t^*}{2+t^*H},$$

which is not directly affected by γ. Since the left-hand side of this expression is a decreasing function of N^* and hence an increasing function of γ while the right-hand side increases with t^*, the commuting cutoff decreases with a weaker distance-decay effect (lower γ). In other words, continual reduction in distance decay in knowledge spillovers can lead to fragmentation of tech clusters. Similar argument applies to α but the force behind a continual change in knowledge spillover externality is less obvious and hence not elaborated.

By contrast, the effect of a larger labor pool is *a priori* ambiguous. Indeed, for a given mass N of intermediate firms, Proposition 2 implies that a higher value of H raises the commuting cutoff t^*. However, Proposition 4 implies that a bigger labor pool also induces a larger equilibrium mass of intermediate firms. This in turn reduces the commuting cutoff. Yet, we show in the Appendix D that the cutoff t^* may decrease with the size of the labor pool when t^* takes intermediate values. Specifically, when the labor pool is sufficiently large, there exist two commuting rates $0 < t_1 < t_2$ such that $\mathrm{d}t^*/\mathrm{d}H < 0$ holds over (t_1, t_2). Thus, when $t^* \in (t_1, t_2)$, a further increase in the labor pool leads to a lower cutoff $t^{**} < t^*$. Should the new cutoff t^{**} fall below the current commuting cost rate t, the city would turn out to host too many workers to sustain its tech cluster.

The above arguments imply:

Proposition 6. (Fragmentation of tech clusters) In a high-tech city, continual reduction in distance decay (falling γ) or in coordination cost (falling ϕ) leads to fragmentation of tech clusters, whereas the impact of continual expansion of the labor pool (rising H) is generally ambiguous.

4.2. Numerical exercises

While our theory provides some insights toward understanding the rise and fall a tech cluster, its size and the R&D investment generated, the effect of continual expansion of the labor pool remains ambiguous. Moreover, one may also wonder how responsive the fragmentation process is to the reduction in distance decay and coordination cost. To address these questions, we source to numerical analysis.

We first normalize $H = 1$. Then, based on Peter and Ruane (2023), the elasticity of substitution σ among inputs is about 3, which gives $\frac{\sigma-1}{\sigma} = 2/3$, leaving a potential rent of 33.3%. The elasticity of production θ with respect to R&D is about 0.12 (Hall et al., 2010). Since the land share is approximately 0.06 for manufacturing (Caselli and Coleman, 2001) and 0.15 for services (Brinkman et al., 2015), we set the land share to be 0.1 and hence set $\beta = 0.9$. Thus, the required condition for strictly concave firm profit function, $\sigma/(\sigma - 1) > \theta + \beta$, is met (precisely, $\sigma/(\sigma - 1) - \theta - \beta = 0.48$). The strength of the Romer externality may range from 10% to 20%. In the case of a high-tech city, we choose the higher end and set $\alpha = 0.2$. While there is no direct estimate of γ, knowledge spillover distance decay should be sufficiently high as empirical evidence suggest strong localization. We thus set $\gamma = 2$. Given these parameters, we calibrate ϕ such that total production-line cost is about 30% of total revenue Y. Specifically, by setting $\phi = 0.01$, the production-line cost share become 0.306, close to the target. Finally, to ensure the formation of a high-tech city, we choose reasonably low commuting cost with $t = 0.01$, under which commuting cost from $z = N/2$ to $z = 0$ is about 10.7% of the wage paid at the center, which falls in the range of 9–11.4% as documented by Redding and Turner (2015).

Under this benchmark parametrization, a high-tech city is formed with $N^* = 9.687$. Notably the mass of workers and firms are normalized based on their unit demand for land. Should an average firm use 500 time more land than a worker, workers per firm should be adjusted to $500H/N^* = 51.6$. The size of the tech cluster is 10.7 and the commuting cutoff is $t^* = T(N^*) = 0.02$, above the commuting cost of $t = 0.01$. In this benchmark, the effect of continual expansion of the labor pool for research and tech production is fragmentation of the tech cluster. The result is robust to reasonable changes in parameters not fully based on previous studies, (α, γ, ϕ) (see Appendix E).

