Research Policy 55 (2026) 105422

Contents lists available at ScienceDirect

Research Policy

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

The software complexity of nations

Sándor Juhász[a,b,c,d], Johannes Wachs[b,c,d,*], Jermain Kaminski[e], César A. Hidalgo[f,g,h]

[a] Corvinus Institute for Advanced Studies, Corvinus University of Budapest
[b] Institute for Data Analytics and Information Systems, Corvinus University of Budapest, Hungary
[c] Centre for Economic and Regional Studies, ELTE, Budapest, Hungary
[d] Complexity Science Hub, Vienna, Austria
[e] School of Business and Economics, Maastricht University, Netherlands
[f] Center for Collective Learning, IAST, Toulouse School of Economics, University of Toulouse Capitole, France
[g] Center for Collective Learning, CIAS, Corvinus University of Budapest, Hungary
[h] Alliance Manchester Business School, University of Manchester, United Kingdom

ARTICLE INFO

Keywords:
Economic complexity
Open source software
Innovation
GitHub
Relatedness

ABSTRACT

Despite the growing importance of the digital sector, research on economic complexity and its implications continues to rely mostly on administrative records—e.g. data on exports, patents, and employment—that have blind spots when it comes to the digital economy. In this paper we use data on the geography of programming languages used in open-source software to extend economic complexity ideas to the digital economy. We estimate a country's software economic complexity index ($ECI^{software}$) and show that it complements the ability of measures of complexity based on trade, patents, and research to account for international differences in GDP per capita, income inequality, and emissions. We also show that open-source software follows the principle of relatedness, meaning that a country's entries and exits in programming languages are partly explained by its current pattern of specialization. Together, these findings help extend economic complexity ideas and their policy implications to the digital economy.

1. Introduction

The study of economic complexity has predominantly relied on administrative records, such as international trade data (Hidalgo et al., 2007; Hidalgo and Hausmann, 2009), patent filings (Balland and Rigby, 2017; Kogler et al., 2013), and employment statistics (Jara-Figueroa et al., 2018; Neffke and Henning, 2013), that while valuable, struggle to capture the importance of the digital economy. This “dark matter” (Greenstein and Nagle, 2014) is important because software capabilities—which are human capital intensive—represent a mobile and transmissible source of economic complexity that is relevant for policy efforts focused on increasing the complexity of economies (Hidalgo, 2023). Yet, despite this evident need, internationally comparable estimates of software-related economic complexity remain limited.

Economic complexity refers to the structure and breadth of productive capabilities embedded or implicit in an economy's industries, products, or workforce (Hidalgo and Hausmann, 2009; Hausmann et al., 2014; Hidalgo, 2021). Methodologically, its modeled using two key concepts: the *economic complexity index (ECI)* and the idea of *relatedness*.

The economic complexity index (ECI) provides a mean to estimate the combined presence of an economy's capabilities without having to define them (Hidalgo and Stojkoski, 2025). It is often used to anticipate macroeconomic outcomes, such as long-term economic growth (Hidalgo and Hausmann, 2009, Domini, 2022, Chávez et al., 2017, Stojkoski et al., 2023a, 2023b), since economies endowed with diverse capabilities can recombine them into more complex and higher value added products (Hidalgo and Stojkoski, 2025). Relatedness asserts that regions and countries diversify into new activities when these share capabilities with those that an economy is currently specialized in (Hidalgo et al., 2007; Neffke et al., 2011; Neffke and Henning, 2013; Hausmann et al., 2014; Hidalgo et al., 2018; Hidalgo, 2021; Balland et al., 2022). For instance, a country with expertise in data analytics and high-performance computing is more likely to expand into fields that build upon that foundation, such as artificial intelligence, than countries lacking these complementary specializations.

While economic complexity methods have expanded to include trade, patents, employment, and research publication data, their application to the digital sector remains limited. Software capabilities are

* Corresponding author at: Institute for Data Analytics and Information Systems, Corvinus University of Budapest, Hungary.
E-mail address: johannes.wachs@uni-corvinus.hu (J. Wachs).

https://doi.org/10.1016/j.respol.2026.105422
Received 28 August 2024; Received in revised form 30 December 2025; Accepted 9 January 2026
Available online 22 January 2026

only partially visible in these metrics and digital capabilities are insufficiently expressed in physical product data (Rahmati et al., 2021; Stojkoski et al., 2024). Code crosses borders through cloud services, downloads, and remote platforms rather than through customs, and digital firms often create local subsidiaries that obscure trade flows even further. Moreover, service trade categories remain notoriously broad (including groupings such as "computer and information services"); and patents record protectable inventions rather than the open knowledge embedded in everyday programming.

Yet, these data limitations are at odds with the growing importance of the digital economy and the role played by open-source software (OSS). IT technologies and software development are predictors of firm productivity, innovation capacity, and economic growth (Brynjolfsson and Hitt, 2003, 1998; Rahmati et al., 2021). Within this sector, OSS libraries have become essential building blocks (Eghbal, 2020), with OSS participation predicting higher entrepreneurial activity (Wright et al., 2023) and value-added productivity in ecosystems with complementary capabilities (Nagle, 2019, 2018; Rock, 2019). In the US alone, annual investment in OSS was estimated to be about $38bn in 2019 (Korkmaz et al., 2024), and government subsidies to OSS generate large returns (Gortmaker, 2025). As it is known for complex and innovative activities (Audretsch and Feldman, 1996; Balland et al., 2020), OSS development is human capital-intensive, geographically concentrated (Wachs et al., 2022), and open to international collaboration (Goldbeck, 2025). This suggests software capabilities may follow spatial patterns distinct from traditional complexity metrics.

Taken together, the growing importance of the digital economy, the key role that open-source software plays in it, and the remaining open questions about the geography of software capabilities, represent a critical gap in economic complexity research. Moreover, it remains unclear whether the "complexity" of the digital economy substitutes or complements traditional complexity metrics. In this paper, we address these gaps by exploring the question: *Do economic complexity measures based on the geography of open-source software production correlate with macroeconomic indicators like GDP* per capita*, inequality, and emissions, complementing complexity measures based on trade, research, and patents?*

In this study, we use data on the geographic distribution of OSS projects hosted on GitHub to generate a national-level software economic complexity index ($ECI^{software}$). Our main specification constructs $ECI^{software}$ from clusters of programming languages frequently used together in repositories. The cluster-based measure summarizes the diversity and sophistication of a country's software capabilities in a way that is comparable across countries and aligned with how developers combine technologies in practice. We then link $ECI^{software}$ to GDP per capita, inequality measured through the Gini coefficient, and CO_2-per-GDP from the World Bank and compare its explanatory power with complexity indices based on trade, patents, and research. *Our analyses show that $ECI^{software}$* captures a digital capability dimension that while correlated with trade-, patent- and research-based complexity measures ($R^2 \sim 0.5$–0.6) adds significant explanatory power in cross-country models of GDP per capita and income inequality. In addition, we show that countries' entries and exits in programming languages follows the principle of relatedness, confirming that digital diversification mirrors path-dependence observed in physical industries.

By incorporating software into the complexity toolbox, we provide evidence that digital specialization is reshaping economic structures and creating new pathways for structural transformation. From a policy perspective, the accessibility and granularity of open-source software data offers a cost-effective and reproducible means to track and potentially enhance economic complexity research, providing policymakers a new route to design interventions focused on fostering digital capabilities. Unlike traditional development strategies focused on infrastructure and physical capital, fostering digital complexity relies more on human capital development and knowledge spillovers within software ecosystems (Apostol and Hernández-Rodríguez, 2024; Balland et al., 2022; Brynjolfsson and Saunders, 2010; Korkmaz et al., 2024), and thus, represents a new frontier for applied and fundamental work in economic geography and economic complexity research.

