Transport Policy 180 (2026) 104036

Contents lists available at ScienceDirect

Transport Policy

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

From connectivity to complexity: The influence of high-speed rail on urban knowledge complexity

Mingming Guan, Yuting Hou [*],[1]

Department of Building and Real Estate, The Hong Kong Polytechnic University, Hong Kong Special Administrative Region of China

ARTICLE INFO

Keywords:
High-speed rail
Knowledge complexity
Agglomeration externalities
Network externalities
Difference-in-differences

ABSTRACT

Knowledge complexity is a key determinant of regional competitiveness, yet the mechanisms and micro-level carriers through which transportation infrastructure shapes it remains insufficiently understood. This study examines the impact of high-speed rail (HSR) on knowledge complexity using patent and socio-economic data from 268 Chinese cities over 2005–2020, applying a multi-period difference-in-differences approach. Results show that HSR significantly enhances urban knowledge complexity, and the findings remain robust after addressing endogeneity concerns. Mechanism analysis reveals that HSR promotes complexity primarily through diversified agglomeration and network externalities, while specialized agglomeration has no significant effect. Moreover, HSR reshapes the relationship between agglomeration and network effects by substituting localized specialization with networked knowledge flows and enhancing the innovative potential of diversification through cross-regional complementarities. At the micro level, HSR triggers knowledge combination through two channels: a sharing mechanism that emphasizes collaborative interactions and collective knowledge externalities, and a matching mechanism that facilitates the strategic acquisition and recombination of external knowledge via technology transfers. By integrating agglomeration and network externality frameworks, this study provides empirical evidence on how HSR shapes urban knowledge complexity. The findings offer China-specific policy implications and transferable insights for regions pursuing innovation-driven growth through improved connectivity.

1. Introduction

As a transformative innovation in transportation, high-speed rail (HSR) has been widely adopted worldwide for its exceptional efficiency and sustainability. Following Japan's inauguration of the Shinkansen in 1964, several European countries, including France and Germany, established extensive HSR networks. As a latecomer, China launched its first HSR line in 2008 and, within just over a decade, has developed the world's largest and most comprehensive HSR network. The economic and social implications of HSR have attracted substantial academic attention. Existing studies have highlighted its profound impacts across multiple dimensions, including accessibility and travel behavior (Chan and Yuan, 2017; Shaw et al., 2014), economic growth (Ishikura, 2025; Yoo et al., 2023; Zhang et al., 2025; Zheng and Kahn, 2013), labor allocation (Feng et al., 2023; Guirao et al., 2018; Heuermann and Schmieder, 2019), and knowledge spillovers (Gao and Zheng, 2020; Hou, 2022; Miwa et al., 2022; Sun et al., 2023; Tang et al., 2021).

Despite the growing literature on HSR and knowledge spillovers, a significant gap remains, as our understanding of the mechanisms through which HSR influences complex knowledge recombination is still limited. Innovation, by its nature, is a complex and dynamic process. Schumpeter (1934) conceptualized innovation as the new combination of production factors. From the perspective of knowledge, knowledge complexity provides a comprehensive indicator for capturing this intricate process of innovation dynamics. It encompasses not only the number of knowledge combinations but also their richness and technological diversity (Balland and Rigby, 2017; Fleming and Sorenson, 2001; Sorenson et al., 2006). Knowledge complexity holds particular importance for regional development policy. Specifically, regions with higher levels of knowledge complexity are better positioned to reorganize complementary capabilities and explore new technological domains, thereby enhancing regional adaptability and long-term

---

[*] Corresponding author.

E-mail addresses: mingming.guan@connect.polyu.hk (M. Guan), yuting.hou@polyu.edu.hk (Y. Hou).

[1] Present/permanent address.: Department of Building and Real Estate, The Hong Kong Polytechnic University, 11 Yuk Choi Road, Hung Hom, Kowloon, Hong Kong.

https://doi.org/10.1016/j.tranpol.2026.104036
Received 30 May 2025; Received in revised form 23 December 2025; Accepted 23 January 2026

competitiveness (Balland and Rigby, 2017; Hidalgo, 2021; Mewes and Broekel, 2022). In light of increasing technological sophistication and the rapid expansion of transportation networks, understanding how HSR affects knowledge complexity is of both theoretical and practical significance. Academically, such inquiry helps uncover how HSR reshapes the evolutionary trajectories of regional knowledge systems, filling an essential gap in the literature. Practically, it provides evidence-based insights for HSR network planning and transport-oriented innovation policy design, which can strengthen the resilience, adaptability, and competitiveness of regional innovation systems.

As a high-speed transportation system, HSR generates a pronounced space–time compression effect, facilitating knowledge diffusion and enhancing knowledge search activities (Dong et al., 2020; Tan and Pan, 2024; Tang et al., 2021; Yao and Li, 2022). Although a conceptual link exists between HSR and knowledge recombination, several empirical gaps remain. First, there is a lack of a comprehensive framework to analyze how HSR influences knowledge complexity. A large body of existing research has predominantly interpreted the impact of HSR development on knowledge spillovers through the lens of localized agglomeration effects (Gao and Zheng, 2020; Hou, 2022; Miwa et al., 2022; Sun et al., 2023), while largely overlooking network externalities generated by HSR interconnectivity. These network effects can substantially reshape knowledge flows beyond the scope of localized agglomeration (Berger and Prawitz, 2024; Dong et al., 2020; Russell et al., 2024; Yao and Li, 2022). Second, prior research has insufficiently examined the interaction between agglomeration economies and network externalities, leaving unresolved questions regarding whether these effects are complementary or substitutive. Third, micro-level mechanisms through which HSR fosters knowledge recombination remain underexplored, particularly the mediating roles of innovative actors such as firms, universities, and research institutions. Addressing these gaps is essential for a holistic understanding of how HSR reshapes complex knowledge dynamics.

To fill the gaps in existing research, this paper empirically examines the impact mechanisms of HSR on knowledge complexity using panel data from 268 prefecture-level cities in China spanning 2005–2020. Methodologically, we combine the difference-in-differences (DID) approach with an instrumental variable (IV) regression to construct an empirical strategy that effectively mitigates endogeneity concerns, thereby providing a more accurate identification of the causal effect of HSR on knowledge complexity. Furthermore, we develop a conceptual framework that integrates agglomeration and network externalities and employ a mediation model to systematically explore the transmission channels through which HSR affects knowledge complexity. Building on this framework, we further investigate how agglomeration and network externalities interact to jointly shape the evolution of knowledge complexity. Finally, we identify the micro-level carrier of these mechanisms, which lies in technological collaboration and transfer among different innovators.

This study contributes to the existing literature in several aspects. First, we develop a conceptual framework that integrates HSR, agglomeration externalities, network externalities, and knowledge complexity into a unified analytical system. This framework elucidates the mechanisms through which HSR influences knowledge complexity, emphasizing the mediating roles of diversified agglomeration and network externalities. In doing so, it deepens our understanding of how HSR-driven connectivity enhances knowledge complexity and extends prior research on the determinants of knowledge complexity (Balland and Rigby, 2017; Boschma, 2017; Broekel et al., 2023; Fernhaber and Patel, 2012; Sorenson et al., 2006). Second, we provide novel empirical evidence for the interactive relationship between agglomeration and network externalities, a subject of long-standing theoretical debate (Burger and Meijers, 2016; Capello, 2000; Huang et al., 2020). Whereas earlier studies on HSR and innovation often examined these two externalities separately (Hou, 2022; Tan and Pan, 2024; Tang et al., 2021), this study empirically identifies their interactive effects, offering a more comprehensive perspective on how multiple externalities jointly shape knowledge complexity. Finally, this study advances the understanding of network externalities by highlighting their micro-level mechanisms. While prior research emphasizes their role in urban growth through borrowing scale effects (Alonso, 1975; Meijers et al., 2016), we show that HSR enhances knowledge complexity via inter-firm collaboration and technology transfer, enabling innovative actors to leverage network externalities for knowledge sharing and matching.

After the introduction, the paper is organized into five analytical sections. Section 2 reviews existing literature to develop research hypotheses, while Section 3 outlines our methodology. The following sections present the empirical framework: variable definition and data collection (Section 4), followed by results, endogeneity treatment, robustness test, mechanisms, and heterogeneous effects (Section 5). Section 6 elaborates on the contributions of the study, outlines its practical policy relevance, and proposes avenues for future scholarly exploration.

2. Theoretical analysis and research hypotheses

2.1. Theoretical analysis and conceptual framework

The relationship between transport and innovation can be traced back to the companion innovation hypothesis, which suggests that improved transport infrastructure stimulates innovative production and facilitates innovation diffusion across industries (Garrison and Souleyrette II, 1996). Knowledge recombination theory further emphasizes that innovation depends on the diversity and interdependence of knowledge components, with greater complexity requiring stronger interconnections and recombination (Fleming and Sorenson, 2001; Schumpeter, 1934). Linking these perspectives, HSR can create favorable conditions for complex knowledge recombination by promoting inter-organizational collaboration and interregional learning. From a mechanism perspective, agglomeration economy theory emphasizes that agglomeration serves as a key channel for knowledge spillovers (Jacobs, 1969; Marshall, 1890). By enhancing accessibility and reducing transaction costs, HSR optimizes the spatial allocation of innovation resources and strengthens knowledge agglomeration. Meanwhile, network externality theory suggests that the value of a network rises exponentially with its connectivity (Burt, 1987; Katz and Shapiro, 1985; Metcalfe, 1995). HSR connectivity expands intercity interactions and bridges previously isolated knowledge bases. At the micro level, HSR influences knowledge complexity through technological collaboration and transfer among diverse innovation actors. Integrating these theoretical insights, this study develops a conceptual framework (Fig. 1) that examines how HSR shapes knowledge complexity.

2.2. High-speed rail, knowledge spillovers, and knowledge complexity

Amidst the ongoing waves of globalization and the information technology revolution, the diffusion of codified knowledge has accelerated at an unprecedented rate. However, the spatial stickiness of complex knowledge remains a significant phenomenon (Balland et al., 2020; Balland and Rigby, 2017), creating inherent barriers to its inter-regional transfer. This persistence largely stems from two key factors. The first aspect lies in the inherent complexity of knowledge components. Knowledge production depends on the recombination of diverse elements, with higher knowledge complexity requiring a greater number of components and stronger interdependencies among them (Fleming and Sorenson, 2001; Sorenson et al., 2006). Second, the social embeddedness of knowledge transfer plays a critical role. Unlike codified explicit knowledge, complex knowledge often contains substantial tacit elements, which are difficult to formalize and require prolonged social interaction and practical engagement for effective transfer (Polanyi, 1962). Empirical studies have demonstrated that the successful transfer of tacit knowledge heavily depends on informal mechanisms

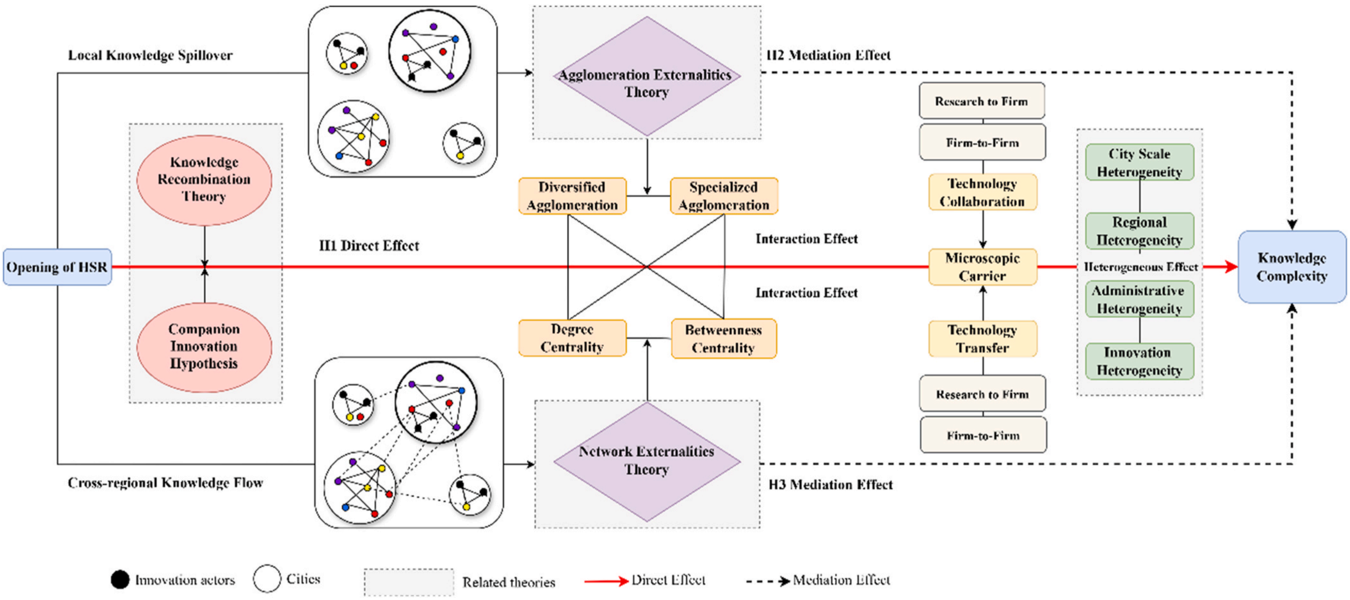


Fig. 1. Conceptual framework of the impact of HSR on knowledge complexity.