We turn next to examine the fragmentation process in response to, hypothetically, a continual 5% reduction annually in distance decay (the solid curve in red in Fig. 2) or coordination cost (the dotted curve in blue) or a continual 5% annual increase in the labor pool (the dash curve in green). One can see from Fig. 2 much quicker fragmentation with coordination cost reduction (15 years) and labor pool expansion (16 years), but slower with distance decay reduction (47 years). Throughout this process, the equilibrium mass of intermediate firms N^*, which stands for the size of the tech cluster, rise sharply in response to coordination cost reduction (the dotted curve in blue in Fig. 3) or labor pool expansion (the dash curve in green), but only modestly in response to distance decay reduction (the solid curve in red). These explain why the tech cluster is less likely to sustain in response to the two former changes.

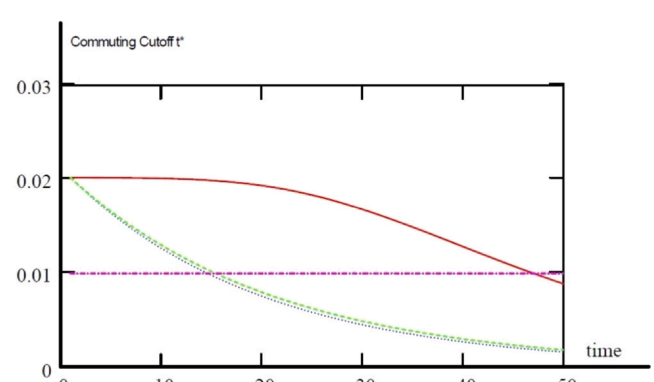

Fig. 2. Fragmentation of tech clusters.

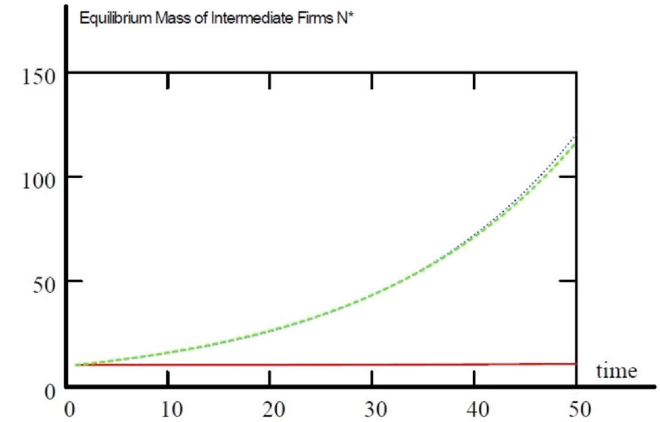

Fig. 3. Growth of tech clusters.

4.3. Taking stock

In summary, the main drivers for the rise and fall and the size of tech clusters and the accompanied R&D investment are:

	t	γ	α	ϕ	H
Rise of tech clusters	−		+	0	+
Size of tech clusters		−	−	−	+
Fall of tech clusters		+		+	?
R&D		+	−	−	+

Notes: "−" stands for a negative effect, "+" a positive one, and "?" an ambiguous outcome.

The findings yield the following takeaways. First, modern and efficient transportation and communication infrastructures, strong knowledge spillovers, sizable pool of workers for research and tech production purposes are key drivers for the rise of a tech cluster. Second, while lower distance decay or coordination cost help expand the size of a tech cluster, continual reduction in either would cause the eventual fall of the tech cluster, leading to fragmentation. Finally, a tech cluster generate more R&D investments with stronger distance decay, weaker spillover externality, a lower coordination cost, or a larger labor pool.

4.4. On tech cluster promotion policies

One may inquire whether the typical tech cluster promotion policies may work effectively. Based on our analysis, the answer is it depends.

One of the most common instruments used by governments is to reduce firms' startup costs which can be translated into the reduction in the coordination cost for setting up the production line. Our analysis above suggests that implementing such a policy need not achieve its goals.