2. Economic complexity and open-source software production

2.1. Complexity, relatedness and the digital sector

Economic complexity involves the use of fine-grained data on activities to capture economic structure and shifts in specialization patterns (Balland et al., 2022; Domini, 2022; Guevara et al., 2016; Hausmann et al., 2014; Hidalgo et al., 2018, 2007; Hidalgo, 2021; Hidalgo and Hausmann, 2009; Hidalgo and Stojkoski, 2025; Poncet and de Waldemar, 2015; Stojkoski et al., 2023b). These structural measures are used to explain variation in macroeconomic outcomes, such as economic growth (Pérez-Balsalobre et al., 2019; Chávez et al., 2017; Domini, 2022; Hausmann et al., 2014; Hidalgo and Hausmann, 2009; Koch, 2021; Poncet and de Waldemar, 2013; Stojkoski et al., 2016, 2023b; Weber et al., 2021), income and gender inequality (Bandeira Morais et al., 2018; Ben Saâd and Assoumou-Ella, 2019; Chu and Hoang, 2020; Hartmann et al., 2017; Lee and Vu, 2019; Sbardella et al., 2017), and emissions (Can and Gozgor, 2017; Doğan et al., 2021; Lapatinas et al., 2019; Mealy and Teytelboym, 2020; Romero and Gramkow, 2021). In the last fifteen years, these methods grew into popular indicators for international and regional development policy (Balland et al., 2022; Hidalgo, 2023, 2021) together with methods designed to explain shifts in specialization, building on the principle of relatedness (Hidalgo et al., 2018): the notion that economies diversify by entering activities that reuse some of their existing capabilities. Relatedness metrics highlight path dependencies and help predict which industries, products, research activities, or technologies are likely to grow or decline in a country, city, or region (Alabdulkareem et al., 2018; Apostol and Hernández-Rodríguez, 2024; Boschma et al., 2013; Guevara et al., 2016; Hidalgo et al., 2018, 2007; Jara-Figueroa et al., 2018; Kogler et al., 2013; Li and Neffke, 2024; Neffke et al., 2011; Neffke and Henning, 2013; Poncet and de Waldemar, 2015). Complexity metrics then provide a comparative estimate of the value of a region's specialization pattern.

But while economic complexity methods enjoy significant adoption in policy and academia, their application is still limited by the availability of fine-grained data. Like the proverbial man looking for his keys under the lamppost, economic complexity efforts thus far have focused on international trade statistics (Hidalgo et al., 2007; Hidalgo and Hausmann, 2009), manufacturing, payroll, firm registry, and employment data for industries (Chávez et al., 2017; Fritz and Manduca, 2021; Gao and Zhou, 2018; Hidalgo, 2021; Jara-Figueroa et al., 2018; Neffke et al., 2011; Neffke and Henning, 2013), data on occupations (Alabdulkareem et al., 2018; Farinha et al., 2019; Jara-Figueroa et al., 2018; Muneepeerakul et al., 2013), patents (Balland and Rigby, 2017; Kogler et al., 2013), and research papers (Chinazzi et al., 2019; Guevara et al., 2016; Stojkoski et al., 2023b). This expansion recently led to the introduction of multidimensional economic complexity (Stojkoski et al., 2023b), the notion that metrics of complexity derived from multiple datasets complement each other to explain macroeconomic outcomes (e.g. trade and patent complexity estimates explain economic growth better together than alone). But with the exception of some recent work on digital trade (Stojkoski et al., 2023a), digital infrastructure (Liang and Tan, 2024), and software components in physical products (Rahmati et al., 2021), the multidimensional expansion of economic complexity is yet to fully reach the digital sector, despite work highlighting the importance of software outside economic complexity research (Shapiro and Varian, 1999; Chattergoon and Kerr, 2022).

For instance Aum and Shin (2024) emphasize the critical role played by software in modern economies, highlighting how it substitutes labor with high elasticity. Branstetter et al. (2019) find that firms, not only technology firms, with greater software intensity measured by patenting activity achieve greater returns to R&D. These results suggest that data on software activity can predict macro level growth. Moreover, the growth of the digital economy and its integration into the offline economy is thought to reduce greenhouse emissions (Liu et al., 2023; Zhang et al., 2024). The impact of digitalization and software production on inequality outcomes is less clear, as unequal access and winner take all dynamics may compound inequality (Arthur, 1994), while growth in access to information and employment opportunities may shrink it (Tian et al., 2025).

In practice the question of how software complexity influences macroeconomic outcomes like growth, inequality and emissions, remains unclear because economic complexity research still suffers from a "digital blind-spot". This is due to the lack of datasets that capture a detailed view of software-related activity (Balland and Rigby, 2017; Chávez et al., 2017; Guevara et al., 2016; Stojkoski et al., 2023b). This gap hinders our ability to apply the insights derived from other datasets to digital industries, making it difficult to—for instance—forecast which digital diversification efforts are more likely to succeed or estimate how software capabilities evolve and cluster over time.

There is in fact some evidence hinting to the notion that data used traditionally to study economic complexity can miss digital capabilities. Economic complexity estimates derived from trade data (Hidalgo and Hausmann, 2009) may not align well with software, which crosses borders through cloud services, downloads, or remote platforms rather than through standard customs channels (Corrado et al., 2005; Stojkoski et al., 2023a). As a result, trade data may systematically underestimate digital activity. Service trade data should be an alternative, but it is notoriously coarse, with categories such as "Computer and Information Services", which are too broad to distinguish basic IT outsourcing from advanced software development. Moreover, software production is often carried out through subsidiaries, obscuring the real geography of capabilities. Furthermore, open-source projects and collaborative code repositories do not appear as discrete tradeable goods (Greenstein and Nagle, 2014; Korkmaz et al., 2024) since many software products are monetized via subscriptions, advertising, or freemium models, making them hard to track in conventional trade records. When it comes to employment statistics, software is also represented through coarse industry categories, such as "Software Publishing," and coarse occupations, such us "Software developers" which provide no information about the programming languages used or the applications created by this segment of the labor force.

In short, it is difficult to describe an economy's digital capabilities using traditional data sources. This limits our understanding of the path-dependent dynamics and sophistication of digital economies. Countries or regions that excel in certain digital fields may not show up clearly in traditional complexity data, undercutting our ability to understand related diversification in their context. More generally, we cannot tell how productive capabilities in this sector relate to important macroeconomic outcomes such as income, growth, inequality and the carbon intensity of economies. Digital or software complexity may complement or substitute classic economic complexity estimates, which are significant predictors of these outcomes. But to understand whether these are complements or substitutes, we need to test these ideas empirically.

2.2. Conceptualizing software complexity

Insofar we have argued that data used to commonly estimate economic complexity fails to capture information about an economy's digital capabilities. But what data can we use to approximate capabilities implicit in the digital economy? Here, we follow a two-pronged approach, building on data on *programming languages* and *software bundles*.

Programming languages provide an unusually fine-grained and consistent trace of digital production. A language is not only a syntax but a technical paradigm formed by an ecosystem of tools, libraries, and conventions that shapes how software is built and maintained (Valverde and Solé, 2015a,b). Language adoption indicates embedded knowledge and skills: familiarity with syntax, common practices, and domain-focused applications (e.g., AI, cybersecurity, or high-performance computing).

Languages are also meaningful categories because their ecosystems exhibit strong social and market dynamics. The value of adopting a language often depends on the availability of complementary assets—libraries, frameworks, documentation, and experienced developers—so technology choices reflect local talent pools and ecosystem maturity rather than purely technical merits (Meyerovich and Rabkin, 2013). These complementarities generate switching costs: the primary barrier to adopting a new language is frequently the surrounding toolchain and library landscape rather than the syntax itself (Shrestha et al., 2022). As a result, language portfolios tend to evolve in path-dependent ways, with organizations moving to technologically proximate ecosystems (e.g., within enterprise stacks or within data science stacks) rather than jumping arbitrarily. For these reasons, programming languages can play a role in software-based comparisons of economies that is analogous to product categories or technology classes in traditional complexity measures: they are observable, comparable across places, and tied to capability accumulation.