(Bathelt, 2007; Pérez-Luño et al., 2019; Storper and Venables, 2004), such as social networks, organizational trust, and face-to-face communication.

Transportation infrastructure serves as a critical enabler in shaping regional innovation ecosystems and facilitating knowledge diffusion. For example, empirical evidence suggests that the expansion of metro networks significantly enhances the intensity of intra-city knowledge collaboration (Koh et al., 2022). Furthermore, modern transportation infrastructure, including new air routes and HSR, generates time-space compression effects. These effects foster cross-regional face-to-face interactions and provide new channels for knowledge spillovers (Bai et al., 2024; Berger and Prawitz, 2024; Dong et al., 2020; Tan and Pan, 2024; Yao and Li, 2022). HSR reduces spatial friction and transaction costs, creating favorable conditions for knowledge spillovers and thereby facilitating the recombination of complex knowledge. Based on the above analysis, we formulate the following hypothesis:

H1. The expansion of HSR exerts a significantly positive influence on urban knowledge complexity.

2.3. High-speed rail, agglomeration externalities, and knowledge complexity

Agglomeration externalities, a central mechanism in explaining the dynamics of knowledge production, refers to the economic advantages firms derive from geographic proximity. These advantages include enhanced knowledge spillovers, increased opportunities for resource sharing, and reduced transaction costs (Marshall, 1890). The Marshall externalities theory posits that knowledge spillovers predominantly occur within industrial clusters characterized by technological similarity or close supply chain linkages (Marshall, 1890). In contrast, Jacobs' theory of externalities underscores that the extent and effectiveness of knowledge spillovers are primarily driven by the interplay and convergence of diverse industries (Jacobs, 1969). The significance of such externalities in shaping regional knowledge production is widely regarded as contingent upon several factors, including spatial scale, the characteristics of knowledge itself, and the specific innovation requirements of local industries (Beaudry and Schiffauerova, 2009; Caragliu et al., 2016; Huang et al., 2020).

The mechanisms through which agglomeration externalities shape knowledge complexity remain contested in the literature. Two dominant yet competing perspectives emerge regarding spatial clustering's role in knowledge recombination. The first perspective emphasizes knowledge diversity as fundamental to innovation ecosystems (Østergaard et al., 2011; Uzzi and Spiro, 2005). Proponents argue that varied knowledge bases provide essential raw materials for recombination, enhancing complexity through cross-pollination of ideas. Agglomeration, in this view, facilitates complexity by concentrating diverse knowledge stocks. Conversely, the second perspective underscores specialization's critical role (Balland et al., 2022; Hidalgo, 2021). This camp maintains that complexity arises not from breadth alone, but from deep, cumulative expertise within focused domains. Specialized agglomerations foster tight-knit professional networks that drive field-specific advances in knowledge complexity. This perspective suggests that knowledge complexity can be viewed as an extension of specialized capabilities.

In the context of economic geography, the sharp decline in transportation costs has enhanced intercity economic integration and simultaneously reshaped the spatial distribution and clustering tendencies of enterprises and human capital (Agrawal et al., 2017; Russell et al., 2024; Zheng and Kahn, 2013). HSR, as a fast and efficient transportation mode, exerts a substantial influence on the intra-urban spatial organization of economic activity. Nevertheless, scholarly debate persists over whether the introduction of HSR fosters patterns of specialized agglomeration or encourages greater diversification. On the one hand, the deployment of HSR can help cities attract more firms and talent with specialized advantages in certain fields, supporting the development of specialized agglomerations (Chen and Guo, 2023; Hou, 2022; Lin, 2017). On the other hand, the expansion of HSR enhances intercity connectivity, allowing firms and talent from diverse industries to gather more easily in a single city, thus accelerating cross-industry knowledge exchange and integration, and fostering diversified agglomeration (Tang et al., 2021; Yang and Ma, 2023). We propose the following hypotheses based on the above analysis and discussion.

H2. The commencement of HSR will promote agglomeration effects within cities, thereby increasing urban knowledge complexity.

H2a. The commencement of HSR promotes urban knowledge complexity through specialized agglomeration.

H2b. The commencement of HSR enhances urban knowledge complexity through diversified agglomeration.

2.4. High-speed rail, network externalities, and knowledge complexity

Network externalities, a concept initially introduced by Katz and Shapiro (1985), describes the phenomenon in which the utility derived from a product or service increases as the number of consumers using that product or service expands. Building on this foundation, Capello (2000) extends the framework to urban and regional networks, proposing that network effects operate through "borrowed size." Specifically, smaller cities or regions can integrate into larger urban networks to benefit from agglomeration economies, including knowledge spillovers and access to larger markets, while mitigating the negative externalities associated with high-density agglomerations (Alonso, 1975; Meijers et al., 2016). Transportation infrastructure is crucial in enabling cities to leverage borrowed size, allowing smaller urban centers to access the agglomeration benefits typically concentrated in large metropolitan areas.

In social network theory, structural hole theory emphasizes the importance of non-redundant ties in enabling the flow of information and the access to resources (Burt, 1987). This theory holds profound implications for innovation networks, where participation significantly enhances knowledge exchange, resource sharing, and collaboration among enterprises, universities, and research institutions, thereby improving innovation performance. An increasing body of empirical evidence highlights the crucial role of collaborative innovation in driving knowledge recombination (Berger and Prawitz, 2024; Uzzi and Spiro, 2005; Xiao et al., 2022). From the perspective of knowledge complexity, the core function of knowledge networks is primarily reflected in facilitating the systematic reorganization of knowledge and enhancing the knowledge absorption capacity of entities. On the one hand, cross-organizational collaboration fosters the integration of diverse knowledge bases, facilitating knowledge recombination (Fleming and Sorenson, 2001; Xiao et al., 2022). On the other hand, cooperative relationships within innovation networks enhance the absorptive capacity of participants (Chatzistamoulou et al., 2022). Collaborative innovation fosters diverse learning processes, such as imitative learning, reverse engineering, and joint R&D, allowing entities to better recognize, absorb, and leverage external knowledge.

The formation and evolution of innovation networks depend not only on geographical proximity but also on efficient transportation and communication infrastructure (Guan et al., 2022; Yao and Li, 2022). China's rapidly growing HSR system exhibits strong network externalities consistent with Metcalfe's Law that network value scales quadratically with node connectivity (Metcalfe, 1995). Thus, as the HSR network expands, its overall value grows exponentially, reinforcing network externalities. HSR facilitates intercity innovation collaboration and strengthens the structural hole effect within innovation networks. By enabling more actors to occupy structural hole positions, HSR allows them to leverage informational and positional advantages to access diverse knowledge and resources. For example, smaller cities can use HSR to establish cooperative relationships with larger cities or innovation hubs, gaining access to shared innovation resources (Dong et al., 2020; Henke et al., 2023; Tang et al., 2021; Yao and Li, 2022). Guided by the theoretical and empirical foundations established above, we advance the third core hypothesis:

H3. The commencement of HSR promotes intercity innovation cooperation, thereby increasing knowledge complexity through network externalities.

3. Methodology

3.1. Baseline regression model

This study leverages the staggered rollout of HSR systems across cities as a quasi-natural experiment and employs a multi-period DID strategy to evaluate the causal effects of HSR on knowledge complexity. This approach effectively simulates a quasi-experimental setting, enabling us to identify the causal impact of HSR on knowledge complexity while controlling for time-invariant unobserved heterogeneity and common temporal shocks. The formula is as follows:

$$Y_{it} = \alpha_0 + \beta_1 HSRDID_{it} + \gamma_k \sum CV_{it} + \mu_i + \nu_{pt} + \varepsilon_{it} \tag{1}$$

Where: Y_{it} is the dependent variable, representing knowledge complexity; $HSRDID_{it}$ is the independent variable, indicating whether HSR has been implemented; CV_{it} represents a set of control variables, which will be detailed below; μ_i captures city fixed effect, while ν_{pt} denotes province-by-year fixed effects, which control for all time-varying shocks common to cities within the same province. ε_{it} is the error term, capturing unobserved random factors.

3.2. Mediation effect model

This study examines the multifaceted mechanisms through which HSR impacts knowledge complexity, focusing on dual channels of influence: agglomeration and network externalities. Building on Baron and Kenny (1986) 's framework, we construct a mediation model to analyze the intervening roles of these dynamics.

$$M_{it} = \delta_0 + \delta_1 HSRDID_{it} + \delta_k \sum CV_{it} + \mu_i + \nu_{pt} + \varepsilon_{it} \tag{2}$$

$$Y_{it} = \gamma_0 + \gamma_1 M_{it} + \gamma_k \sum CV_{it} + \mu_i + \nu_{pt} + \varepsilon_{it} \tag{3}$$

$$Y_{it} = \varphi_0 + \varphi_1 HSRDID_{it} + \varphi_2 M_{it} + \varphi_k \sum CV_{it} + \mu_i + \nu_{pt} + \varepsilon_{it} \tag{4}$$

We examine mediation effects through a three-step procedure, where M_{it} captures agglomeration and network externalities, and control variables follow Model (1). Step 1: Estimate Model (2) to test whether $HSRDID_{it}$ significantly affects M_{it} (coefficient δ_1). Step 2: Estimate Model (3) to assess whether M_{it} influences Y_{it} (coefficient γ_1). Step 3: Estimate Model (4) to evaluate the direct effects of $HSRDID_{it}$ (coefficient φ_1) and M_{it} (coefficient φ_2) on Y_{it}. Mediation is established if: (1) δ_1 and γ_1 are significant, and (2) φ_1 diminishes or loses significance in Model (4). Full mediation occurs if φ_1 becomes insignificant, while partial mediation is inferred if φ_1 remains significant but attenuated.

4. Variables and data

4.1. Study area and data sources

This study focuses on China's urban system and its HSR network as the research setting. The nation's HSR infrastructure, having evolved

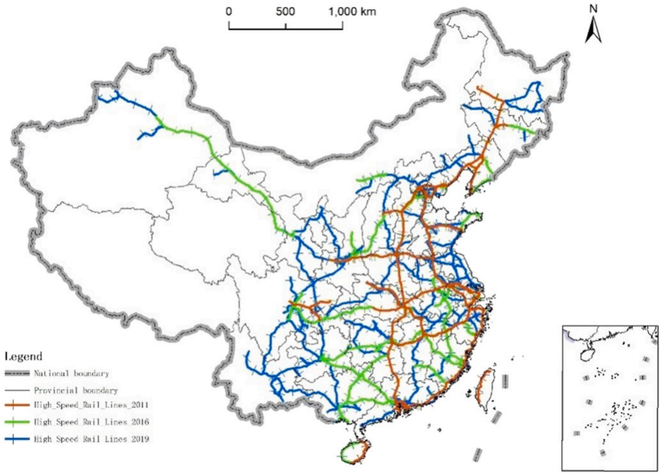


Fig. 2. Distribution of China's HSR network in 2011, 2016, and 2019.

from regional corridors to the world's most extensive network (Fig. 2), offers a compelling quasi-natural experiment for examining the co-evolution of HSR and knowledge complexity. This study focuses on China's prefecture-level cities and constructs a panel dataset for the period 2005 to 2020. To ensure comparability of the data, the following sample selection and processing procedures are applied: First, the centrally administered municipalities (Beijing, Shanghai, Chongqing, and Tianjin) were removed from the analysis due to their distinct roles within China's economic system. These cities function as provincial-level units, and as such, lack direct comparability with other prefecture-level cities. Second, to maintain sample consistency, cities located along the Qinhuangdao–Shenyang passenger line, including Qinhuangdao, Huludao, Jinzhou, Panjin, Anshan, and Shenyang, were excluded from the analysis.[2] Finally, to ensure data quality, cities with a significant number of missing values (e.g., Lhasa, Bijie, and Tongren) were excluded from the analysis. As a result, a balanced panel dataset comprising 268 prefecture-level cities is constructed.

Regarding data sources, patent text data were obtained from the China National Intellectual Property Administration (CNIPA). Specifically, invention patent applications are used as the primary raw data for measuring knowledge complexity, as they are widely recognized as high-quality innovations and among the most reliable indicators of technological advancement (Acs et al., 2002; Fleming and Sorenson, 2001; Griliches, 1998; Higham et al., 2021). Data on HSR openings were collected from the China National Railway Administration, while socio-economic variables were sourced from the China City Statistical Yearbook (2006–2021) and the statistical bureaus of relevant provinces. Information on patent collaboration and patent transfer was obtained from three sources: CNIPA, the Tianyancha database, and GAODE Map. By integrating applicant and assignee data from CNIPA, firm-level information from Tianyancha, and geocoding services from GAODE Map, we were able to accurately determine the detailed geographic locations of each patent entity.