By contrast, subsidizing land may incentivize firms to cluster in a large tech cluster. More specifically, with subsidy Δr, the equilibrium land rent becomes $r_j(z) = \max\{r_j^F(z) + \Delta r, r_j^W(z)\}$. This generates a discontinuity in firms' bid rents at the border between industrial and residential areas, thus enabling firms to outbid workers and to occupy more land within the inner area of the high-tech city. Having illustrated briefly the working of this land subsidy policy, a full welfare analysis is required in order to pin down optimal subsidy by maintaining neutral government revenues.

5. Concluding remarks

In this paper, we have developed a spatial equilibrium model of a high-tech city in the presence of positive knowledge spillovers between monopolistically competitive intermediate firms. We have shown that a tech cluster hosts an intermediate sector in which firms' TFP is mainly driven by the aggregate R&D expenditure. With this proviso, a high-tech city is more likely to form under strongly localized knowledge spillovers, skilled labor abundance, and low commuting costs. We have also shown why continual improvements in infrastructure and communication technology that lowers coordination costs or mitigates distance decay in knowledge spillovers may lead to eventual fragmentation of tech clusters.

We have considered a closed city model in which the total population is fixed. A promising step for future research is to determine the equilibrium population size of the tech cluster by using an open city setting in which workers are free to migrate in and out while the utility level is exogenously given by the best option available in the rest of the world. In this way, firms' incentive to cluster will interact with workers' incentive to migrate, potentially leading to richer equilibrium outcomes. Another important extension is to dig further into the black box of spillovers by assuming with Davis and Dingel (2019) that face-to-face contacts across workers are costly and freely chosen, and to determine the condition for a high-tech city to emerge in such a social environment. Equally important, like in Behrens et al. (2014), our setting should be extended to deal with an urban system that involves heterogeneous workers, as well as cities that host a tech cluster and cities that do not. These are likely to be rewarding avenues for future research.

Declaration of Competing Interests

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Appendix A. Second-order condition

To show that firms' profit functions are strictly concave in (R, L), we compute the second derivatives of (10):

$$\frac{\partial^2\pi}{\partial R^2(z)} = \theta\left(\theta\frac{\sigma-1}{\sigma}-1\right)\left(\frac{\sigma-1}{\sigma}\right)^2\frac{x^{\frac{\sigma-1}{\sigma}}(z)}{R(z)^2} < 0,$$

$$\frac{\partial^2\pi}{\partial L^2(z)} = \beta\left(\beta\frac{\sigma-1}{\sigma}-1\right)\left(\frac{\sigma-1}{\sigma}\right)^2\frac{x^{\frac{\sigma-1}{\sigma}}(z)}{L(z)^2} < 0,$$

$$\frac{\partial^2\pi}{\partial R(z)\partial L(z)} = \beta\theta\left(\frac{\sigma-1}{\sigma}\right)^3\frac{x^{\frac{\sigma-1}{\sigma}}(z)}{L(z)R(z)}.$$

Since the Hessian matrix $H(z)$ of $\pi(z)$ is given by,

$$H = \begin{pmatrix} \theta\left(\theta\frac{\sigma-1}{\sigma}-1\right)\left(\frac{\sigma-1}{\sigma}\right)^2\frac{x^{\frac{\sigma-1}{\sigma}}(z)}{R(z)^2} & \beta\theta\left(\frac{\sigma-1}{\sigma}\right)^3\frac{x^{\frac{\sigma-1}{\sigma}}(z)}{L(z)R(z)} \\ \beta\theta\left(\frac{\sigma-1}{\sigma}\right)^3\frac{1}{L(i,z)R(i,z)}x_j^{\frac{\sigma-1}{\sigma}}(i, z) & \beta\left(\beta\frac{\sigma-1}{\sigma}-1\right)\left(\frac{\sigma-1}{\sigma}\right)^2\frac{x^{\frac{\sigma-1}{\sigma}}(z)}{L(z)R(z)} \end{pmatrix}.$$