Languages, however, are not the natural "activity unit" of software production: most modern software systems rely on bundles of languages that are used together as part of a coherent development stack (e.g., front-end web, data science, low-level systems). Treating each language as an independent activity risks fragmenting what practitioners and firms would recognize as a single capability bundle. To align the measurement unit with how software diversification is typically conceptualized—around software genres, use cases, and ecosystems rather than individual technologies—we aggregate languages into clusters based on their revealed co-use within repositories (Boudreau, 2012; Cennamo and Santaló, 2019). The key idea is that repeated co-use identifies stable bundles of complementary capabilities: languages that are frequently used together tend to be part of the same development stack, and these stacks are closer to the activities whose diversification and sophistication economic complexity methods are designed to capture. in patent-based complexity, patent classes are already higher-level, use-oriented groupings rather than the underlying set of technologies used to produce the patent. Analogously, our co-use clusters summarize software capability bundles rather than individual syntaxes, while still being grounded in observable production choices.

In the empirical analysis, we therefore treat languages as the underlying building blocks and use software bundles as the main unit of observation. We construct these clusters using a project-level dataset of all public GitHub repositories active up to 2024 and the set of programming languages used in each repository. These clusters are interpretable as capability bundles—e.g., a front-end web stack (HTML/CSS/JavaScript), a data science stack (Python/Jupyter Notebook), or low-level systems tooling (C/Assembly/Makefile)—and provide a tractable and stable basis for country-level specialization patterns. We additionally compute versions based on individual languages, theoretically defined language groupings, and GitHub topics; these are used only as robustness checks and reported in the Supplementary Information.

2.3. Scope and contribution

Traditional approaches to economic complexity overlook much of the software sector's intangible and rapidly evolving nature. Programming languages, in particular clusters of languages defined by complementary use, offer a way to fill this gap by reflecting embedded knowledge, illustrating specialized skills, and revealing path-dependent growth patterns.

Specifically, we address economic complexity's digital gap by using data on the country level geographic distribution of programming languages and bundles used in OSS projects to estimate economic complexity for the software sector and explore the principle of relatedness in the context of OSS. This work does not aim to introduce a new method to estimate economic complexity, but simply to apply an existing method to new data and explore the complementarity of these estimates to those derived from well-known data sources (product exports, patents, and research publications). We acknowledge that there has been considerable work exploring alternative mathematical definitions of economic complexity, such as the transformational complexity measure (Natera and Castellacci, 2021), the Log Product Diversity (Inoua, 2023), the Ability index (Bustos and Yıldırım, 2022), and the Fitness complexity (Tacchella et al., 2012). Unlike these contributions, our paper does not involve the introduction of a new mathematical definition but the application of the Hidalgo and Hausmann (2009) definition of economic complexity to open-source software data.

In the next section we present the data and methods used to calculate these indicators and then explore their ability to explain international variance in GDP per capita, income inequality, and emissions that is unaccounted for by measures of complexity based on trade, patents, and research papers. We then construct a network of related open-source software bundles to explore the principle of relatedness in the context of software.

3. Data and the construction of economic complexity measures

We begin by describing the data sources and methods used to construct the country–activity matrices used in the complexity analysis. A key step is that we treat programming languages as an observable building blocks of software production but aggregate them into the software bundles (a.k.a. technology stacks) used in practice. We then apply the standard economic complexity methods to this country–bundle matrix. Finally, we construct a software bundle relatedness network to test the principle of relatedness.

We use data on the geography of open-source software provided by the GitHub Innovation Graph (GHIG).[1] GitHub is the leading platform for OSS development, with over 100 million users worldwide. The dataset presents the number of GitHub users pushing code—uploading local code from a developer's machine to an online repository—by country and programming language on a quarterly basis starting from Q1 2020 and continuing until Q4 2023. GHIG data assigns software contributions to countries based on the IP address of the developer. This data provides a more accurate measure of a location's software activity than sources relying on self-reported locations, which are known to suffer from bias (Hecht et al., 2011). After completing the basic data cleaning procedures explained in section S1 of the Supplementary information, we are left with a sample of 163 countries and 150 programming languages for the period of 2020–2023.

To define the activity categories used in our main $\mathrm{ECI}^{\mathrm{software}}$ specification, we group programming languages into clusters based on their complementary use within repositories. We build these clusters from a separate project-level dataset constructed as follows. First, we identified GitHub repositories that were active in 2024 using GHArchive. Second, for each active repository we queried the GitHub GraphQL API to retrieve its set of programming languages. Repositories typically contain multiple languages; we restrict attention to the set of languages that overlap with the 150 languages retained in the GitHub Innovation Graph (GHIG) sample.

We then construct weighted language occurrence and *co*-occurrence counts in a way that prevents highly polyglot repositories from dominating similarity estimates. For each repository with n distinct in-scope languages, we assign each language a weight of $1/n$, so that the total language weight contributed by a repository adds to 1. For each unordered language pair within the repository, we assign a weight of $2/[n(n-1)]$, so that the total pair weight also adds to 1 for repositories with $n > 1$. Aggregating these weights across repositories yields (i) weighted marginal counts c_l for each language l, and (ii) weighted co-occurrence counts $c_{l\,l'}$ for each pair (l, l'). From these counts we compute cosine similarity between languages. For languages l and l', cosine similarity is defined as:

$$s_{l\,l'} = \frac{c_{l\,l'}}{\sqrt{c_l}\sqrt{c_{l'}}}$$

We convert similarity to distance as: $d_{l\,l'} = 1 - s_{l\,l'}$, and apply hierarchical agglomerative clustering to this distance matrix (linkage as implemented in our code). We obtain our baseline partition by cutting the dendrogram at a distance threshold chosen to yield an interpretable set of clusters (59 in the baseline). Each programming language is assigned to exactly one software bundle or co-use cluster.

Finally, we map GHIG language-level country activity into a country–bundle matrix by summing over languages within each bundle. Let X_{cl} denote the number of developers in country c pushing code in language l (from GHIG). For each cluster k, we define:

$$X_{ck} = \sum_{l \in k} X_{cl}$$

This country–bundle matrix X_{ck} is the main input to our construction of $\mathrm{ECI}^{\mathrm{software}}$ below. In the Supplementary Information (S1, S3, S4), we present three alternative operationalizations of $\mathrm{ECI}^{\mathrm{software}}$, based on individual languages, theoretical clusters of languages derived from the computer science literature, and topics (user tags of project content).

We estimate the Economic Complexity Index (ECI) using the standard technique introduced by (Hidalgo and Hausmann, 2009). Let X_{ck} be a matrix counting the number of developers with an IP in country c pushing code to GitHub in software bundle k. We use X_{ck} to derive the matrix of specialization or revealed comparative advantage R_{ck} as:

$$R_{ck} = \frac{X_{ck} X}{X_c X_k},$$

where omitted indexes have been added over (e.g. $X_c = \sum_k X_{ck}$). We then binarize the matrix R_{ck} to generate the matrix $M_{ck} = 1$ if $R_{ck} \geq 1$ or 0 otherwise. Finally, we let the economic complexity index of a country c (ECI_c) and the software bundle complexity index of an activity k (PCI_k) be defined as the steady state of the map:

$$\mathrm{ECI}_c = \frac{1}{M_c} \sum_k M_{ck} \mathrm{PCI}_k$$

$$\mathrm{PCI}_k = \frac{1}{M_k} \sum_c M_{ck} \mathrm{ECI}_c$$

As is customary, we normalize ECI and PCI values by subtracting their respective mean and dividing them by their standard deviation.