4.2. *Variable definitions*

4.2.1. *Dependent variable*

Knowledge complexity captures the structural diversity, specialization, and interdependence of knowledge components within and across technological domains, reflecting the organization of a knowledge system (Balland and Rigby, 2017; Fleming and Sorenson, 2001). The measurement of knowledge complexity varies depending on the theoretical perspective and research purpose. Building on prior studies (Fleming and Sorenson, 2001; Ning et al., 2023), this study adopts the element combination method to directly measure knowledge complexity. The theoretical rationale of this approach is rooted in complexity economics and evolutionary theory. Specifically, the stronger the interdependence among knowledge elements, the more difficult knowledge recombination becomes, implying a higher level of complexity. Compared with traditional indicators such as patent counts, R&D expenditure, or patent citations, knowledge complexity emphasizes the structural richness and systemic interdependence of knowledge systems and serves as a more comprehensive and explanatory indicator of innovation activities (Balland and Rigby, 2017; Broekel, 2019; Hidalgo, 2021). The formula is as follows:

$$KCI_{c,t} = \frac{1}{\sqrt{N_{c,t}}}\sum_{l\in c,t} KCI_l = \frac{1}{\sqrt{N_{c,t}}}\sum_{l\in c,t}\frac{SC_l}{EA_{i(l),t}} \tag{5}$$

Where: KCI stands for the urban knowledge complexity index. KCL_l measures patent l's technological interdependence, where SC_l is the count of technological subclasses of patent, $EA_{i,t}$ denotes the recombinational ease of technology category i in year t. $N_{c,t}$ is the total number of patents in city c in year t. In other words, $KCI_{c,t}$ is a mean per patent, not a sum; therefore, it does not mechanically grow with city patent counts.

This study employs the four-digit International Patent Classification (IPC) system to identify recombination relationships among technological elements. The ease of recombination is defined as follows:

$$E_{i,t} = \frac{N_{sc,i,t}}{N_{pat,i,t}} \tag{6}$$

Where: $N_{sc,i,t}$ is the number of other technological categories combined with subclass i between 2000 and 2020, $N_{pat,i,t}$ represents the total number of patents containing technological subclass i during the same period. The value of $E_{i,t}$ changes over time. A higher value indicates a greater diversity of technologies being combined, while a lower value suggests fewer technological categories are being recombined.

4.2.2. *Independent variable*

This paper employs a DID model to estimate the causal effect of HSR on knowledge complexity. Accordingly, the treatment variable, HSRDID, is coded as a binary indicator, with a value of 1 representing cities connected by HSR and 0 representing non-connected cities. While a simple binary variable provides a straightforward identification of treatment status, it may underestimate the heterogeneous effects of HSR connectivity and network centrality across cities. To address this limitation, we construct alternative measures of HSR in the robustness test, which capture the connectivity and centrality of HSR connections more accurately.

4.2.3. *Mediation variables*

Specialized agglomeration. To quantify technological specialization patterns, we employ the Herfindahl-Hirschman index (HHI) maximum values across cities as our specialized agglomeration metric, calculated as:

$$SPE_{c,t} = \max\left[\frac{\left(T_{i,k}/T_k\right)}{\left(T_{i,t}/T_t\right)}\right] \tag{7}$$

Specifically, $T_{i,k}$ denotes the number of technologies in category k for city i; T_k is the national total in category k. $T_{i,t}$ and T_t represent the total technologies in city i and nationally, respectively. Technologies are classified using two-digit IPC codes to ensure consistent cross-city comparisons.

Diversified agglomeration. Building on the framework proposed by Frenken et al. (2007), we distinguish between related variety (RV) and unrelated variety (UV). RV measures the extent of relatedness within a city's knowledge base, offering a reflection of knowledge diversity across various levels of sectoral aggregation. The formula is as follows:

$$RV = \sum P_g * \left(\sum_{j\in S_g}\frac{P_j}{P_g} * ln\left(\frac{P_j}{P_g}\right)\right) \tag{8}$$

The methodology incorporates two core patent distribution metrics: P_g measures two-digit IPC category shares within cities, while P_j assesses four-digit subclass distributions within each category. These jointly inform the RV index, an entropy-based measure of technological diversification.

UV measures the diversity of knowledge that lacks a cognitive basis,

---

[2] These cities were excluded for two reasons. First, according to the Design Specification for High-Speed Railways in China (TB10621-2014), HSR refers to passenger-dedicated lines with a design speed of at least 250 km/h and an initial operating speed of no less than 200 km/h. The Qinhuangdao–Shenyang line, however, operated at an initial speed of only 160 km/h, which falls short of this standard and therefore does not qualify as HSR in a strict sense. Second, this line commenced operation in October 2003, preceding our sample period of 2005–2020. Including it could thus introduce inconsistencies into the temporal framework of the study.

reflecting differences across broader sectoral categories. The formula for UV is as follows:

$$UV = \sum_{g} P_g * \ln\left(\frac{1}{P_g}\right) \tag{9}$$

Where: P_g denotes the proportion of patent IPC two-digit categories within the city's technological classification. UV represents the entropy value of technological categories.

Network externalities. Network externalities can be systematically evaluated using various quantitative metrics, among which weighted degree and betweenness centrality represent two established measurement instruments (Huang et al., 2020; Tang et al., 2021). To operate this analytical framework, we construct a patent collaboration network specifically designed for quantifying these two critical indicators. Detailed methodological specifications and technical implementation procedures for network construction are provided in Appendix A.

Weighted degree (WD) serves as a network analysis metric that evaluates nodal prominence by aggregating the intensity of its linkages with neighboring nodes. This measure reflects a node's capacity to access and influence network resources. The mathematical formulation of WD is provided below:

$$WD_i^w = \sum_{j \in N(i)} w_{ij} \tag{10}$$

where: WD_i^w represents the weighted degree of node i, w_{ij} denotes the number of joint patent applications between node i and node j.

Betweenness centrality (BC) measures a node's brokerage potential in network connectivity, formally defined as the fraction of shortest paths passing through it relative to all possible shortest paths in the network. Formally expressed as:

$$BC(i) = \sum_{s \neq i \neq t} \frac{\sigma_{ij}(i)}{\sigma_{ij}} \tag{11}$$

where: σ_{st} is the total count of shortest paths between nodes i and j; $\sigma_{st}(i)$ is the count of shortest paths through node i.

4.2.4. Control variables

Building on previous studies, this study controls for a range of potential confounding factors. First, regarding population and economic conditions, we include population size and GDP per capita to capture market size and development disparities, thereby accounting for potential city-scale effects. Second, regarding innovation inputs and the policy environment, we introduce variables such as R&D intensity, human capital, and innovation policies to capture resource endowment and institutional support. Third, to account for potential digitalization effects on knowledge complexity, we incorporate indicators of digital infrastructure, including measures of the digital economy and information infrastructure. In addition, structural variables, including foreign direct investment and financial development, are included to represent openness and financial capacity. Finally, as prior studies have shown that other transportation modes, such as highways and air transport, significantly affect knowledge diffusion and innovation outcomes (Agrawal et al., 2017; Bai et al., 2024; Wang et al., 2018), we further control for highway and air transportation to mitigate potential confounding effects. The detailed calculation methods for these variables are reported in Table 1.

5. Results analysis

5.1. Descriptive statistics and correlation analysis

Table 2 presents the summary statistics for key variables. The knowledge complexity index (KCI) exhibits substantial cross-city variation, ranging from 0.01 to 47.6, reflecting pronounced regional disparities in knowledge complexity. The dichotomous treatment indicator HSRDID exhibits sufficient variation across both cross-sectional and temporal dimensions, satisfying the common support requirement for causal identification. The correlation matrix presented in Fig. 3 demonstrates that all pairwise correlation coefficients remain below the conventional threshold of 0.8. This initial assessment is further corroborated by variance inflation factor (VIF) analysis, which yields a mean VIF of 2.42, substantially below the commonly accepted threshold of 5.0. The maximum individual VIF of 4.14 provides additional confirmation that multicollinearity does not pose a significant concern for our regression specifications.

Table 2
Summary statistics for model variables.

Variable	Obs	Mean	Std. Dev.	Min	Max
KCI	4288	3.07	5.17	0.01	47.60
HSRDID	4288	0.37	0.48	0.00	1.00
pop	4288	423.07	264.93	18.20	2094.70
pergdp	4288	41922.75	31666.92	2730.00	256877.00
rdi	4288	0.01	0.02	0.00	0.21
hc	4288	8.95	0.81	6.12	11.54
inpolicy	4288	1.24	1.99	0.00	21.00
fin	4288	2.21	1.10	0.51	21.30
fdi	4288	0.02	0.02	0.00	0.20
infoinfa	4288	71.49	88.84	0.00	931.10
digital	4288	0.42	1.07	0.00	18.57
motor	4288	7052.10	10590.95	12.00	195597.00
air	4288	1932959.00	6145060.00	0.00	73400000.00
SPE	4288	8.89	12.39	1.28	277.90
RV	4288	4.73	2.50	0.00	13.61
UV	4288	3.47	0.53	0.00	4.14
WD	4288	415.22	1362.67	0.00	20818.00
BC	4288	66.27	232.78	0.00	3316.78

Table 1
Definitions and calculation methods of control variables.

Variable Name	Abbreviation	Measurement	Reference
Population size	pop	Total number of permanent residents	(Broekel et al., 2023; Tang et al., 2021)
Economic development level	pergdp	GDP per capita	(Broekel et al., 2023; Gao and Zheng, 2020; Tang et al., 2021)
R&D intensity	rdi	Share of government science expenditure in total fiscal budget	(Broekel et al., 2023; Hou, 2022; Tang et al., 2021)
Human capital	hc	Average years of schooling (interpolated where census data missing)	Rodríguez-Pose et al. (2021)
Innovation policy	inpolicy	Number of National High-Tech Industrial Development Zones	/
Digital economy	digital	Number of newly registered digital-related firms per capita	/
Information infrastructure	infoinfa	Number of broadband internet interfaces per 100 residents	Wan et al. (2024)
Foreign direct investment	fdi	Ratio of actually utilized foreign capital to GDP	Ning et al. (2016)
Financial development	fin	Credit intermediation ratio (deposits + loans of financial institutions/GDP)	Li and Li (2022)
Highway transport	motor	Highway passenger traffic	(Agrawal et al., 2017; Wang et al., 2018)
Aviation transport	air	Civil aviation passenger turnover	Bai et al. (2024)

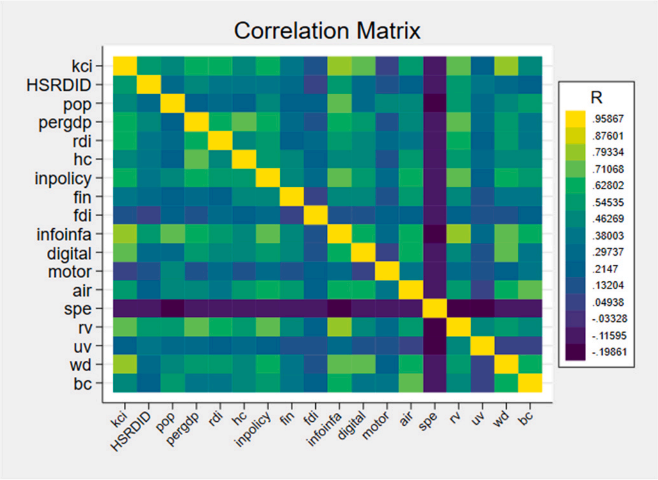


Fig. 3. Correlation matrix of variables.

5.2. *Baseline regression result*

Table 3 presents the DID estimates of HSR construction on knowledge complexity. Column (1) shows the baseline model including only city and year fixed effects, while columns (2) to (4) sequentially add controls for population size, economic development, innovation inputs, digital infrastructure, and alternative transportation modes to mitigate potential omitted variable bias. Column (5) further introduces province–year fixed effects to account for unobserved shocks at the provincial level. All models cluster robust standard errors at the city level. The coefficient of HSRDID is positive and statistically significant across all specifications. In the fully specified model (column 5), the coefficient is 0.081 and significant at the 1 % level, indicating that, after controlling for multiple confounding factors, cities with HSR exhibit knowledge complexity approximately 8.1 % higher than non-HSR cities. These results provide strong empirical support for Hypothesis 1, suggesting that HSR serves as a catalyst for enhancing knowledge recombination capacity. Extending prior research emphasizing HSR's role in stimulating innovation and knowledge spillovers (Gao and Zheng, 2020; Hou, 2022; Miwa et al., 2022; Sun et al., 2023), our findings demonstrate that HSR not only increases the volume of knowledge exchange but also structurally reshapes regional innovation systems by reorganizing knowledge elements.

5.3. *Endogeneity treatment*

5.3.1. *Dealing with reverse causality*

There may exist a reverse causal relationship between HSR implementation and knowledge complexity, as cities with higher knowledge complexity are potentially more likely to receive HSR connections. To address this endogeneity, we implement an instrumental variable (IV) strategy leveraging historical transportation infrastructure, following established practice in transport economics (Baum-Snow et al., 2017; Zheng and Kahn, 2013; Zhu, 2021). Specifically, we construct two historical IVs: $IV\_1931rail$, a dummy indicating whether a city had a railway station in 1931, and IV_Post, the number of post stations in a city during the Ming Dynasty. Since both IVs are inherently cross-sectional and thus would be absorbed by fixed effects in a panel setting, we follow Duflo and Pande (2007) and introduce temporal variation by interacting each historical IVs with a full set of year variables. Formally, the time-varying IVs are defined as: $IV\_1931railnum = IV\_1931rail \times year$ and $IV\_postnum = IV\_Post \times year$. This multiplicative specification

Table 3
Baseline regression results of HSR on knowledge complexity.