the second-order condition holds if and only if $|H(z)| > 0$, that is,

$$|H| = \theta\beta\left(\theta\frac{\sigma-1}{\sigma}-1\right)\left(\frac{\sigma-1}{\sigma}\right)^4\frac{x^{2\frac{\sigma-1}{\sigma}}(z)}{L(z)^2R(z)^2}\left(\beta\frac{\sigma-1}{\sigma}-1\right) - \left[\beta\theta\left(\frac{\sigma-1}{\sigma}\right)^3\frac{x^{\frac{\sigma-1}{\sigma}}(z)}{L(z)R(z)}\right]^2,$$

or, after simplifications,

$$|H| = \beta\theta\left(\frac{\sigma-1}{\sigma}\right)^5\left(\frac{\sigma}{\sigma-1}-\theta-\beta\right)\frac{x_j^{2\frac{\sigma-1}{\sigma}}(i, z)}{R(i, z)^2L(i, z)^2} > 0,$$

which is equivalent to (16). Q.E.D.

B. Firm's bid rent

We begin by determining the value of the constant of integration by equalizing firms' and workers' bit rents at $N/2$ (condition (a)):

$$r^F(N/2) = \left(\frac{\sigma}{\sigma-1}-\theta-\beta\right)\frac{L^*(N)}{\beta}w(N/2) = 1 + \frac{tH}{2} = r^W(N/2). \tag{B.1}$$

Plugging (18) evaluated at $z = N/2$ into (B.1), we obtain:

$$\left(\frac{\sigma}{\sigma-1}-\theta-\beta\right)\frac{CL^*}{\beta}\left(\frac{1-e^{-\gamma N}}{\gamma}\right)^{\frac{\alpha(\sigma-1)}{\sigma}} = 1 + \frac{tH}{2}, \tag{B.2}$$

which pins down the unique value of the constant of integration C.

Plugging (18) in (22) for $z = 0$ and using (B.2) to replace CL^* yields (28).

C. Proof of Proposition 2

Evaluating (18) at $z = 0$ yields:

$$w(0) = C\left(2\frac{1-e^{-\gamma N/2}}{\gamma}\right)^{\frac{\alpha(\sigma-1)}{\sigma}}.$$

Plugging this expression into (5), we obtain:

$$x(N/2)^{\frac{\sigma-1}{\sigma}} = \sigma\phi C\left(2\frac{1-e^{-\gamma N/2}}{\gamma}\right)^{\frac{\alpha(\sigma-1)}{\sigma}}. \tag{C.1}$$

Using (6), we obtain the equilibrium output at $N/2$:

$$x^*(N/2) = \frac{\theta}{\beta}(L^*)^{1+\beta}N^\alpha S(N/2)^\alpha = \frac{\theta}{\beta}(L^*)^{1+\beta}N^\alpha\left[\frac{1}{\gamma}(1-e^{-\gamma N})\right]^\alpha, \tag{C.2}$$

Combining (C.1) and (C.2) leads to

$$C^\sigma = \left(\frac{\theta}{\beta}\right)^{(\alpha+\theta)(\sigma-1)}\cdot\frac{1}{(\sigma\varphi)^\sigma}\left[\frac{1-e^{-\gamma N}}{2(1-e^{-\gamma N/2})}\right]^{\frac{\alpha(\sigma-1)}{\sigma}}N^{\alpha(\sigma-1)}\cdot(L^*)^{(\alpha+\beta+\theta)(\sigma-1)}.$$

Substituting C^σ into (19) and simplifying yields:

$$L^* = \frac{\sigma\phi}{\beta}\left(\frac{\sigma}{\sigma-1}\right)^2\left[\frac{2(1-e^{-\gamma N/2})}{1-e^{-\gamma N}}\right]^{\frac{\alpha(\sigma-1)}{\sigma}}. \tag{C.3}$$

Equalizing (20) and (C.3) leads to (30) that pins down the equilibrium mass of intermediate firms.