There are several interpretations of ECI. In the context of a supply side production function, it is a method to recover an economy's capabilities from a matrix of geographic specialization (Hidalgo and Stojkoski, 2025). ECI is also a spectral-clustering method that identifies whether an economy belongs to the high- or low-capability cluster, by assigning a number to each economy and to each activity that minimizes the distance between the number assigned to each economy and the numbers assigned to each activity (Bottai et al., 2024; Mealy et al., 2019; Servedio et al., 2024). That is, it provides an optimal one-factor split of the specialization matrix. From an intuitive perspective, the capability interpretation of economic complexity simply means that higher complexity economies tend to be endowed with more of the complementary factors of production needed to specialize in activities.

We compare ECI indicators derived from open-source software ($\mathrm{ECI}^{\mathrm{software}}$) with the multidimensional economic complexity data compiled by (Stojkoski et al., 2023b), which uses trade data from the Observatory of Economic Complexity (oec.world), patent data from the World Intellectual Property Organization's International Patent System, and research publication data from SCImago Journal & Country Rank portal. These datasets are described in detail in section S5 of the Supplementary information.

[1] GitHub Innovation Graph https://github.com/github/innovationgraph

We explore the ability of ECI^{software} to complement traditional economic complexity measures in explaining international variation in GDP per capita, income inequality, and emissions. All macroeconomic indicators are derived from the Databank of The World Bank. We use simple cross-sectional Ordinary Least Squares (OLS) models, based on around 90 observations, since the relatively short coverage of the GHIG data (four years) limits our analysis to controlled correlation tests.

We test the principle of relatedness following the approach introduced in the product space (Hidalgo et al., 2007), which starts from the same specialization matrix (*M*) we used to derive measures of economic complexity. Formally, we define the proximity between two software bundles *k* and *k'* as the minimum of the two conditional probabilities that a country specialized in one is also specialized in the other:

$$\phi_{kk'} = \frac{\sum_c M_{ck} M_{ck'}}{\max(M_k, M_{k'})}$$

And define the relatedness between a country *c* and a software bundle *k* as:

$$\omega_{ck} = \frac{\sum_{k'} M_{ck'} \phi_{kk'}}{\phi_k}$$

Where again, missing indices have been added over (e.g. $\phi_k = \sum_{k'} \phi_{kk'}$).

To assess whether countries are more likely to enter software bundles related to their existing portfolio of open-source software specializations, we run linear probability models with country and language-cluster fixed effects. We estimate relatedness using 2020 data and say that a country enters a software bundle if they were not specialized in that software bundle (RCA $<$ 1) in 2020 and 2021 and then gained comparative advantage (RCA $\geq$1) in 2022 and 2023 (e.g. M_{ck} = {0,0,1,1} for the years 2020 to 2023). Our models predict entry as a function of relatedness and software bundle ubiquity.

4. Results

4.1. Software and economic complexity

We begin our analysis by comparing our estimate of economic complexity based on the geography of programming languages clusters (ECI^{software}), with published estimates of economic complexity based on physical product exports (ECI^{trade}), patents (ECI^{technology}), and research publications (ECI^{research}) (Stojkoski et al., 2023b).

Fig. 1A compares four specialization matrices (*M*) where countries are sorted by diversity (number of products, software bundles they specialize in, etc.) and columns are sorted by ubiquity (number of countries specialized in each software bundle, product, etc.). Much like the specialization matrices for trade, patents, and research papers, the *country-software bundle* matrix exhibits a nested structure (Bustos et al., 2012; Mariani et al., 2019), meaning that low diversity economies tend to specialize in a subset of ubiquitous activities found in more diverse economies.

Fig. 1B shows a map of ECI^{software}-based ranking of countries constructed from the country-software bundle matrix and Fig. 1C compares ECI^{software} with the three other ECI measures, showing that the geography of software complexity is different from that expressed in data on products, patents, and research publications. For instance, Russia (RUS), a well-known natural resource exporter with a low ECI^{trade} score (0.112 on a normalized [−1,1] scale), scores much higher in ECI^{software} (0.872

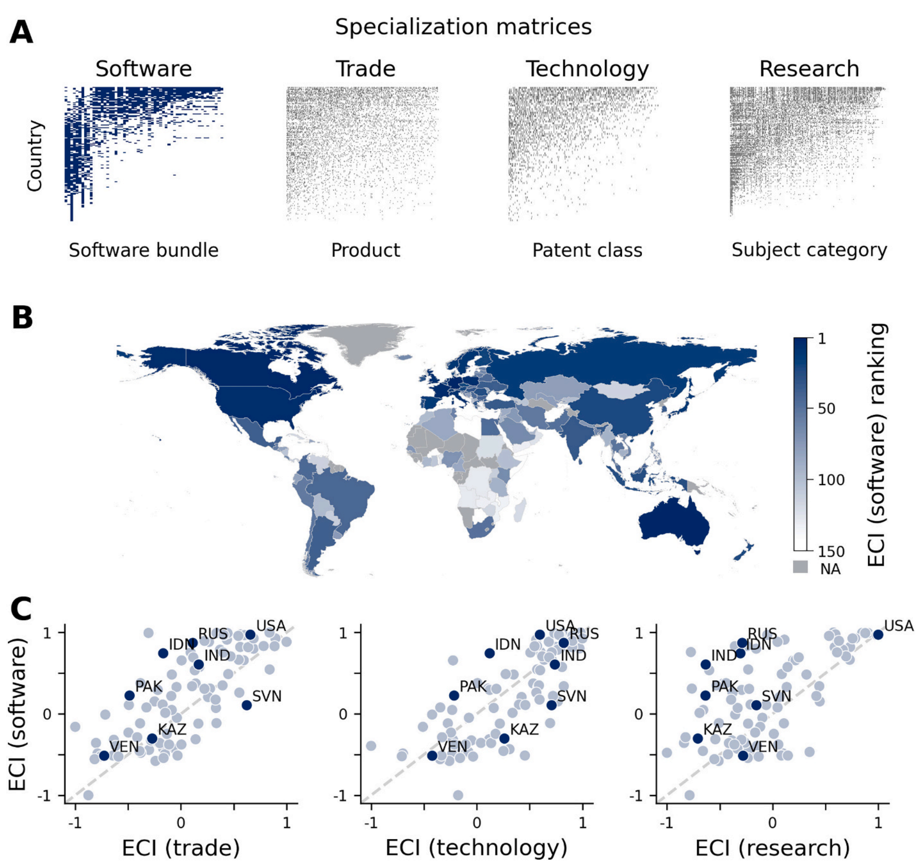

Fig. 1. **A** Specialization matrices for countries and software bundles, products, patents, and research papers. **B** Geographic distribution of software economic complexity (ECI^{software}). **C** Comparison between ECI^{software} and ECI^{trade}, ECI^{technology}, and ECI^{research} respectively (R^2 = 0.576, *p*-value <0.001, R^2 = 0.620, p-value <0.001 and R^2 = 0.346, p-value <0.001). For visualization purposes, ECI values are normalized to a scale of [−1,1]. All ECI measures presented above are calculated using 2020 data only.

on a normalized [−1,1] scale). Similarly, India (IND) scores much higher in $ECI^{software}$ (0.606) than in $ECI^{research}$ (−0.633). The contrast between software and the other dimensions is highlighted by cases such as Indonesia (IDN) and Pakistan (PAK), which rank relatively high in $ECI^{software}$ (0.872 and 0.225) despite scoring much lower in the other ECI measures. Section S6 of the Supplementary information presents a table comparing the values of $ECI^{software}$, ECI^{trade}, $ECI^{technology}$, and $ECI^{research}$ for all countries in our sample.

Next, we explore whether $ECI^{software}$ complements other measures of economic complexity in explaining international variation in GDP per capita, income inequality, and emissions. Descriptive statistics for the key variables are presented in section S7 of the Supplementary information.