	(1)	(2)	(3)	(4)	(5)
	lnKCI	lnKCI	lnKCI	lnKCI	lnKCI
HSRDID	0.186***	0.109***	0.103***	0.104***	0.081***
	(0.025)	(0.019)	(0.018)	(0.018)	(0.016)
lnpop		0.917***	0.592***	0.584***	0.445***
		(0.149)	(0.126)	(0.126)	(0.115)
lngdp		0.055	0.076	0.090*	0.054
		(0.059)	(0.049)	(0.048)	(0.050)
lnrdi		0.142***	0.125***	0.125***	0.054***
		(0.014)	(0.013)	(0.013)	(0.010)
lnhc		1.403	1.182	1.192	0.858
		(0.871)	(0.794)	(0.782)	(0.616)
lninpolicy		0.218***	0.160***	0.162***	0.118***
		(0.032)	(0.028)	(0.028)	(0.022)
lnfin		0.036	0.085*	0.092*	0.033
		(0.058)	(0.050)	(0.049)	(0.041)
lnfdi		−1.299*	−0.392	−0.302	−0.885
		(0.759)	(0.671)	(0.678)	(0.613)
lninfoinfa			0.002***	0.002***	0.001***
			(0.000)	(0.000)	(0.000)
lndigital			0.180***	0.174***	0.346***
			(0.054)	(0.054)	(0.074)
lnmotor				−0.029**	−0.005
				(0.014)	(0.014)
lnair				−0.001	0.000
				(0.002)	(0.002)
Year fixed effects	YES	YES	YES	YES	NO
City fixed effects	YES	YES	YES	YES	YES
Province-Year fixed effects	NO	NO	NO	NO	YES
Constant	0.902***	−7.527***	−5.612***	−5.488***	−4.044***
	(0.009)	(2.095)	(1.907)	(1.889)	(1.540)
Observations	4288	4288	4288	4288	4272
r2_a	0.923	0.949	0.956	0.956	0.972

Note: Robust standard errors (in parentheses) are clustered at the city level. *, **, and *** denote statistical significance at the 10 %, 5 %, and 1 % levels, respectively. The same notation applies to the subsequent regression tables and is therefore not repeated.

preserves the exogenous historical component of the IVs while generating the time variation required for panel estimation, thereby ensuring relevance without compromising the exclusion restriction.

Subsequently, we re-estimate model (1) using a two-stage least squares (2SLS) approach, with the results presented in Table 4. Columns (2) and (6) report the first-stage estimates, showing that both *IV_1931railnum* and *IV_postnum* are positively and significantly associated with HSRDID. This indicates a strong correlation between the locations of modern HSR stations and the layout of historical transportation infrastructure. Regarding the validity of the IVs, both first-stage regressions show that the IVs are strongly correlated with HSR implementation, with F-statistics of 109.57 and 280.11, respectively, well above the conventional threshold of 10. Columns (4) and (8) report the second-stage estimation results, where the coefficients of HSRDID remain positive and statistically significant, and are substantially larger than the baseline results. These results indicate that, after addressing potential reverse causality, HSR continues to exert a significant positive effect on knowledge complexity, further corroborating the robustness of our findings.

5.3.2. *Dealing with sample selection bias*

The spatial distribution of HSR stations across Chinese cities reflects deliberate planning shaped by geographic, economic, and administrative considerations. This non-random placement raises potential endogeneity concerns, as station locations may correlate with underlying urban characteristics. To address potential sample selection bias, we implement a propensity score matching (PSM) strategy, estimating each city's likelihood of receiving treatment and matching treated and control cities to simulate a counterfactual scenario approximating random assignment. In the baseline specification, we adopt one-to-one nearest-neighbor matching with replacement to balance covariates and ensure comparability. As shown in Table 5, standardized biases across most covariates fall below 20 % after matching, with bias reduction exceeding 60 % and mean differences becoming statistically insignificant. These results indicate that PSM substantially enhances comparability, ensuring the robustness and validity of the subsequent DID estimations.

Table 6 presents the results of re-estimating model (1) using the matched sample. Columns (1) and (2) show that the coefficients of HSRDID are positive and statistically significant, consistent with the results of the baseline regression. To ensure robustness and rule out sensitivity to the matching method, we conduct additional checks using three alternative PSM techniques: one-to-one nearest-neighbor matching with replacement and a 0.01 caliper (columns 3–4), kernel matching (columns 5–6), and radius caliper matching (columns 7–8). Across all methods, the coefficient of HSRDID remains positive and significant. These findings confirm that, after controlling for potential sample selection bias, the positive effect of HSR on knowledge complexity is robust, further reinforcing the credibility of our conclusions.

5.4. *Robustness test*

5.4.1. *Parallel trend test*

The DID framework's validity crucially depends on satisfying the parallel trends assumption, which requires comparable pre-treatment trajectories between treatment and control groups. Following the methodology of Beck et al. (2010), we address potential pre-existing trends through mean-differencing of regression coefficients and confidence intervals across all pre-treatment periods. Fig. 4 presents the results of the parallel trends test. Before the introduction of HSR, the coefficients fluctuate around zero without a clear trend, confirming that treatment and control regions followed similar trajectories in knowledge complexity. Following the opening of HSR, the coefficients rise significantly and continuously, indicating that HSR expansion has a sustained and cumulative positive impact on regional knowledge complexity over time.

5.4.2. *Placebo test*

To validate the credibility of our results, we conduct a placebo test in line with the method proposed by Ferrara et al. (2012). By generating 300 randomized pseudo-treatment assignments and re-estimating Model (1), we assess whether the observed effects persist under artificial conditions. Fig. 5 reports the results of the placebo test. The estimated coefficients obtained from random reassignment of treatment years are symmetrically distributed around zero, and most of the simulated effects are statistically insignificant, as indicated by the high p-values. This suggests that the observed positive impact of HSR on knowledge complexity is unlikely to be driven by random chance, thereby confirming the robustness and causal validity of the baseline results.

5.4.3. *Alternative measures of dependent variable*

In the baseline regression models, knowledge complexity is measured using the element combination method. To further assess the robustness of our findings, we also employ the mapping approach proposed by Balland and Rigby (2017) to construct an alternative knowledge complexity index (KCIMORT). The theoretical rationale of this approach is rooted in product space theory, which posits that a technology exhibits higher knowledge complexity if its spatial distribution is concentrated and produced by only a few regions, indicating greater production difficulty. The detailed computation procedure is provided in Appendix B. To verify the robustness of our findings, we compare the statistical properties of KCI and KCIMORT. As shown in Appendix C, the two indicators exhibit distinct distributional characteristics and a relatively low Pearson correlation (r = 0.059), suggesting that they capture complementary dimensions of knowledge complexity. Columns (1) and (2) of Table 7 present the estimation results based on the alternative dependent variable. The coefficient of HSRDID remains positive and statistically significant. This finding demonstrates that the positive effect of HSR on knowledge complexity is robust and insensitive to the choice

Table 4
Results of IV (2SLS) estimation.

	First-stage regression		Second-stage regression		First-stage regression		Second-stage regression	
	(1) HSRDID	(2) HSRDID	(3) lnKCI	(4) lnKCI	(5) HSRDID	(6) HSRDID	(7) lnKCI	(8) lnKCI
HSRDID			0.205*** (0.045)	0.147*** (0.028)			0.458*** (0.080)	0.256*** (0.048)
IV_postnum	0.092*** (0.009)	0.093*** (0.009)						
IV_1931railnum					0.522*** (0.029)	0.522*** (0.031)		
Control variables	NO	YES	NO	YES	NO	YES	NO	YES
City fixed effects	YES	YES	YES	YES	YES	YES	YES	YES
Province-Year fixed effects	YES	YES	YES	YES	YES	YES	YES	YES
Observations	4272	4272	4272	4272	4272	4272	4272	4272
F test	106.88	109.57			321.06	280.11		

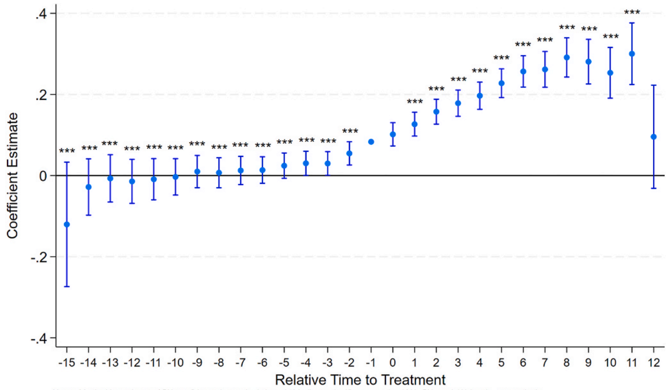
Table 5
Results of the balance test for PSM.

Variable	Matched	Treated Mean	Control Mean	%bias	%reduct|bias|	*t*-test t	p>|t|	V(T)/V(C)
lnpop	U	5.9577	5.2931	105.7		26.58	0.000	0.90[a]
	M	5.8131	5.7733	6.3	94.0	2.67	0.008	1.10[a]
lnpergdp	U	10.414	10.229	24.4		6.19	0.000	0.84[a]
	M	10.26	10.161	13.0	46.5	5.41	0.000	1.02
lnrdi	U	−4.6202	−5.1877	65.9		15.07	0.000	1.68[a]
	M	−4.8318	−4.9984	19.3	70.6	7.88	0.000	1.67[a]
lnhc	U	2.1923	2.168	27.3		6.79	0.000	0.96
	M	2.1718	2.15	24.5	10.3	11.24	0.000	1.38[a]
lninpolicy	U	0.64004	0.24887	73.0		15.77	0.000	2.76[a]
	M	0.43711	0.28492	28.4	61.1	12.78	0.000	1.54[a]
lnfin	U	0.71371	0.61842	25.0		5.82	0.000	1.45[a]
	M	0.62727	0.58541	11.0	56.1	4.49	0.000	1.50[a]
lnfdi	U	0.01941	0.00779	79.0		16.76	0.000	3.33[a]
	M	0.01392	0.01953	−38.2	51.6	−12.46	0.000	0.40[a]
lninfoinfa	U	79.984	30.481	69.7		14.07	0.000	7.26[a]
	M	52.457	56.76	−6.1	91.3	−2.78	0.005	0.63[a]
lndigital	U	0.2713	0.15309	38.2		8.08	0.000	3.42[a]
	M	0.17813	0.15446	7.6	80.0	4.07	0.000	1.91[a]
lnmotor	U	8.503	7.6345	88.3		22.52	0.000	0.83[a]
	M	8.2983	8.0989	20.3	77.0	9.38	0.000	1.27[a]
lnair	U	6.5844	4.9625	24.8		5.84	0.000	1.35[a]
	M	5.1271	3.468	25.4	−2.3	10.24	0.000	1.30[a]

[a] If variance ratio outside [0.94; 1.07] for U and [0.86; 1.17] for M.

Table 6
DID estimation results using different PSM matching methods.

	1:1 Nearest-Neighbor Matching		1:1 Nearest-Neighbor Matching (Caliper)		Kernel Matching		Radius Caliper Matching	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	lnKCI	lnKCI	lnKCI	lnKCI	lnKCI	lnKCI	lnKCI	lnKCI
HSRDID	0.075***	0.040***	0.069***	0.036**	0.074***	0.039***	0.071***	0.037***
	(0.017)	(0.014)	(0.017)	(0.014)	(0.017)	(0.014)	(0.017)	(0.014)
Control variables	NO	YES	NO	YES	NO	YES	NO	YES
City fixed effects	YES	YES	YES	YES	YES	YES	YES	YES
Province-Year fixed effects	YES	YES	YES	YES	YES	YES	YES	YES
Constant	0.790***	−0.380	0.794***	−0.094	0.792***	−0.240	0.793***	−0.155
	(0.005)	(1.168)	(0.005)	(1.188)	(0.005)	(1.180)	(0.005)	(1.182)
Observations	3473	3473	3426	3426	3462	3462	3453	3453
r2_a	0.962	0.970	0.962	0.970	0.962	0.970	0.962	0.970



Notes: Vertical bars denote 95% confidence intervals. *, *, and ** indicate significance at the 10%, 5%, and 1% levels, respectively

Fig. 4. Parallel trend test plot.

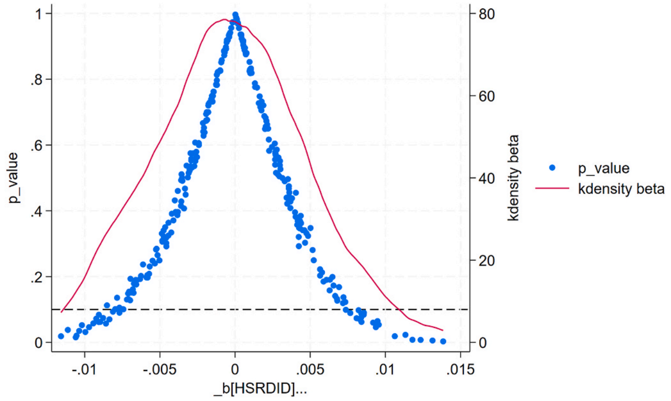


Fig. 5. Placebo test results.

of complexity indicator, providing further confidence in the validity of our empirical conclusions.

5.4.4. *Alternative measures of independent variable*

To further ensure the robustness of our results, we re-estimate the model by adopting an alternative measure of the key explanatory variable. Specifically, two more fine-grained indicators are introduced to replace the binary HSR treatment variable, providing a more accurate characterization of cities' connectivity and centrality within the HSR network. The first, the HSR centrality index (HSRcentrality), is constructed using intercity train timetable data from 2013 to 2020. This index accounts for both the number and directionality of HSR services between cities, offering a comprehensive representation of each city's hub position within China's national HSR network. Detailed

Table 7
Robustness test results using alternative variables.