D. The impact of H on $T(N^*)$

Differentiating (30) and (32) with respect to H yields at N^*:

$$F'(N)N' = \frac{N-HN'}{N^2},$$

$$\left(\frac{\sigma-1}{\sigma}\right)^2\frac{\beta^2}{(\beta+\theta)\sigma\phi}\frac{N-HN'}{N^2} = \frac{(t^{*'}N+t^*N')(2+t^*H)-t^*N(t^{*'}H+t^*)}{(2+t^*H)^2},$$

or, after simplifications,

$$N' = \frac{1}{NF'(N)+\frac{H}{N}},$$

$$\frac{2t^{*'}N}{(2+t^*H)^2} = \frac{Nt^{*2}}{(2+t^*H)^2} + \left(\frac{\sigma-1}{\sigma}\right)^2\frac{\beta^2}{(\beta+\theta)\sigma\phi}\frac{1}{N} - \left[\left(\frac{\sigma-1}{\sigma}\right)^2\frac{\beta^2}{(\beta+\theta)\sigma\phi}\frac{H}{N^2}+\frac{t^*}{(2+t^*H)}\right]N'.$$

Combining these two equations and simplifying, we obtain:

$$2t^{*'}N\left(NF'(N)+\frac{H}{N}\right) = N^2t^{*2}F'(N) + \left(\frac{\sigma-1}{\sigma}\right)^2\frac{\beta^2(2+t^*H)^2}{(\beta+\theta)\sigma\phi}F'(N) - 2t^*,$$

or, equivalently,

$$\begin{aligned} 2(N^2F'(N)+H)t^{*'} &= \Psi(t^*) \\ &\equiv F'(N)\left[N^2+H^2\left(\tfrac{\sigma-1}{\sigma}\right)^2\tfrac{\beta^2}{(\beta+\theta)\sigma\phi}\right]t^{*2} \\ &\quad -2\left[1-2H\left(\tfrac{\sigma-1}{\sigma}\right)^2\tfrac{\beta^2}{(\beta+\theta)\sigma\phi}F'(N)\right]t^* + 4\left(\tfrac{\sigma-1}{\sigma}\right)^2\tfrac{\beta^2}{(\beta+\theta)\sigma\phi}F'(N). \end{aligned} \tag{D.1}$$

Since $\Psi(t^*)$ is a quadratic function of t^*, it has at most two roots t_1 and t_2 with $t_1 < t_2$. We now determine conditions for both roots are positive. Note $\Psi(t^*)$ is minimized at

$$\bar{t} = \frac{1}{F'(N)}\frac{1-2H\left(\frac{\sigma-1}{\sigma}\right)^2\frac{\beta^2}{(\beta+\theta)\sigma\phi}F'(N)}{N^2+H^2\left(\frac{\sigma-1}{\sigma}\right)^2\frac{\beta^2}{(\beta+\theta)\sigma\phi}}.$$

Since $\Psi(0) > 0$ and $\Psi(\cdot)$ is convex, $0 < t_1 < t_2$ holds if $\bar{t} > 0$ and $\Psi(\bar{t}) < 0$. Since $F'(N) > 0$, the former holds if

$$F'(N) < \left(\frac{\sigma}{\sigma-1}\right)^2\frac{(\beta+\theta)\sigma\phi}{2H\beta^2}. \tag{D.2}$$

Plugging $\bar{t}$ in (D.1), we obtain after simplifications:

$$2\left[N^2+H^2\left(\frac{\sigma-1}{\sigma}\right)^2\frac{\beta^2}{(\beta+\theta)\sigma\phi}\right]t^{*'}(\bar{t}) = 4\left(\frac{\sigma-1}{\sigma}\right)^2\frac{\beta^2}{(\beta+\theta)\sigma\phi} - \frac{1}{F'(N)(N^2F'(N)+H)}. \tag{D.3}$$

Since (D.2) implies that the highest admissible value of $F'(N)$ is given by,

$$F'(N) = \left(\frac{\sigma}{\sigma - 1}\right)^2 \frac{(\beta + \theta)\sigma\phi}{2H\beta^2},$$

it then follows from (D.3) that $t^{*'}(\bar{t}) < 0$ when H is sufficiently large. By implication of (A.1), $\Psi(\bar{t}) < 0$ when H is large enough.