Table 1 shows that the correlation between $ECI^{software}$ and GDP per capita remains strong after controlling for other estimates of economic complexity. In fact, $ECI^{software}$ works out to be as good as ECI^{trade} at explaining international variations in GDP per capita in the complete model (column 8). This validates $ECI^{software}$ as a complementary indicator by showing that there is information about international variations in GDP per capita contained in $ECI^{software}$ that is not redundant with the information captured by the other ECIs. Moreover, the robustness of results across different model specifications suggests $ECI^{software}$ is a reliable and consistent predictor. We also note that in this model ECI^{trade} remains statistically significant across specifications, but $ECI^{technology}$ and $ECI^{software}$ lose their significance in the full models, suggesting that the information about international variations in GDP per capita carried by them is redundant with the information available in $ECI^{software}$ and ECI^{trade}.

Economic complexity indicators often show patterns of spatial clustering, as illustrated in Fig. 1A. Moran's I confirms spatial autocorrelation (global Moran's $I = 0.483$, $p < 0.01$), suggesting that countries with similar $ECI^{software}$ values are geographically proximate, deviating significantly from a random distribution (Salinas, 2021). To address potential endogeneity issues and illustrate the robustness of our results, we provide instrumental variable (IV) regressions, following the identification strategy in Stojkoski et al. (2023b). Detailed explanation and all the related regression results can be found in section S8 of the Supplementary information. The IV regressions in models (2) and (10) of Table 1 show results comparable to our baseline estimations.

Next, we look at the ability of $ECI^{software}$ to explain international variations in income inequality (Table 2). Since official data on income inequality are infrequently published, and Gini coefficients vary slowly over time, we use the average Gini coefficient from the 2020–2022 period. Despite the more limited sample, we find the same negative and significant relationship between income inequality and $ECI^{software}$. In fact, $ECI^{software}$ remains strong, negative, and significant across all specifications. We also find $ECI^{research}$ remains significant, albeit with a positive coefficient.

Finally, we look at the intensity of greenhouse gas emissions (emissions per unit of GDP per capita) (Table 3). This is a particularly interesting outcome for $ECI^{software}$ because compared to the physical economy, software and information technologies are expected to be a less carbon-intensive way to generate GDP (Ciuriak and Ptashkina, 2020; Haberl et al., 2020; Hubacek et al., 2021; Romero and Gramkow, 2021; Stojkoski et al., 2023a; Wang and Zhang, 2021; Wiedenhofer et al., 2020).

Our results suggest that software complexity is negatively associated with emissions per unit of GDP in simpler specifications. However, in full models that account for multiple dimensions of complexity, this effect becomes statistically insignificant. This pattern indicates that $ECI^{software}$ and $ECI^{research}$ may share overlapping explanatory power. The variance inflation factor (VIF) analysis (section S14) suggests some degree of collinearity between software and research complexity. While economies with high software complexity tend to have high research complexity (their individual effects on emissions seem to operate through distinct mechanisms, as evidenced by a non-significant interaction term we tested1. 0 separately). One interpretation of these findings is that $ECI^{research}$ absorbs part of the explanatory power of $ECI^{software}$ in predicting emissions, since research-driven economies may be more likely to invest in low-carbon technologies and knowledge-intensive, low-emission industries.

Correlating $ECI^{software}$ with income inequality and emissions intensity allows us to test the Kuznets hypotheses. In section S9 of the Supplementary information, we present regressions including a squared term for GDP per capita. The results support the Kuznets hypothesis for income inequality, indicating an inverted U-shaped relationship, but show little evidence of such a pattern for emissions intensity.

4.2. Related diversification in open-source software

Having validated $ECI^{software}$ as a complementary measure of economic complexity, we now explore whether changes in the software specialization of countries is subject to the principle of relatedness: the notion that economies are more likely to enter—and less likely to exit—related activities (Autant-Bernard, 2001; Guevara et al., 2016; Hidalgo et al., 2018, 2007; Jaffe, 1986; Neffke et al., 2011; Neffke and Henning, 2013).

Table 4 present our linear probability models predicting entry events as a function of relatedness and the ubiquity of a software bundle or language cluster. We also include country and bundle fixed effects and employ clustered standard errors by country to account for within-country correlations over time, ensuring robust and reliable standard errors in our regression models. Estimations based on logit models can be found in section S10 of the Supplementary information.

Table 1
$ECI^{software}$ and GDP per capita (2020) in a multidimensional setting. Robust standard errors in parentheses. Significance codes: *$p < 0.1$, **$p < 0.05$, ***$p < 0.01$.

	GDP per capita (log)									
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
$ECI^{software}$	0.343***	0.358***				0.180***	0.192***	0.338***	0.125***	0.169***
	(0.025)	(0.026)				(0.037)	(0.037)	(0.037)	(0.044)	(0.043)
ECI^{trade}			0.337***			0.222***			0.190***	0.177***
			(0.028)			(0.037)			(0.046)	(0.045)
$ECI^{technology}$				0.266***			0.156***		0.063*	0.051
				(0.021)			(0.029)		(0.035)	(0.036)
$ECI^{research}$					0.140***			0.006	0.022	0.013
					(0.025)			(0.028)	(0.026)	(0.025)
Population (ln)	−0.146***	−0.150***	−0.079***	−0.103***	−0.066***	−0.117***	−0.133***	−0.145***	−0.122***	−0.120***
	(0.017)	(0.017)	(0.015)	(0.019)	(0.020)	(0.014)	(0.017)	(0.019)	(0.016)	(0.016)
Natural resources (ln)	0.015	0.018	0.023*	−0.018	−0.037**	0.034***	0.007	0.015	0.028**	0.031**
	(0.012)	(0.013)	(0.013)	(0.012)	(0.018)	(0.012)	(0.011)	(0.012)	(0.014)	(0.014)
Instrumental variable	No	Yes	No	No	No	No	No	No	No	Yes
Observations	93	93	93	93	93	93	93	93	93	93
R^2	0.648	0.647	0.693	0.654	0.374	0.753	0.711	0.648	0.764	0.762

Table 2
$ECI^{software}$ and income inequality in a multidimensional setting. ECI estimates are based on 2020 data, while the dependent variable is the average Gini coefficient between 2020 and 2022. Robust standard errors in parentheses. Significance codes: *$p < 0.1$, **$p < 0.05$, ***$p < 0.01$.

	Gini coefficient									
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
$ECI^{software}$	−1.038***	−1.054***				−0.905**	−1.033***	−0.981***	−0.920**	−0.966**
	(0.353)	(0.413)				(0.358)	(0.409)	(0.349)	(0.381)	(0.416)
ECI^{trade}			−0.679**			−0.500*			−0.359	−0.354
			(0.289)			(0.275)			(0.293)	(0.294)
$ECI^{technology}$				−0.219			−0.013		0.061	0.069
				(0.253)			(0.288)		(0.285)	(0.281)
$ECI^{research}$					0.419**			0.387**	0.332**	0.331**
					(0.158)			(0.144)	(0.153)	(0.154)
GDP per capita (ln)	0.905***	0.918**	0.612*	0.262	−0.330	1.219***	0.914***	0.521	0.759**	0.787**
	(0.350)	(0.389)	(0.322)	(0.324)	(0.249)	(0.357)	(0.350)	(0.344)	(0.343)	(0.367)
Population (ln)	0.455***	0.460***	0.222**	0.177*	0.090	0.481***	0.456***	0.401***	0.422***	0.435***
	(0.129)	(0.146)	(0.088)	(0.091)	(0.078)	(0.127)	(0.125)	(0.116)	(0.113)	(0.128)
Natural resources (ln)	0.250**	0.248**	0.286**	0.354***	0.400***	0.224*	0.251**	0.313***	0.279**	0.274**
	(0.109)	(0.112)	(0.117)	(0.112)	(0.092)	(0.117)	(0.113)	(0.097)	(0.117)	(0.121)
Instrumental variable	No	Yes	No	No	No	No	No	No	No	Yes
Observations	48	48	48	48	48	48	48	48	48	48
R^2	0.409	0.409	0.357	0.299	0.376	0.445	0.409	0.484	0.499	0.499

Table 3
$ECI^{software}$ and greenhouse gas emission intensity (2020) in a multidimensional setting. Robust standard errors in parentheses. Significance codes: *$p < 0.1$, **$p < 0.05$, ***$p < 0.01$.