	(1)	(2)	(3)	(4)	(5)	(6)
	lnKCIMORT	lnKCIMORT	lnKCI	lnKCI	lnKCI	lnKCI
HSRDID	0.040*	0.037*				
	(0.022)	(0.022)				
lnHSRcentrality			0.016***	0.020***		
			(0.005)	(0.004)		
lnHSRconnectivity					0.023**	0.022**
					(0.011)	(0.011)
Control variables	NO	YES	NO	YES	NO	YES
City fixed effects	YES	YES	YES	YES	YES	YES
Province-Year fixed effects	YES	YES	YES	YES	YES	YES
Constant	3.338***	2.857	1.508***	−0.245	1.513***	0.718
	(0.008)	(2.769)	(0.010)	(2.061)	(0.015)	(2.775)
Observations	4272	4272	2136	2144	2136	2136
r2_a	0.732	0.734	0.971	0.959	0.970	0.972

computational procedures are provided in Appendix D. The second indicator, the HSR connectivity index (HSRconnectivity), measures the total number of HSR routes directly linked to each city's stations, capturing the breadth of its network connections. Data are drawn from a database compiled by Davis et al. (2025). Columns (3)–(8) of Table 7 present the estimation results. Both lnHSRcentrality and lnHSRconnectivity yield positive and statistically significant coefficients, consistent with the baseline findings, confirming that HSR network connectivity and centrality significantly enhance knowledge complexity.

5.5. Mechanism analysis

5.5.1. Agglomeration externalities mechanism

Columns (1)–(4) of Table 8 present the results using specialized agglomeration as the mediating variable. As shown in column (2), the coefficient of HSRDID is 0.720 but statistically insignificant, suggesting that the expansion of HSR has not significantly reshaped specialized agglomeration patterns. This result contrasts with previous studies arguing that HSR development promotes industrial specialization (Chen and Guo, 2023; Hou, 2022; Miwa et al., 2022; Sun et al., 2023). The inconsistency may imply that HSR primarily facilitates innovation by enhancing intercity knowledge exchange rather than by strengthening industry-specific knowledge bases. When lnSPE is introduced into the model (column 4), its coefficient becomes significantly negative, while the coefficient of HSRDID declines from 0.081 in the baseline regression to 0.077. This outcome indicates that specialized agglomeration does not serve as an effective channel through which HSR enhances knowledge complexity. Moreover, the insignificant association between lnSPE and knowledge complexity suggests that specialization does not exert a clear influence on complex knowledge recombination. This finding contradicts Balland et al. (2022), who emphasize the positive contribution of specialization to economic complexity, but aligns with Martin and Sunley (2022), who contend that excessive specialization can induce path dependence and constrain regional innovation capacity. Overall, the results do not support the hypothesized pathway “HSR → specialized agglomeration → knowledge complexity,” and hypothesis H2a is therefore not supported.

Columns (5)–(8) of Table 8 report the results using related variety as the mediating variable. As shown in column (6), the coefficient on HSRDID is positive but not statistically significant, indicating that the expansion of the HSR network has only a limited impact on promoting diversification. In column (8), the coefficient of lnRV is positive and significant. This further implies that interrelated knowledge can be more readily absorbed, recombined, and transformed into new knowledge, consistent with the theoretical perspectives of related variety (Balland et al., 2019; Frenken et al., 2007). Although column (8) shows that the coefficient of HSRDID decreases from 0.081 to 0.075 while lnRV remains significantly positive, this reduction cannot be interpreted as evidence of mediation because the essential first-stage relationship, namely the effect of HSR on lnRV, is not established. Therefore, within the mediation analysis framework adopted in this study, the proposed pathway “HSR → related variety → knowledge complexity” is not empirically supported, and hypothesis H2b is consequently not validated.

Columns (9)–(12) of Table 8 present the results with unrelated variety as the mediating variable. Specifically, in column (10), the coefficient of HSRDID is significantly negative, indicating that the introduction of HSR reduces unrelated variety. Furthermore, in column (12), when lnUV is included in the model, its coefficient remains significantly negative, while the coefficient of HSRDID decreases from 0.081 to 0.074 but remains statistically significant. This indicates that unrelated variety partially mediates the effect of HSR on knowledge complexity. The negative coefficient of lnUV suggests that unrelated variety suppresses knowledge recombination, thereby constraining increases in knowledge complexity. This finding contrasts with prior literature suggesting that unrelated variety may provide reservoirs for breakthrough innovation (Castaldi et al., 2016). However, due to the large inherent technological distance between unrelated fields, the coordination and integration costs often outweigh potential benefits, impeding complex knowledge formation. Overall, the evidence supports the pathway “HSR → reduction in unrelated variety → knowledge complexity” (see Fig. 6), suggesting that HSR indirectly mitigates the negative influence of unrelated variety on knowledge complexity by reducing internal unrelated variety.

5.5.2. Network externalities mechanism

Table 9 presents the results of the mediation analysis with network externalities. Specifically, in column (2), the coefficient of HSRDID is not significant, whereas in column (6), it is significantly positive. A potential explanation is that HSR development significantly increases a node's direct innovation connections (degree centrality) but has no significant effect on its strategic position as an intermediary or bridge within the network (betweenness centrality). These results suggest that the introduction of HSR enhances a city's connectivity within broader innovation networks. Furthermore, in columns (4) and (8), when lnBC and lnWD are included in the model, both coefficients are significantly positive, while the coefficient of HSRDID decreases from 0.081 to 0.076 and 0.075 but remains statistically significant. These results suggest that network externalities partially mediate the effect of HSR on knowledge complexity. The positive and significant coefficients of lnWD indicate that network centrality positively influences knowledge complexity, consistent with network externality theory (Burger and Meijers, 2016; Capello, 2000; Katz and Shapiro, 1985). In short, enhancing network centrality strengthens a city's intermediary role within innovation networks and improves its capacity for knowledge diffusion, intercity collaboration, and technological recombination. Overall, the evidence

Table 8
Mediation effects of agglomeration externalities.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
	lnSPE	lnSPE	lnKCI	lnKCI	lnRV	lnRV	lnKCI	lnKCI	lnUV	lnUV	lnKCI	lnKCI
HSRDID	0.932**	0.720	0.119***	0.077***	0.060*	0.010	0.113***	0.075***	−0.052**	−0.026*	0.108***	0.074***
	(0.436)	(0.446)	(0.020)	(0.016)	(0.034)	(0.054)	(0.017)	(0.013)	(0.025)	(0.015)	(0.017)	(0.014)
lnSPE			0.001**	0.000								
			(0.001)	(0.000)								
lnRV							0.161***	0.150***				
							(0.011)	(0.007)				
lnUV											−0.240***	−0.146***
											(0.022)	(0.014)
Control variables	NO	YES	NO	YES	NO	YES	NO	YES	NO	YES	NO	YES
City fixed effects	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES
Province-Year fixed effects	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES
Constant	8.137***	19.719	0.918***	−4.527***	−9.733***	−16.025***	0.170***	−2.122	3.493***	−0.744	1.765***	−4.230***
	(0.163)	(43.416)	(0.009)	(1.550)	(2.250)	(4.216)	(0.054)	(1.363)	(0.009)	(1.016)	(0.074)	(1.403)
Observations	4272	4272	4272	4272	4272	4272	4272	4272	4272	4272	4272	4272
r2_a	0.340	0.342	0.957	0.972	0.947	0.949	0.970	0.983	0.707	0.735	0.963	0.974

supports the pathway “HSR → increased network centrality → enhanced knowledge complexity” (see Fig. 6), demonstrating that HSR expansion not only improves physical accessibility but also reinforces the connectivity and integration of urban innovation networks. Consequently, hypothesis H3 is confirmed.

5.5.3. Interaction effects between agglomeration externalities and network externalities

To further uncover how HSR influences knowledge complexity through the interaction between agglomeration externalities and network externalities, this study introduces an interaction term between the two mediating dimensions within the mediation analysis framework. The regression results of the interaction effects are reported in Table 10. Furthermore, to provide a more intuitive illustration of the substitutive and complementary relationships between agglomeration and network externalities, Fig. 7 presents the corresponding conceptual framework.

In columns (2) and (8) of Table 10, the interaction terms lnSPE × lnBC and lnSPE × lnWD are both significantly negative, indicating a substitutive relationship between specialized agglomeration and network externalities. According to agglomeration economics, specialized agglomeration fosters innovation through shared labor markets, supply-demand matching, and knowledge spillovers among local actors (Marshall, 1890). However, the expansion of HSR substantially improves accessibility, enabling cities to embed more deeply within cross-regional innovation networks. As innovation actors engage in these networks, the marginal benefits of local specialization decline, since knowledge flows are no longer confined to local agglomeration but can also leverage intercity knowledge pipelines. This finding aligns with Burger and Meijers (2016), who suggest that network externalities across geographic boundaries can partially substitute for local agglomeration economies. It also resonates with the “borrowed scale” and “networked city-region” frameworks (Capello, 2000; Meijers et al., 2016), highlighting that even in the absence of strong local agglomeration economics, cities can access external knowledge and economic gains via strategic network positions.

In columns (4) and (10) of Table 10, the interaction terms lnRV × lnBC and lnRV × lnWD are positive and statistically significant, indicating a complementary relationship between related variety and network externalities. Mechanistically, related variety enhances innovation by increasing technological proximity and avoiding path dependency (Balland et al., 2019; Frenken et al., 2007). With HSR expansion, intercity spatial linkages are strengthened, enabling firms to access knowledge across a broader innovation network and absorb non-local complementary resources, thereby boosting knowledge complexity. Additionally, in columns (6) and (12) of Table 10, the interaction terms lnUV × lnBC and lnUV × lnWD are also positive and significant, suggesting a complementary effect between unrelated variety and network externalities. The underlying mechanism is that HSR facilitates external interactions and cross-domain collaboration, creating richer channels for knowledge flows. This partially mitigates the learning and integration challenges posed by unrelated variety. This finding aligns with the “local buzz and global pipelines” framework (Bathelt, 2007; Wang et al., 2025), indicating that intensive local knowledge interactions, combined with knowledge flows through external network connections, jointly foster innovation.

5.5.4. Knowledge complexity carriers: technology collaboration and transfer

To further unpack the micro-level mechanisms through which HSR affects knowledge complexity, this study examines the process from the perspective of interactions among innovation actors. Specifically, we focus on four micro-level channels: firm-to-firm technology collaboration (FFTC), research-to-firm technology collaboration (RFTC), firm-to-firm technology transfer (FFTT), and research-to-firm technology transfer (RFTT). Detailed measurement procedures are provided in Appendix E. Based on these indicators; we employ a mediation model to

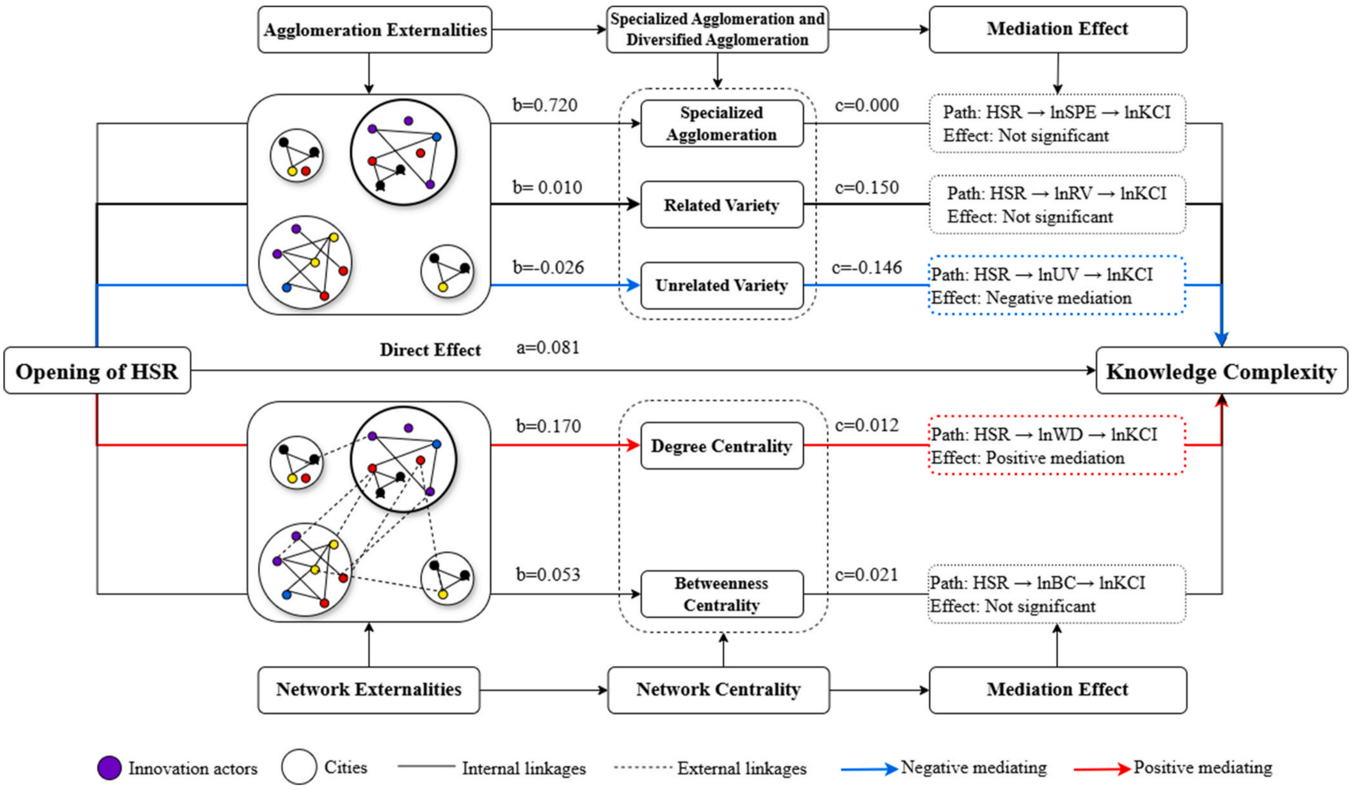


Fig. 6. Conceptual diagram of agglomeration externalities and network externalities pathway.