E. Robustness checks

Three parameters set in the quantitative analysis are not fully based on previous studies or calibrated to observable data, namely, (α, γ, ϕ). In this appendix, we perform sensitivity analysis regarding $\pm$ 20% changes in each of these parameters. The results are summarized in the following:

	city size	commuting cutoff	years to fragmentation with 5% annual		
	$N^* + H$	t^*	reduction in γ	reduction in ϕ	increase in H
Benchmark	10.687	0.020	47	15	16
$\alpha \cdot 1.2$	10.527	0.025	52	19	20
$\alpha \cdot 0.8$	10.850	0.016	41	10	11
$\gamma \cdot 1.2$	10.686	0.020	51	15	16
$\gamma \cdot 0.8$	10.688	0.020	43	15	16
$\phi \cdot 1.2$	9.073	0.024	49	19	20
$\phi \cdot 0.8$	13.109	0.016	45	10	11

Thus, the main findings all remain valid.

References

Aghion, P., Jaravel, X., 2015. Knowledge spillovers, innovation and growth. Econ. J. 125, 533–573.
Arzaghi, M., Henderson, J.V., 2008. Networking off Madison Avenue. Rev. Econ. Stud. 75, 1011–1038.
Audretsch, D., Feldman, M., 2004. Knowledge spillovers and the geography of innovation. In: Henderson, J.V., Thisse, J.-F. (Eds.), Handbook of Urban and Regional Economics, Volume 4. North-Holland, Amsterdam, pp. 2713–2739.
Becker, G.S., Murphy, K.M., 1992. The division of labor, coordination costs, and knowledge. Q. J. Econ. 107, 1137–1160.
Behrens, K., Duranton, G., Robert-Nicoud, F., 2014. Productive cities: sorting, selection and agglomeration. J. Political Econ. 122, 507–553.
Belleflamme, P., Picard, P., Thisse, J.-F., 2000. An economic theory of regional clusters. J. Urban Econ. 48, 158–184.
Berliant, M., Peng, S.-K., Wang, P., 2002. Production externalities and urban configuration. J. Econ. Theory 104, 275–303.
Brinkman, J., Coen-Pirani, D., Sieg, H., 2015. Firm dynamics in an urban economy. Int. Econ. Rev. 56, 1135–1164.
Buzard, K., Carlino, G.A., Hunt, R.M., Carr, J.K., Smith, T.E., 2017. The agglomeration of American R&D labs. J. Urban Econ. 101, 14–26.
Carlino, G.A., Kerr, W.R., 2015. Agglomeration and innovation. In: Duranton, G., Henderson, J.V., Strange, W. (Eds.), Handbook of Regional and Urban Economics. Volume 5. Elsevier, Amsterdam, pp. 349–404.
Caselli, F., Coleman II, W.J., 2001. The U.S. structural transformation and regional convergence: a reinterpretation. J. Political Econ. 109, 584–616.
Castells, M., Hall, P.G., 1994. Technopoles of the World: The Making of Twenty-first-century Industrial Complexes. Routledge, London.
Chen, T.-J., 2008. The emergence of Hsinchu Science Park as an IT cluster. In: Yusuf, S., Nabeshima, K., Yamashita, S. (Eds.), Growing Industrial Clusters in Asia: Serendipity and Science. The World Bank, Washington, DC, pp. 67–89.
Chyi, Y.L., Lai, Y.M., Liu, W.H., 2012. Knowledge spillovers and firm performance in the high-technology industrial cluster. Res. Policy 41, 556–564.
Davis, D.R., Dingel, J., 2019. A spatial knowledge economy. Am. Econ. Rev. 109, 153–170.
de Palma, A., Papageorgiou, Y.Y., Thisse, J.-F., Ushchev, P., 2019. About the origin of cities. J. Urban Econ. 111, 1–13.
Desmet, K., Rossi-Hansberg, E., 2012. Innovation in space. Am. Econ. Rev. Pap. Proc. 102, 447–452.
Desmet, K., Rossi-Hansberg, E., 2014. Spatial development. Am. Econ. Rev. 104, 1211–1243.
Duranton, G., Martin, Ph., Mayer, T., Mayneris, F., 2010. The Economics of Clusters. Oxford University Press, Oxford.
Fujita, M., 1989. Urban Economic Theory. Cambridge University Press, Cambridge.