	Emission per GDP (log)									
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
$ECI^{software}$	−0.115***	−0.112**				−0.118***	−0.106**	−0.079*	−0.072	−0.059
	(0.041)	(0.043)				(0.043)	(0.047)	(0.044)	(0.050)	(0.052)
ECI^{trade}			−0.021			0.012			0.001	−0.001
			(0.040)			(0.040)			(0.042)	(0.042)
$ECI^{technology}$				−0.052			−0.016		−0.014	−0.017
				(0.033)			(0.038)		(0.039)	(0.039)
$ECI^{research}$					−0.064***			−0.046**	−0.046**	−0.048**
					(0.020)			(0.021)	(0.021)	(0.022)
GDP per capita (ln)	0.011	0.009	−0.051	−0.020	−0.031	0.004	0.019	0.013	0.019	0.016
	(0.027)	(0.028)	(0.032)	(0.030)	(0.024)	(0.034)	(0.029)	(0.026)	(0.034)	(0.034)
Population (ln)	0.031*	0.030	−0.005	0.006	−0.002	0.030	0.032*	0.024	0.025	0.022
	(0.018)	(0.018)	(0.014)	(0.016)	(0.013)	(0.018)	(0.018)	(0.018)	(0.018)	(0.018)
Natural resources (ln)	0.054***	0.055***	0.066***	0.067***	0.062***	0.056***	0.055***	0.053***	0.054***	0.055***
	(0.013)	(0.014)	(0.015)	(0.012)	(0.012)	(0.014)	(0.013)	(0.013)	(0.015)	(0.015)
Instrumental variable	No	Yes	No	No	No	No	No	No	No	Yes
Observations	92	92	92	92	92	92	92	92	92	92
R^2	0.553	0.553	0.506	0.521	0.557	0.553	0.554	0.576	0.577	0.577

Table 4
Entry models on countries gaining revealed comparative advantage (RCA ≥ 1) in software bundles (2020−2023). Standard errors are clustered at the country level. Significance codes: *$p < 0.1$, **$p < 0.05$, ***$p < 0.01$.

	Entry						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Relatedness density	0.154**	0.349**	0.282***	0.429**		0.171**	0.328**
	(0.072)	(0.133)	(0.097)	(0.174)		(0.079)	(0.134)
Ubiquity					−0.006	−0.012	−0.012
					(0.009)	(0.010)	(0.010)
Country FE	No	Yes	No	Yes	No	No	Yes
Software bundle FE	No	No	Yes	Yes	No	No	No
Observations	764	764	764	764	764	764	764
R^2	0.013	0.187	0.118	0.271	0.001	0.016	0.189

Table 4 suggests that open-source software specialization follows the principle of relatedness, with countries being more likely to specialize in software bundles that are related to those they are currently specialized in. The negative and significant effect of bundle ubiquity indicates that countries are less likely to enter common language bundles, which is reasonable since many countries already have comparative advantage in them. While relatedness in the case of OSS behaves similarly across both simpler and more complex models, its explanatory power remains limited, with a baseline R^2 of about 1%. We suggest a few reasons why this is still a significant finding. First, entry is a rare event: we observe 42 entrances vs 722 non-entrances. Second, the R-squared values of the models with country and language-cluster fixed effects are much higher (27%) and the estimate of the effect of relatedness on entry is about three times as large as in the baseline model (0.154 vs 0.429). Third, similar levels of explanatory power are observed in other papers testing the principle of relatedness (for example see Balland et al., 2018; and for a general overview see Li and Neffke, 2024). Interpreting the effect size also indicates the significance of relatedness as a correlate of entry. The mean of the relatedness measure in the full sample is 0.326, with a standard deviation of 0.168. Moving from the mean to one standard deviation above it is associated with a 7.2–percentage-point increase in the probability of entry, nearly double the base rate of entry of 5–6% to about 12–13%.

Fig. 2 shows the network of related software bundles following the visualization approach of (Hidalgo et al., 2007). Fig. 2A highlights a few example software bundles, with labels listing all programming languages within each. We then focus on the entry and exit patterns of three countries on Fig. 2B. In each case, entries occur into bundles that are adjacent to existing specializations, while exits tend to occur out of more weakly connected bundles.

Fig. 2B highlights contrasting dynamics in countries' software capability portfolios, measured as entries and exits in revealed comparative advantage (RCA) across software bundles. China exhibits multiple entries, consistent with an expanding and diversifying software profile: it is increasingly likely to develop comparative advantage in additional capability bundles, suggesting active broadening of its OSS specializations. Great Britain shows comparatively few transitions, indicating a more stable specialization structure over the period—its portfolio appears to evolve gradually, with limited reallocation across bundles. Russia, in contrast, displays several exits, consistent with a contraction or relative weakening of specialization in a set of capability bundles, likely related to large scale emigration of software developers in the wake of the 2022 invasion of Ukraine (Wachs, 2023).

We then explore the principle of relatedness in the context of exits (Table 5). We consider exits as countries that were specialized in a software bundle (RCA $\geq$ 1) in 2020 and 2021 and later lost their comparative advantage (RCA $<$ 1) in 2022 and 2023 (e.g. $M_{cl} = \{1,1,0,0\}$ for the years going from 2020 to 2023). The negative and significant effect of relatedness across both simpler and more complex models indicates that countries are less likely to lose their advantage in software bundles that are related to those they currently specialize in. Again, the effects of relatedness are overall mild ($R^2 < 3\%$ on the baseline model) but are robust to the inclusion of country and bundle fixed-effects, showing that they go beyond what can explained based on the statistic characteristics of a country or bundle.

4.3. Robustness checks and alternative approaches

We verify the consistency of our findings through multiple alternative specifications and modeling strategies. First, we confirm that the main results hold when varying RCA thresholds or applying Tobit regressions to account for the nature of the dependent variables (see section S10 and S11 in the Supplementary information). We also verify that restricting the sample to countries with fully available macroeconomic data does not alter the significance or direction of our coefficients, indicating that sample selection does not drive our conclusions (see section S13 in the Supplementary information). Further, to address potential statistical concerns, we check for multicollinearity through VIF analyses and remove mathematical dependencies from key variables, ensuring that the variables used are valid and adequately capture different dimensions of complexity (see section S14 in the Supplementary information for more details).

Second, we go back to our alternative definitions of $ECI^{software}$ to show that our conclusions hold when we define software complexity on different basis, either by grouping languages into theoretical clusters (e.g., web-oriented or system-level languages; see S3) or by using a measure based on topics (S4), or simply by consider languages themselves (S1). We find that even when we change the unit of observation to topics, $ECI^{software}$ remains positively correlated with GDP per capita and negatively correlated with income inequality.