Table 9
Mediation effects of network externalities.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	lnBC	lnBC	lnKCI	lnKCI	lnWD	lnWD	lnKCI	lnKCI
HSRDID	0.093	0.053	0.118***	0.076***	0.172**	0.170**	0.118***	0.075***
	(0.060)	(0.058)	(0.020)	(0.015)	(0.070)	(0.070)	(0.020)	(0.016)
lnBC			0.028***	0.021***				
			(0.005)	(0.003)				
lnWD							0.013**	0.012**
							(0.006)	(0.005)
Control variables	NO	YES	NO	YES	NO	YES	NO	YES
Province-Year fixed effects	YES	YES	YES	YES	YES	YES	YES	YES
City fixed effects	YES	YES	YES	YES	YES	YES	YES	YES
Constant	1.570***	−4.158	0.884***	−4.435***	3.613***	−14.881**	0.882***	−4.348***
	(0.023)	(5.112)	(0.011)	(1.513)	(0.026)	(6.102)	(0.023)	(1.532)
Observations	4272	4272	4272	4272	4272	4272	4272	4272
r2_a	0.812	0.814	0.958	0.973	0.902	0.903	0.957	0.972

empirically test the role of these channels in driving HSR-induced knowledge complexity. The results are reported in Tables 11 and 12, respectively.

Table 11 reports the mediating effects of FFTC and RFTC. In column (2), the coefficient of HSRDID is positive and significant, indicating that HSR promotes inter-firm technological collaboration. In column (4), when lnFFTC is included, both HSRDID and lnFFTC remain significant, with the coefficient of HSRDID decreasing from 0.081 in the baseline regression to 0.078, suggesting that part of HSR's effect on knowledge complexity operates through FFTC. Similarly, in column (6), the coefficient of HSRDID is significantly positive, indicating that HSR enhances industry-university-research collaboration. In column (8), with lnRFTC included, the coefficient of HSRDID decreases from 0.081 to 0.077, while lnRFTC remains significantly positive. This further supports the partial mediation mechanism, suggesting that HSR enhances knowledge complexity primarily by strengthening collaboration among universities, industries, and research institutions. Overall, these results indicate that HSR fosters both horizontal (inter-firm) and vertical (industry–research) technological collaboration, accelerating knowledge recombination. This mechanism can be conceptualized as a “knowledge-sharing mechanism,” whereby the collaboration networks facilitated by HSR further incentivize innovation actors to share knowledge within the innovation network. The pathway can be summarized as follows: HSR expansion → collaboration among innovation actors → formation of collaboration networks → enhancement of knowledge complexity.

Table 12 presents the results for FFTT and RFTT as mediating channels. In column (2), the coefficient of HSRDID is significantly positive, demonstrating that HSR promotes inter-firm technology transfer. In column (4), with lnFFTT included, lnFFTT is significantly positive, and the coefficient of HSRDID decreases from 0.081 to 0.076 but remains significant, indicating a partial mediating effect of inter-firm

Table 10
Interaction effects between agglomeration and network externalities.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
	lnKCI	lnKCI	lnKCI	lnKCI	lnKCI	lnKCI	lnKCI	lnKCI	lnKCI	lnKCI	lnKCI	lnKCI
HSRDID	0.118***	0.080***	0.058***	0.056***	0.103***	0.076***	0.114***	0.078***	0.034***	0.033***	0.094***	0.068***
	(0.019)	(0.015)	(0.010)	(0.010)	(0.009)	(0.008)	(0.019)	(0.015)	(0.008)	(0.008)	(0.017)	(0.014)
lnSPE	0.003***	0.001***					0.006***	0.003***				
	(0.001)	(0.000)					(0.001)	(0.001)				
lnBC	0.043***	0.030***	−0.131***	−0.091***	−0.059***	−0.016						
	(0.006)	(0.004)	(0.008)	(0.008)	(0.022)	(0.019)						
lnSPE × lnBC	−0.002***	−0.001***										
	(0.001)	(0.000)										
lnRV			0.114***	0.120***					0.055***	0.058***		
			(0.008)	(0.006)					(0.008)	(0.006)		
lnRV × lnBC			0.031***	0.022***								
			(0.001)	(0.002)								
lnUV					−0.246***	−0.149***					−0.286***	−0.183***
					(0.010)	(0.009)					(0.024)	(0.016)
lnUV × lnBC					0.023***	0.010*						
					(0.006)	(0.005)						
lnWD							0.032***	0.022***	−0.102***	−0.097***	−0.092***	−0.067***
							(0.006)	(0.005)	(0.005)	(0.005)	(0.025)	(0.023)
lnSPE × lnWD							−0.002***	−0.001***				
							(0.000)	(0.000)				
lnRV × lnWD									0.027***	0.025***		
									(0.001)	(0.001)		
lnUV × lnWD											0.030***	0.022***
											(0.007)	(0.007)
Control variables	NO	YES	NO	YES	NO	YES	NO	YES	NO	YES	NO	YES
Province-Year fixed effects	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES
City fixed effects	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES
Constant	0.856***	−4.037***	0.277***	−0.589	1.749***	−3.676***	0.816***	−4.006***	0.472***	0.113	1.871***	−3.362**
	(0.012)	(1.486)	(0.039)	(1.021)	(0.036)	(0.613)	(0.024)	(1.483)	(0.037)	(0.692)	(0.085)	(1.353)
Observations	4272	4272	4272	4272	4272	4272	4272	4272	4272	4272	4272	4272
r2_a	0.959	0.973	0.985	0.987	0.964	0.975	0.959	0.973	0.991	0.991	0.964	0.974

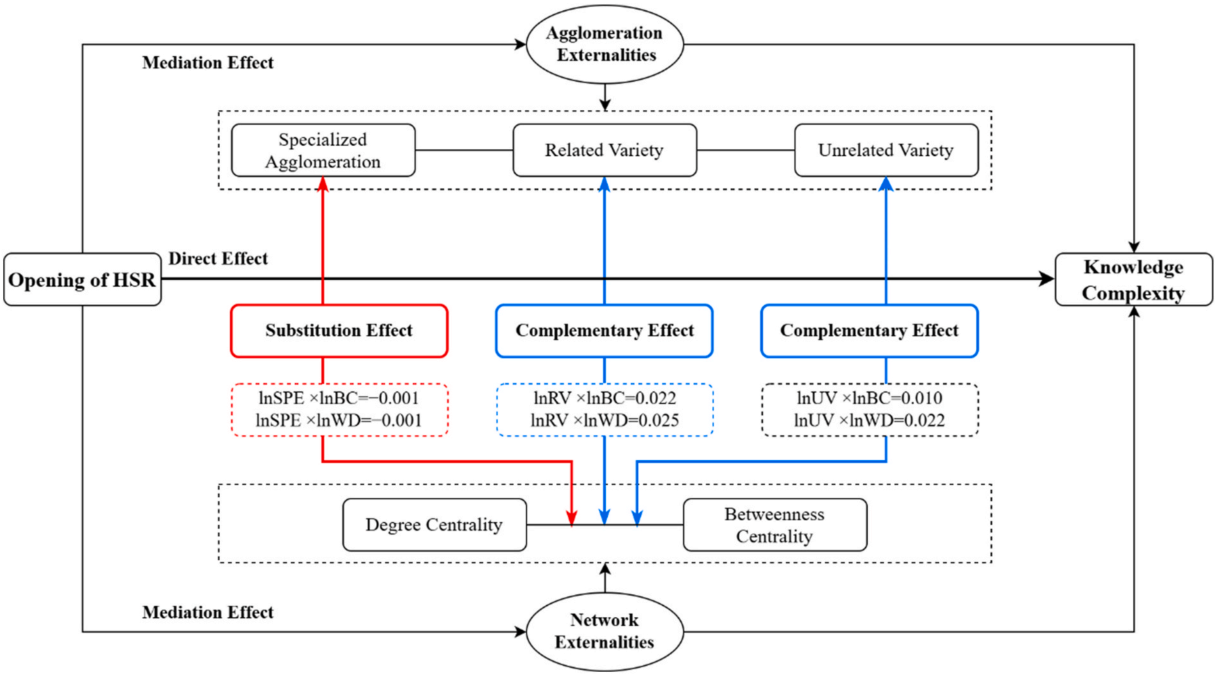


Fig. 7. Interaction between agglomeration externalities and network externalities.

Table 11
Effects of technology collaboration as mediating channels.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	lnFFTC	lnFFTC	lnKCI	lnKCI	lnRFTC	lnRFTC	lnKCI	lnKCI
HSRDID	0.237*** (0.077)	0.167** (0.072)	0.111*** (0.019)	0.078*** (0.016)	0.127** (0.052)	0.086* (0.049)	0.111*** (0.019)	0.077*** (0.015)
lnFFTC			0.041*** (0.007)	0.017*** (0.005)				
lnRFTC							0.072*** (0.009)	0.042*** (0.006)
Control variables	NO	YES	NO	YES	NO	YES	NO	YES
Province-Year fixed effects	YES	YES	YES	YES	YES	YES	YES	YES
City fixed effects	YES	YES	YES	YES	YES	YES	YES	YES
Constant	2.464*** (0.029)	−21.717*** (5.456)	0.826*** (0.020)	−3.669** (1.495)	1.808*** (0.019)	−14.204*** (4.219)	0.796*** (0.020)	−3.452** (1.448)
Observations	4272	4272	4272	4272	4272	4272	4272	4272
r2_a	0.854	0.862	0.958	0.972	0.859	0.865	0.960	0.973

technology transfer. Similarly, in column (6), the coefficient of HSRDID is positive and statistically significant, indicating that HSR development facilitates technology transfer from research institutions to firms. In column (8), after incorporating lnRFTT, its coefficient remains positive and statistically significant, while the direct effect of HSRDID decreases yet retains statistical significance. This provides further empirical

Table 12
Effects of technology transfer as mediating channels.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	lnFFTT	lnFFTT	lnKCI	lnKCI	lnRFTT	lnRFTT	lnKCI	lnKCI
HSRDID	0.233*** (0.056)	0.183*** (0.054)	0.111*** (0.019)	0.076*** (0.015)	0.324*** (0.062)	0.226*** (0.057)	0.094*** (0.017)	0.072*** (0.015)
lnFFTT			0.041*** (0.006)	0.022*** (0.004)				
lnRFTT							0.080*** (0.008)	0.038*** (0.005)
Control variables	NO	YES	NO	YES	NO	YES	NO	YES
Province-Year fixed effects	YES	YES	YES	YES	YES	YES	YES	YES
City fixed effects	YES	YES	YES	YES	YES	YES	YES	YES
Constant	2.453*** (0.021)	−11.710*** (4.084)	0.827*** (0.017)	−3.781** (1.515)	1.180*** (0.023)	−7.070* (4.267)	0.833*** (0.014)	−3.777*** (1.444)
Observations	4272	4272	4272	4272	4272	4272	4272	4272
r2_a	0.883	0.887	0.958	0.972	0.795	0.819	0.962	0.973

support for the mediating mechanism, suggesting that technology transfer from research institutions to firms constitutes an important pathway through which HSR enhances knowledge complexity. Collectively, these findings suggest that HSR indirectly promotes knowledge recombination through technology transfer among innovation actors. Mechanistically, this process can be described as a "knowledge-matching mechanism": HSR expansion reduces spatial frictions among innovation actors, enabling them to more efficiently identify, exchange, and integrate complementary technological resources across broader markets. The pathway can be summarized as: HSR expansion → technology transfer among innovation actors → formation of technology transfer networks → increased knowledge complexity.

5.6. *Heterogeneity discussion*

To examine whether the effect of HSR openings on knowledge complexity varies across cities of different sizes, we classify cities based on the population criteria of the Seventh National Population Census (2021). Cities with a population exceeding one million across its constituent districts are categorized as large cities, while the remainder are classified as small and medium-sized cities. The results are reported in columns (1)–(4) of Table 13. The coefficients of HSRDID are positive and statistically significant at the 1 % level in both groups. It is noteworthy that the estimated coefficient for medium and small cities ($\beta = 0.032$) is smaller than that for large cities ($\beta = 0.084$). Moreover, the Fisher permutation test ($p = 0.005$) indicates that this difference is statistically significant, suggesting heterogeneous effects of HSR on knowledge complexity across cities of different sizes. The underlying mechanism likely lies in the fact that large cities possess stronger foundational advantages in innovation-related factors, enabling them to more effectively translate the knowledge spillovers generated by HSR into increases in complexity. By contrast, the relatively weaker innovation systems and absorptive capacities of medium and small cities dampen the marginal impact of comparable improvements in accessibility.