Fujita, M., Ogawa, F., 1982. Multiple equilibria and structural transition of non-monocentric urban configurations. Reg. Sci. Urban Econ. 12, 161–196.
Gornig, M., Schiersch, A., 2024. Agglomeration economies: different effects on TFP in high-tech and low-tech industries. Reg. Stud. 1–12.
Grashof, N., 2020. Firm-specific cluster effects: a meta-analysis. Pap. Reg. Sci. 99, 1237–1261.
Greenstone, M., Hornbeck, R., Moretti, E., 2010. Identifying agglomeration spillovers: evidence from winners and losers of large plant openings. J. Political Econ. 118, 536–598.
Hall, B.J.M., Mairesse, Mohnen, P., 2010. Measuring the returns to R&D. In: Hall, B.H., Rosenberg, N. (Eds.), Handbook of the Economics of Innovation. Volume 2. North-Holland, Amsterdam, pp. 1033–1082.
Helmers, C., 2019. Choose the neighbor before the house: agglomeration externalities in a UK science park. J. Econ. Geogr. 19, 31–55.
Katz, B., Bradley, J., 2013. The Metropolitan Revolution. How Cities and Metros Are Fixing Our Broken Politics and Fragile Economy. Brookings, Washington, DC.
Kerr, W.R., Robert-Nicoud, F., 2020. Tech clusters. J. Econ. Perspect. 34, 50–76.
Klepper, S., 2010. The origin and growth of industry clusters: the making of Silicon Valley and Detroit. J. Urban Econ. 67, 15–32.
Li, Y., Zhu, K., 2017. Spatial dependence and heterogeneity in the location processes of new high-tech firms in Nanjing, China. Pap. Reg. Sci. 96, 519–536.
Lucas Jr., R.E., 1988. On the mechanics of economic development. J. Monet. Econ. 22, 3–22.
Lucas Jr., R.E., Rossi-Hansberg, E., 2002. On the internal structure of cities. Econometrica 70, 1445–1476.
Lychagin, S., Pinkse, J., Slade, M.E., Van Reenen, J., 2016. Spillovers in space: does geography matter? J. Ind. Econ. 64, 295–335.
Marshall, A., 1890. Principles of Economics. Macmillan, London 8th edition published in 1920.
McCarthy, I.P., Silvestre, B.S., vonNordenflycht, A., Breznitz, S.M., 2018. A typology of university research park strategies: what parks do and why it matters. J. Eng. Technol. Manag. 47, 110–122.
Porter, M.E., 2000. Location, competition, and economic development: local clusters in a global economy. Econ. Dev. Q. 14, 15–34.
Ogawa, M., Fujita, M., 1980. Equilibrium land use patterns in a nonmonocentric city. J. Reg. Sci. 20, 455–475.
O'Hara, D.J., 1977. Location of firms within a square central business district. J. Political Econ. 85, 1189–1207.
Peter, A. and C. Ruane (2023)The aggregate importance of intermediate input substitutability. NBER Working Paper No. 31233.
Redding, S.J., Turner, M.A., 2015. Transportation costs and the spatial organization of economic activity. In: Duranton, G., Henderson, J.V., Strange, W.C. (Eds.), Handbook of Regional and Urban Economics, Volume 5. North-Holland, Amsterdam, pp. 1339–1398.
Romer, P., 1986. Increasing returns and long-run growth. J. Political Econ. 94, 1002–1037.
Saxenian, A., 1994. Regional Advantage: Culture and Competition in Silicon Valley and Route 128. Harvard University Press, Cambridge, MA.
Siegel, D.S., Westhead, P., Wright, M., 2003. Assessing the impact of university science parks on research productivity: exploratory firm-level evidence from the United Kingdom. Int. J. Ind. Organ. 21, 1357–1369.
Smith, T., 1978. A cost-efficieny principle of spatial interaction behavior. Reg. Sci. Urban Econ. 8, 313–337.