Our findings on the relationship between $ECI^{software}$ and macroeconomic indicators are based on cross-sectional regressions. In section S15

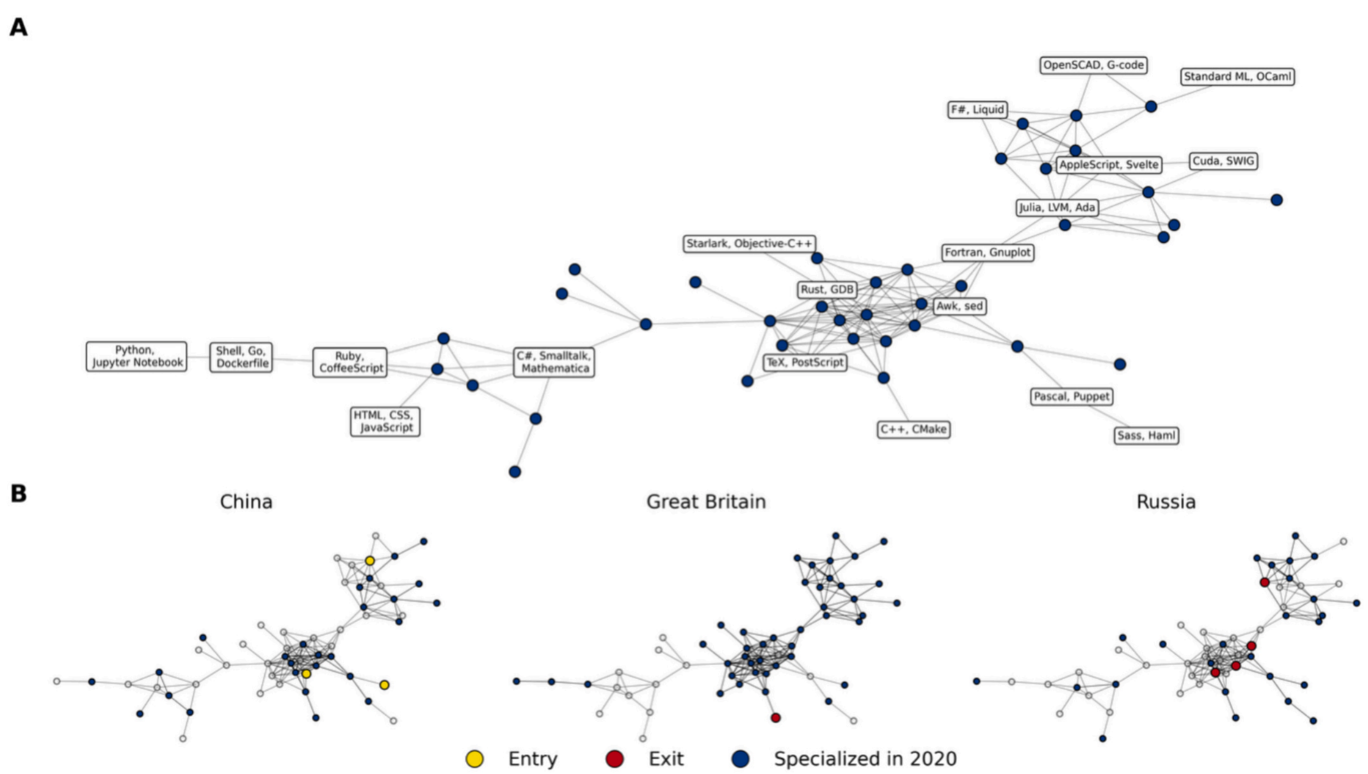

Fig. 2. (A) Network representation of software bundle relatedness. **(B)** Changes in revealed comparative advantage (RCA) in programming languages clusters (2020–2023) in China, Great Britain, and Russia. Dark blue nodes indicate specialization in 2020–2021 (RCA $\geq$1), while yellow nodes indicate subsequent (2022–2023) specialization in software bundles, and red nodes indicate exits. Countries are more likely to specialize in new software bundles adjacent to their previous specializations. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Table 5
Exit models on countries losing revealed comparative advantage (RCA < 1) in software bundles (2020–2023). Standard errors are clustered at the country level. Significance codes: *$p < 0.1$, **$p < 0.05$, ***$p < 0.01$.

	Exit						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Relatedness density	−0.160***	−0.405***	−0.190***	−0.285**		−0.223***	−0.348***
	(0.033)	(0.105)	(0.044)	(0.116)		(0.043)	(0.099)
Ubiquity					−0.006	−0.027***	−0.018**
					(0.006)	(0.008)	(0.009)
Country FE	No	Yes	No	Yes	No	No	Yes
Software bundle FE	No	No	Yes	Yes	No	No	No
Observations	1544	1544	1544	1544	1544	1544	1544
R^2	0.023	0.185	0.116	0.257	0.001	0.035	0.187

of the Supplementary information, we replicate GDP growth models in the style of Hidalgo and Hausmann (2009). However, this is not recommended due to the limited time span of available data (2020–2023), since measures of complexity are structural measures that are connected to long term growth (so we should not expect significance in short time periods dominated by other dynamics, such as the covid bounce-back in this case). As expected, we find that neither $ECI^{software}$ nor ECI^{trade} significantly predicts GDP growth. Structural measures such as $ECI^{software}$ tend to be stable over time, whereas short-term growth outcomes are more volatile. Supporting this, we find that $ECI^{software}$ remains highly stable across years, with correlations exceeding 0.92 (see section S16 of the Supplementary information), suggesting its predictive power may become more apparent over longer time horizons. Additionally, we provide an extensive explanation of our instrumental variable approach, including extended models and tests in section S8 of the Supplementary information. However, testing for potential endogeneity using instruments for other complexity measures—or between complexity measures themselves, such as $ECI^{software}$ and $ECI^{technology}$—was beyond the scope of this paper. Together, these tests demonstrate that our main results are stable and robust, even when we account for alternative definitions, model specifications, and potential sources of bias.

5. Discussion

Here we expanded the study of economic complexity to include the software sector by leveraging recently published data on the geography of open-source software (OSS). By relying on the IP addresses of the developers contributing to OSS projects, instead of on self-reported locations (which can suffer from reporting bias (Hecht et al., 2011)), we were able to construct estimates of the geographic distribution of open-source software language knowledge for 100+ programming languages and use them to create internationally comparable estimates of economic complexity for the software sector and to study OSS's diffusion in the context of the principle of relatedness. Our study provides a cross-country measure of software economic complexity and demonstrates it complements well-established ECI metrics based on trade, patents and research.

Building on prior studies linking software specialization to broader skill formation and productivity gains (Brynjolfsson and Hitt, 2003; Nagle, 2019, 2018; Wright et al., 2023), our results indicate that countries with higher software-based economic complexity may be better equipped to generate inclusive growth—thereby reducing inequality. This aligns with research showing that knowledge-intensive economies can create wider opportunities for high-skilled labor, mitigating income disparities (Hartmann et al., 2017). Although not consistently significant across all models, the observed negative association between software complexity and emissions aligns directionally with prior evidence that digitally driven economies may reduce their reliance on resource-intensive activities (Haberl et al., 2020; Stojkoski et al., 2024). These points suggest that software complexity could serve as a policy-relevant indicator for steering economies towards less environmentally taxing activities. In sum, our study contributes to the literature by offering both an empirical measure of software capabilities and an interpretation, consistent with earlier scholarship, of how these capabilities might shape pathways of inclusive and sustainable growth.

We also found that $ECI^{software}$ complements other measures of economic complexity when explaining macro-outcomes. One plausible interpretation of this complementarity is that the overlap between these different activities is not exhaustive, and hence, the differences among them are informative. Patent data includes many non-software activities, such as patents in biotech or the life sciences. Similarly, research publication data also includes many non-software related sectors, such as publications in history or philosophy. Also, open-source software data may provide some additional granularity that might not be available in the other data sources. For example, OSS data involves hundreds of unique languages, which provide a resolution over the software sector that is larger than the one captured in research publication data. The idea that correlated measures of complexity can prove to be complementary is at the core of the idea of multidimensional complexity (Stojkoski et al., 2023a, 2023b), which is based on the idea that information on the geography of different activities (products, patents, papers, software, etc.) captures different levels of detail making them mutually reinforcing. In simple terms, they fill each other's "gaps."