To further identify the regional heterogeneity of HSR effects, this study classifies the sample into eastern, central and western regions according to China's administrative divisions and conducts separate subgroup regressions. The results are reported in columns (5)–(8) of Table 13. Specifically, the coefficients of HSRDID are significantly positive across all regions. Moreover, the coefficient for the central and western regions ($\beta = 0.089$) is notably higher than that for the eastern region ($\beta = 0.051$), and the Fisher permutation test ($P = 0.035$) confirms that this difference is statistically significant. This finding suggests that HSR exerts a significant latecomer advantage amplification effect on knowledge complexity. The underlying mechanism is that HSR greatly enhances accessibility in central and western regions, thereby reducing the cost of knowledge exchange with eastern innovation hubs. This cost reduction facilitates cross-regional knowledge spillovers and technology diffusion, which in turn provides critical support for local innovators to engage in efficient knowledge recombination. These findings align with prior research emphasizing HSR's role in fostering inclusive regional innovation and reducing intercity knowledge disparities (Hou, 2022; Tang et al., 2021).

To investigate whether the impact of HSR on knowledge complexity differs across cities with varying administrative hierarchies, this study divides the sample into two groups, provincial capital cities and non-provincial capital cities, and conducts separate regressions for each. The results, reported in columns (1)–(4) of Table 14, show that the coefficients of HSRDID are positive and significant in both groups. Although the coefficient is slightly higher for non-provincial capitals, the Fisher permutation test ($p = 0.375$) indicates that the difference is not statistically significant, suggesting that HSR exerts a similar positive influence on knowledge complexity regardless of administrative rank. This finding enriches existing research emphasizing the "political bias" in China's innovation system. Prior studies argue that provincial capital cities typically benefit from preferential access to innovation resources such as R&D funding, policy support, and human capital (Andersson et al., 2014; Yao et al., 2020; Yuan and Han, 2021). However, the results of this study suggest that the expansion of the HSR network has mitigated such disparities by improving intercity connectivity and the mobility of innovation factors, thereby enhancing knowledge spillovers and reducing dependence on administratively allocated resources.

To investigate whether the impact of HSR on knowledge complexity varies according to cities' initial innovation capacity, this study categorizes cities into two subgroups based on the median number of invention patent applications: high-innovation-capacity cities and low-innovation-capacity cities. Separate regressions are conducted for each subgroup, with results presented in columns (5)–(8) of Table 14. The results indicate that the coefficients of HSRDID are positive and statistically significant across both groups. Although the coefficient of HSRDID for high-innovation-capacity cities is slightly larger than that for low-innovation-capacity cities, the Fisher permutation test ($p = 0.119$) shows that this difference is not statistically significant. This suggests that the positive effect of HSR on knowledge complexity is generally consistent, regardless of a city's baseline innovation capacity. The underlying mechanism lies in the universal nature of the knowledge diffusion effects generated by transportation infrastructure. Specifically, HSR not only reinforces the knowledge recombination capacity of established innovation centers but also provides cities with weaker innovation foundations greater opportunities to access external knowledge networks.

6. Conclusion and discussion

This study leverages the expansion of China's HSR network as a quasi-natural experiment to examine its causal impact on urban knowledge complexity, employing a DID approach combined with IV regression. Our findings robustly indicate that HSR development

Table 13
Results of heterogeneity analysis by subgroups.

	City Scale Heterogeneity				Regional Heterogeneity Analysis			
	Large cities		Small and medium-sized cities		Eastern region		Central and western regions	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	lnKCI	lnKCI	lnKCI	lnKCI	lnKCI	lnKCI	lnKCI	lnKCI
HSRDID	0.114***	0.084***	0.150***	0.032**	0.102***	0.051**	0.132***	0.089***
	(0.036)	(0.028)	(0.029)	(0.016)	(0.027)	(0.022)	(0.028)	(0.021)
Control variables	NO	YES	NO	YES	NO	YES	NO	YES
Province-Year fixed effects	YES	YES	YES	YES	YES	YES	YES	YES
City fixed effects	YES	YES	YES	YES	YES	YES	YES	YES
Constant	1.293***	−6.539	0.728***	−1.741	1.115***	−2.474	0.807***	−4.128***
	(0.018)	(3.949)	(0.009)	(1.412)	(0.012)	(3.168)	(0.009)	(1.577)
Observations	1376	1376	2800	2768	1680	1680	2592	2592
r2_a	0.969	0.980	0.910	0.969	0.964	0.977	0.947	0.968

Table 14
Results of heterogeneity analysis by subgroups.

	Administrative Hierarchy Heterogeneity				Innovation Capability Heterogeneity			
	Provincial capital cities		Non-provincial capital cities		High innovation capacity		Low innovation capacity	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	lnKCI	lnKCI	lnKCI	lnKCI	lnKCI	lnKCI	lnKCI	lnKCI
HSRDID	0.052	0.067*	0.178***	0.086***	0.060***	0.060***	0.043**	0.035**
	(0.041)	(0.038)	(0.026)	(0.019)	(0.022)	(0.020)	(0.018)	(0.016)
Control variables	NO	YES	NO	YES	NO	YES	NO	YES
Province-Year fixed effects	NO	NO	NO	NO	YES	YES	YES	YES
Year fixed effects	YES	YES	YES	YES	NO	NO	NO	NO
City fixed effects	YES	YES	YES	YES	YES	YES	YES	YES
Constant	1.560***	−1.224	0.846***	−5.089**	1.258***	−7.848***	0.637***	−0.866
	(0.024)	(5.497)	(0.009)	(1.968)	(0.011)	(2.300)	(0.005)	(1.264)
Observations	400	400	3888	3888	2007	2007	2084	2084
r2_a	0.975	0.983	0.917	0.951	0.975	0.981	0.962	0.966

significantly enhances urban knowledge complexity, with results remaining consistent across multiple sensitivity and robustness tests. The heterogeneity analysis shows that the positive effect of HSR on knowledge complexity is particularly pronounced in large cities as well as in central and western regions. Mechanistically, we identify two pathways: from agglomeration externalities, HSR primarily influences knowledge complexity through unrelated variety rather than specialization; from network externalities, HSR-driven intercity connectivity fosters the formation of innovation networks, promoting knowledge recombination. The interaction between network externalities and diversified agglomeration externalities produces complementary effects, enhancing the combinatorial potential of regional knowledge portfolios, whereas network externalities and specialized agglomeration exhibit substitution effects. At the micro level, technological collaboration and transfer between firms and research institutions serve as key channels for these mechanisms, indicating that innovation actors can leverage HSR to facilitate knowledge sharing, matching, and recombination.

This study advances the dialogue between transport economics, economic geography, and innovation management, offering insights across multiple dimensions. First, it develops a conceptual framework that integrates HSR, knowledge complexity, network externalities, and agglomeration externalities to explain how HSR reshapes knowledge complexity, complementing prior research that emphasizes local agglomeration effects (Garrison and Souleyrette II, 1996; Hou, 2022; Miwa et al., 2022; Sun et al., 2023). Second, the study provides new empirical evidence for debates on agglomeration economies and network externalities, revealing both substitutive effects between specialized agglomeration and network externalities, supporting earlier findings (Burger and Meijers, 2016; Huang et al., 2020; Wang et al., 2025), and complementary relationships between diversified agglomeration and network externalities, deepening understanding of their nuanced interactions. Finally, this study uncovers the micro-mechanisms through which HSR influences knowledge complexity, illustrating how different innovation actors leverage the HSR network to share and match knowledge. From the perspective of actors' behaviors and interactions, this finding provides strong micro-level evidence supporting the view that transportation infrastructure facilitates knowledge collaboration and search (Berger and Prawitz, 2024; Tan and Pan, 2024).

The findings of this study yield significant policy implications that extend beyond China's context, offering valuable insights for other nations contemplating HSR investments. For China, the study underscores the transformative potential of HSR networks in advancing knowledge complexity, with key policy recommendations emerging: First, strategic optimization of HSR network planning should be prioritized to maximize positive externalities. Emphasis should be placed on expanding HSR connectivity to smaller urban centers and less-developed regions (e.g., central and western provinces), thereby enhancing access to innovative resources across geographical divides. Second, policymakers should actively facilitate the formation of regional innovation consortia among HSR-connected cities. Such alliances, potentially piloted in major urban agglomerations like the Yangtze River Delta, Pearl River Delta and Chengdu-Chongqing economic circle, could institutionalize mechanisms for shared innovation resources, cooperative R&D initiatives, and technology diffusion.

China's experience offers insights that can be extended to other parts of the world, particularly emerging economies grappling with regional development imbalances. The study demonstrates that HSR has the potential to bridge innovation divides by integrating peripheral regions with core knowledge hubs. Strategically incorporating HSR into national development frameworks can yield multiple benefits: stimulating lagging regional economies, cultivating area-specific knowledge specializations, and nurturing cross-regional innovation ecosystems. A universal policy principle emerges from these findings: HSR planning should consciously support dual objectives of enhanced human mobility and systematic knowledge circulation. This can be achieved through two complementary approaches: first, by establishing formal structures for intercity innovation cooperation along HSR corridors; second, by ensuring that physical infrastructure designs actively facilitate both efficient HSR and seamless knowledge exchange between connected urban centers.

While this study elucidates the macro-level relationship between HSR development and urban knowledge complexity, several limitations merit consideration to guide future scholarship. A primary constraint stems from the absence of micro-level analysis, particularly regarding the specific channels through which HSR facilitates knowledge creation and diffusion among innovation actors. Subsequent investigations could employ hierarchical modeling or longitudinal firm-level data to disentangle these underlying mechanisms. Furthermore, the current analysis does not account for potential synergies or substitution effects between HSR and other transportation infrastructures, such as highway or air networks. Comparative studies across different national contexts could yield valuable insights into how institutional environments moderate HSR's innovation effects. Such refined examinations would substantially advance our theoretical understanding of transportation infrastructure's role in knowledge ecosystem development.

CRediT authorship contribution statement

Mingming Guan: Writing – original draft, Visualization, Methodology, Data curation, Conceptualization. **Yuting Hou:** Writing – review & editing, Supervision, Conceptualization.

Declaration of generative AI and AI-assisted technologies in the writing process

During the preparation of this work, the authors used ChatGPT/ Grammarly solely for language editing and readability improvement of certain sections. All substantive content, research design, analysis, and conclusions were entirely generated by the authors. The authors carefully reviewed and edited all AI-assisted portions and take full responsibility for the final content of the publication.

Acknowledgements

This research received no specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Appendix A. Supplementary material

Supplementary material to this article can be found online at https://doi.org/10.1016/j.tranpol.2026.104036.

Data availability

Data will be made available on request.

References

Acs, Z.J., Anselin, L., Varga, A., 2002. Patents and innovation counts as measures of regional production of new knowledge. Res. Pol. 31, 1069–1085. https://doi.org/10.1016/S0048-7333(01)00184-6.

Agrawal, A., Galasso, A., Oettl, A., 2017. Roads and innovation. Rev. Econ. Stat. 99, 417–434.

Alonso, W., 1975. Urban zero population growth. In: The No-Growth Society. Routledge.

Andersson, D.E., Gunessee, S., Matthiessen, C.W., Find, S., 2014. The geography of Chinese science. Environ. Plann. 46, 2950–2971. https://doi.org/10.1068/a130283p.

Bai, J., Jianqiu, Jin, W., Zhou, S., 2024. Proximity and knowledge spillovers: evidence from the introduction of new airline routes. Manag. Sci. 70, 7464–7485. https://doi.org/10.1287/mnsc.2021.01717.

Balland, P.-A., Boschma, R., Crespo, J., Rigby, D.L., 2019. Smart specialization policy in the european union: relatedness, knowledge complexity and regional diversification. Reg. Stud. 53, 1252–1268. https://doi.org/10.1080/00343404.2018.1437900.

Balland, P.-A., Broekel, T., Diodato, D., Giuliani, E., Hausmann, R., O'Clery, N., Rigby, D., 2022. The new paradigm of economic complexity. Res. Pol. 51, 104450. https://doi.org/10.1016/j.respol.2021.104450.

Balland, P.-A., Jara-Figueroa, C., Petralia, S.G., Steijn, M.P., Rigby, D.L., Hidalgo, C.A., 2020. Complex economic activities concentrate in large cities. Nat. Hum. Behav. 4, 248–254.

Balland, P.-A., Rigby, D., 2017. The geography of complex knowledge. Econ. Geogr. 93, 1–23. https://doi.org/10.1080/00130095.2016.1205947.

Baron, R.M., Kenny, D.A., 1986. The moderator–mediator variable distinction in social psychological research: conceptual, strategic, and statistical considerations. J. Personality Soc. Psychol. 51, 1173.

Bathelt, H., 2007. Buzz-and-Pipeline dynamics: towards a knowledge-based multiplier model of clusters. Geography Compass 1, 1282–1298. https://doi.org/10.1111/j.1749-8198.2007.00070.x.

Baum-Snow, N., Brandt, L., Henderson, J.V., Turner, M.A., Zhang, Q., 2017. Roads, railroads, and decentralization of Chinese cities. Rev. Econ. Stat. 99, 435–448.

Beaudry, C., Schiffauerova, A., 2009. Who's right, marshall or jacobs? The localization versus urbanization debate. Res. Pol. 38, 318–337.

Beck, T., Levine, R., Levkov, A., 2010. Big bad banks? The winners and losers from bank deregulation in the United States. J. Finance 65, 1637–1667. https://doi.org/10.1111/j.1540-6261.2010.01589.x.

Berger, T., Prawitz, E., 2024. Collaboration and connectivity: historical evidence from patent records. J. Urban Econ. 139, 103629.

Boschma, R., 2017. Relatedness as driver of regional diversification: a research agenda. Reg. Stud. 51, 351–364. https://doi.org/10.1080/00343404.2016.1254767.

Broekel, T., 2019. Using structural diversity to measure the complexity of technologies. PLoS One 14, e0216856. https://doi.org/10.1371/journal.pone.0216856.

Broekel, T., Knuepling, L., Mewes, L., 2023. Boosting, sorting and complexity—urban scaling of innovation around the world. J. Econ. Geogr. 23, 979–1016. https://doi.org/10.1093/jeg/lbad006.

Burger, M.J., Meijers, E.J., 2016. Agglomerations and the rise of urban network externalities. Pap. Reg. Sci. 95, 5–16. https://doi.org/10.1111/pirs.12223.

Burt, R.S., 1987. Social contagion and innovation: cohesion versus structural equivalence. Am. J. Sociol. 92, 1287–1335. https://doi.org/10.1086/228667.

Capello, R., 2000. The City network paradigm: measuring urban network externalities. Urban Stud. 37, 1925–1945. https://doi.org/10.1080/713707232.

Caragliu, A., De Dominicis, L., De Groot, H.L.F., 2016. Both marshall and jacobs were right. Econ. Geogr. 92, 87–111. https://doi.org/10.1080/00130095.2015.1094371.

Castaldi, C., Frenken, K., Los, B., 2016. Related variety, unrelated variety and technological breakthroughs: an analysis of US state-level patenting. In: Evolutionary Economic Geography. Routledge.

Chan, C.-S., Yuan, J., 2017. Changing travel behaviour of high-speed rail passengers in China. Asia Pac. J. Tourism Res. 22, 1221–1237. https://doi.org/10.1080/10941665.2017.1391303.

Chatzistamoulou, N., Kounetas, K., Tsekouras, K., 2022. Technological hierarchies and learning: spillovers, complexity, relatedness, and the moderating role of absorptive capacity. Technol. Forecast. Soc. Change 183, 121925.

Chen, J., Guo, G., 2023. Diversification or specialization: measuring the impact of high-speed rail connection on technological diversity in China. J. Innov. Knowl. 8, 100306.

Davis, S., Qian, M., Zeng, W., 2025. A comprehensive GIS database for China's Surface transport network with implications for transport and socioeconomics. https://doi.org/10.2139/ssrn.5344880.

Dong, X., Zheng, S., Kahn, M.E., 2020. The role of transportation speed in facilitating high skilled teamwork across cities. J. Urban Econ. 115, 103212.

Duflo, E., Pande, R., 2007. Dams. Q. J. Econ. 122, 601–646.

Feng, Q., Chen, Z., Cheng, C., Chang, H., 2023. Impact of high-speed rail on high-skilled labor mobility in China. Transp. Policy 133, 64–74.

Fernhaber, S.A., Patel, P.C., 2012. How do young firms manage product portfolio complexity? The role of absorptive capacity and ambidexterity. Strateg. Manag. J. 33, 1516–1539. https://doi.org/10.1002/smj.1994.

Ferrara, E.L., Chong, A., Duryea, S., 2012. Soap operas and fertility: evidence from Brazil. Am. Econ. J. Appl. Econ. 4, 1–31. https://doi.org/10.1257/app.4.4.1.

Fleming, L., Sorenson, O., 2001. Technology as a complex adaptive system: evidence from patent data. Res. Pol. 30, 1019–1039. https://doi.org/10.1016/S0048-7333(00)00135-9.

Frenken, K., Van Oort, F., Verburg, T., 2007. Related variety, unrelated variety and regional economic growth. Reg. Stud. 41, 685–697. https://doi.org/10.1080/00343400601120296.

Gao, Y., Zheng, J., 2020. The impact of high-speed rail on innovation: an empirical test of the companion innovation hypothesis of transportation improvement with China's manufacturing firms. World Dev. 127, 104838.

Garrison, W.L., Souleyrette II, R.R., 1996. Transportation, innovation, and development: the companion innovation hypothesis. Logist. Transport Rev. 32, 5.

Griliches, Z., 1998. Patent Statistics as Economic Indicators: a Survey.

Guan, M., Wu, S., Liu, C., 2022. Comparing China's urban aviation and innovation networks. Growth Change 53, 470–486. https://doi.org/10.1111/grow.12593.

Guirao, B., Campa, J.L., Casado-Sanz, N., 2018. Labour mobility between cities and metropolitan integration: the role of high speed rail commuting in Spain. Cities 78, 140–154.

Henke, I., Moyano, A., Pagliara, F., 2023. Influence of high-speed rail on the decentralisation of events from big metropolitan areas to smaller intermediate cities. Soc. Econ. Plann. Sci. 85, 101453.

Heuermann, D.F., Schmieder, J.F., 2019. The effect of infrastructure on worker mobility: evidence from high-speed rail expansion in Germany. J. Econ. Geogr. 19, 335–372.

Hidalgo, C.A., 2021. Economic complexity theory and applications. Nat. Rev. Phys. 3, 92–113. https://doi.org/10.1038/s42254-020-00275-1.

Higham, K., de Rassenfosse, G., Jaffe, A.B., 2021. Patent quality: towards a systematic framework for analysis and measurement. Res. Pol. 50, 104215. https://doi.org/10.1016/j.respol.2021.104215.

Hou, Y., 2022. Agglomeration spillover, accessibility by high-speed rail, and urban innovation in China: a focus on the electronic information industry. Habitat Int. 126, 102618.

Huang, Y., Hong, T., Ma, T., 2020. Urban network externalities, agglomeration economies and urban economic growth. Cities 107, 102882. https://doi.org/10.1016/j.cities.2020.102882.

Ishikura, T., 2025. Estimation of the impact of the new high-speed rail in Japan from a spatial economic perspective. Transp. Policy 171, 214–224. https://doi.org/10.1016/j.tranpol.2025.05.027.

Jacobs, J., 1969. The Economy of Cities. Vintage.

Katz, M.L., Shapiro, C., 1985. Network externalities, competition, and compatibility. Am. Econ. Rev. 75, 424–440.

Koh, Y., Li, J., Xu, J., 2022. Subway, collaborative matching, and innovation. Rev. Econ. Stat. 1–45.

Li, J., Li, B., 2022. Digital inclusive finance and urban innovation: evidence from China. Rev. Dev. Econ. 26, 1010–1034. https://doi.org/10.1111/rode.12846.

Lin, Y., 2017. Travel costs and urban specialization patterns: evidence from China's high speed railway system. J. Urban Econ. 98, 98–123.

Marshall, A., 1890. Principles of Economics, eighth ed. Macmillan, London, p. 3. Published in 1920.

Martin, R., Sunley, P., 2022. Making history matter more in evolutionary economic geography. ZFW – Advances in Economic Geography 66, 65–80. https://doi.org/10.1515/zfw-2022-0014.

Meijers, E.J., Burger, M.J., Hoogerbrugge, M.M., 2016. Borrowing size in networks of cities: city size, network connectivity and metropolitan functions in Europe. Pap. Reg. Sci. 95, 181–199.

Metcalfe, R.M., 1995. Metcalfe's law: a network becomes more valuable as it reaches more users. InfoWorld 17, 53–54.

Mewes, L., Broekel, T., 2022. Technological complexity and economic growth of regions. Res. Pol. 51, 104156.

Miwa, N., Bhatt, A., Morikawa, S., Kato, H., 2022. High-Speed rail and the knowledge economy: evidence from Japan. Transport. Res. Pol. Pract. 159, 398–416. https://doi.org/10.1016/j.tra.2022.01.019.

Ning, L., Guo, R., Chen, K., 2023. Does FDI bring knowledge externalities for host country firms to develop complex technologies? The catalytic role of overseas returnee clustering structures. Res. Pol. 52, 104767. https://doi.org/10.1016/j.respol.2023.104767.

Ning, L., Wang, F., Li, J., 2016. Urban innovation, regional externalities of foreign direct investment and industrial agglomeration: evidence from Chinese cities. Res. Pol. 45, 830–843.

Østergaard, C.R., Timmermans, B., Kristinsson, K., 2011. Does a different view create something new? The effect of employee diversity on innovation. Res. Pol. 40, 500–509.

Pérez-Luño, A., Alegre, J., Valle-Cabrera, R., 2019. The role of tacit knowledge in connecting knowledge exchange and combination with innovation. Technol. Anal. Strat. Manag. 31, 186–198. https://doi.org/10.1080/09537325.2018.1492712.

Polanyi, M., 1962. Tacit knowing: its bearing on some problems of philosophy. Rev. Mod. Phys. 34, 601–616. https://doi.org/10.1103/RevModPhys.34.601.

Rodríguez-Pose, A., Wilkie, C., Zhang, M., 2021. Innovating in “lagging” cities: a comparative exploration of the dynamics of innovation in Chinese cities. Appl. Geogr. 132, 102475. https://doi.org/10.1016/j.apgeog.2021.102475.

Russell, C.J., Chen, C.-L., Wong, C.-Y., 2024. Bridging the gap between high-speed rail transport studies and cluster economics through social knowledge exchange: future research potential. Transp. Rev. 44, 1103–1127. https://doi.org/10.1080/01441647.2024.2366201.

Schumpeter, J.A., 1934. The Theory of Economic Development: an Inquiry into Profits, Capital, Credit, Interest, and the Business Cycle [Translated from the German by R. Opie].

Shaw, S.-L., Fang, Z., Lu, S., Tao, R., 2014. Impacts of high speed rail on railroad network accessibility in China. Journal of Transport Geography, Changing Landscapes of Transport and Logistics in China 40, 112–122. https://doi.org/10.1016/j.jtrangeo.2014.03.010.

Sorenson, O., Rivkin, J.W., Fleming, L., 2006. Complexity, networks and knowledge flow. Res. Pol. 35, 994–1017. https://doi.org/10.1016/j.respol.2006.05.002.

Storper, M., Venables, A.J., 2004. Buzz: face-to-face contact and the urban economy. J. Econ. Geogr. 4, 351–370.

Sun, X., Yan, S., Liu, T., Wang, J., 2023. The impact of high-speed rail on urban economy: synergy with urban agglomeration policy. Transp. Policy 130, 141–154. https://doi.org/10.1016/j.tranpol.2022.11.004.

Tan, R., Pan, L., 2024. How high-speed railway expands beyond local knowledge search in interdisciplinary innovation: evidence from China. Transp. Policy S0967070X24002877. https://doi.org/10.1016/j.tranpol.2024.10.013.

Tang, C., Guan, M., Dou, J., 2021. Understanding the impact of High Speed Railway on urban innovation performance from the perspective of agglomeration externalities and network externalities. Technol. Soc. 67, 101760.

Uzzi, B., Spiro, J., 2005. Collaboration and creativity: the small world problem. Am. J. Sociol. 111, 447–504. https://doi.org/10.1086/432782.

Wan, J., Xie, Q., Fan, X., 2024. The impact of transportation and information infrastructure on urban productivity: evidence from 256 cities in China. Struct. Change Econ. Dynam. 68, 384–392. https://doi.org/10.1016/j.strueco.2023.11.008.

Wang, X., Xie, Z., Zhang, X., Huang, Y., 2018. Roads to innovation: Firm-level evidence from people's Republic of China (PRC). China Econ. Rev. 49, 154–170.

Wang, Y., Wang, G., Chen, G., 2025. Network externalities of the innovation network in China's five urban agglomerations: based on “buzz-and-pipeline” theory. Humanit. Soc. Sci. Commun. 12, 1096. https://doi.org/10.1057/s41599-025-05191-2.

Xiao, T., Makhija, M., Karim, S., 2022. A knowledge recombination perspective of innovation: review and new research directions. J. Manag. 48, 1724–1777. https://doi.org/10.1177/01492063211055982.

Yang, Y., Ma, G., 2023. How can HSR promote inter-city collaborative innovation across regional borders? Cities 138, 104367. https://doi.org/10.1016/j.cities.2023.104367.

Yao, L., Li, J., 2022. Intercity innovation collaboration and the role of high-speed rail connections: evidence from Chinese co-patent data. Reg. Stud. 56, 1845–1857. https://doi.org/10.1080/00343404.2021.2008340.

Yao, L., Li, Jun, Li, Jian, 2020. Urban innovation and intercity patent collaboration: a network analysis of China's national innovation system. Technol. Forecast. Soc. Change 160, 120185.

Yoo, S., Kumagai, J., Kawasaki, K., Hong, S., Zhang, B., Shimamura, T., Managi, S., 2023. Double-edged trains: economic outcomes and regional disparity of high-speed railways. Transp. Policy 133, 120–133. https://doi.org/10.1016/j.tranpol.2023.01.016.

Yuan, Y., Han, Z., 2021. Structural characteristics and proximity comparison of China's urban innovation cooperation network. PLoS One 16, e0255443.

Zhang, Q., Li, Z., Pang, Y., 2025. Can the opening of high-speed rail (HSR) stimulate residents' consumption? –An interpretation based on the clustering effect of innovation factors. Transp. Policy 163, 384–393. https://doi.org/10.1016/j.tranpol.2025.01.015.

Zheng, S., Kahn, M.E., 2013. China's bullet trains facilitate market integration and mitigate the cost of megacity growth. Proc. Natl. Acad. Sci. U.S.A 110. https://doi.org/10.1073/pnas.1209247110.

Zhu, P., 2021. Does high-speed rail stimulate urban land growth? Experience from China. Transport. Res. Transport Environ. 98, 102974.