But what can we make of these findings? First, that economic complexity measures derived from OSS production do indeed correlate significantly with GDP, inequality, and emissions suggests that software complexity can suggest productive diversification directions. The literature on economic development is rife with work advising economies to diversify towards more complex economic activities (Balland et al., 2018; Hausmann et al., 2014; Hidalgo, 2023). High economic complexity activities are associated with better wages and may face less competition in international markets than the production of more ubiquitous commodities. The question that remains is whether this advice can translate to software. We argue that many of the unique aspects of software make it especially attractive for specific kinds of diversification strategies.

Unlike physical products, software relies less on immobile factors, such as large manufacturing or processing plants and natural resources. At the same time, software outputs are highly tradable (OECD, 2023; Stojkoski et al., 2024) and digital products are known to be—on average—of relatively high complexity compared to physical products (Stojkoski et al., 2024). Further, transformer models on platforms like Hugging Face make deep learning accessible with pre-trained models that require significantly fewer resources (Wolf et al., 2019). This means that software provides new opportunities for structural upgrading that are less reliant on physical factors of production and more reliant on efforts to attract human capital. Combined with our finding that diversification in software follows the principle of relatedness, policymakers should seek to attract experts in complex software technologies most related to current areas of strength.

Future research could explore how AI-driven productivity gains might alter the rate at which regions diversify into more sophisticated software niches—and whether that facilitates or hinders upward movement in the digital value chain.

While our study suggests how to estimate, validate, and use measures of economic complexity based on software, it is also subject to several important limitations that may affect the interpretation of our results. First, because our data exclusively captures OSS activity on GitHub, we may underestimate important proprietary or closed-source capabilities—and overlook OSS activity on other platforms. This can lead us to systematically undervalue software complexity in certain economies (for instance, where non-GitHub or closed-source development is predominant). Even OSS projects hosted outside of GitHub are also different on average, for example they are more likely to be academic (Trujillo et al., 2022). Moreover, our assumption that GitHub-based OSS specialization reflects broader digital skills—while supported by research on OSS's role in innovation— may still introduces measurement error. Ultimately, some countries may possess stronger sofOSStware capabilities than our metrics reveal, which could influence the strength of the observed correlations with macroeconomic outcomes.

Second, applying product-complexity methods to programming languages poses conceptual challenges. We treat languages as distinct units of analysis, a choice which offers clear interpretability but simplifies the complex relationships between them. For instance, languages may relate through complementary usage (e.g., HTML and CSS) rather than hierarchical supply chains, meaning the "distance" between them may not perfectly map onto traditional complexity notions. We explored alternative specifications, such as considering individual languages or theoretical clusters instead of bundles as the basis for the ECI calculation in our robustness checks (see Supplementary Information). While these aggregations largely confirm our results, we retain the software bundle approach in our main analysis for its robustness. Ultimately, path-dependent software diversification may follow different patterns than those in manufacturing, and more granular data (e.g., at the project or framework level) will be valuable for future work.

Nevertheless, despite these limitations, our work represents a valuable step towards extending economic complexity analysis to the digital realm, offering insights into the geographic distribution of software capabilities and their potential impact on macroeconomic outcomes. Software complexity is a significant complement to trade, research, and technology complexity measures because it covers a specific and important class of capabilities; this is demonstrated by its ability to extend the predictive power of models of key macro-outcomes including growth, inequality, and emission intensity. As the digital economy continues to evolve, further research integrating diverse data sources will be crucial. Understanding how emerging technologies, particularly in artificial intelligence (Daniotti et al., 2025; Del Rio-Chanona et al., 2024), may alter the nature of software capabilities and pathways for diversification remains a key challenge for the future.

CRediT authorship contribution statement

Sándor Juhász: Writing – original draft, Formal analysis, Data curation, Conceptualization. **Johannes Wachs:** Writing – original draft, Formal analysis, Data curation, Conceptualization. **Jermain Kaminski:** Writing – original draft, Formal analysis, Data curation, Conceptualization. **César A. Hidalgo:** Writing – original draft, Formal analysis, Conceptualization.

Declaration of competing interest

César A. Hidalgo is a founder of Datawheel LLC and a creator of the Observatory of Economic Complexity (oec.world). All other authors have nothing to declare.

Acknowledgements

César A. Hidalgo acknowledges funding by the European Union project 101086712-LearnData-HORIZON-WIDERA-2022-TALENTS-01 financed by European Research Executive Agency (REA) (https://cordis.europa.eu/project/id/101086712), IAST funding from the French National Research Agency (ANR) under grant ANR-17-EURE-0010 (Investissements d'Avenir program), and the European Lighthouse of AI for Sustainability [grant number 101120237-HORIZON-CL4-2022-HUMAN-02]. Sándor Juhász's work was supported by the European Union's Marie Skłodowska-Curie Postdoctoral Fellowship Program (SUPPED, grant number 101062606). Johannes Wachs acknowledges support from the Hungarian National Scientific Fund (OTKA FK 145960). Jermain Kaminski acknowledges the support of the Dutch Research Council (NWO) through the Starter Grant. The authors would like to thank Viktor Stojkoski for support in the multidimensional estimates of economic complexity and the econometric models. César A. Hidalgo is a founder of Datawheel LLC and a creator of the Observatory of Economic Complexity (oec.world).

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.respol.2026.105422.

Data availability

All code and data are available in this GitHub repository.

References

Alabdulkareem, A., Frank, M.R., Sun, L., AlShebli, B., Hidalgo, C., Rahwan, I., 2018. Unpacking the polarization of workplace skills. Sci. Adv. 4, eaao6030. https://doi.org/10.1126/sciadv.aao6030.

Apostol, S., Hernández-Rodríguez, E., 2024. Digitalisation in European regions: unravelling the impact of relatedness and complexity on digital technology adoption and productivity growth. Ind. Innov. 32, 772–801. https://doi.org/10.1080/13662716.2024.2423731.

Arthur, W.B., 1994. Increasing Returns and Path Dependence in the Economy. University of Michigan Press, Ann Arbor, MI.

Audretsch, D.B., Feldman, M.P., 1996. R&D spillovers and the geography of innovation and production. Am. Econ. Rev. 86, 630–640.

Aum, S., Shin, Y., 2024. Is Software Eating the World? (Working Paper, vol. No. 32591 National Bureau of Economic Research, Cambridge, MA. https://doi.org/10.3386/w32591.

Autant-Bernard, C., 2001. The geography of knowledge spillovers and technological proximity. Econ. Innov. New Technol. 10, 237–254. https://doi.org/10.1080/10438590100000010.

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

Balland, P.-A., Boschma, R., Crespo, J., Rigby, D.L., 2018. 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., Jara-Figueroa, C., Petralia, S.G., Steijn, M.P.A., Rigby, D.L., Hidalgo, C.A., 2020. Complex economic activities concentrate in large cities. Nat. Hum. Behav. 4. https://doi.org/10.1038/s41562-019-0803-3.

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

Bandeira Morais, M., Swart, J., Jordaan, J.A., 2018. Economic Complexity and Inequality: Does Productive Structure Affect Regional Wage Differentials in Brazil? (Working Paper No. 18-11) Utrecht University School of Economics, Utrecht.

Ben Saâd, M., Assoumou-Ella, G., 2019. Economic complexity and gender inequality in education: an empirical study. Econ. Bull. 39, 321–334 https://hal.science/hal-03426719/.

Boschma, R., Minondo, A., Navarro, M., 2013. The emergence of new industries at the regional level in Spain: a proximity approach based on product relatedness. Econ. Geogr. 89, 29–51. https://doi.org/10.1111/j.1944-8287.2012.01170.x.

Bottai, C., Di Iorio, J., Iori, M., 2024. Reinterpreting economic complexity: a co-clustering approach. arXiv preprint. https://doi.org/10.48550/arXiv.2406.16199 arXiv:2406.16199.

Boudreau, K.J., 2012. Let a thousand flowers bloom? An early look at large numbers of software app developers and patterns of innovation. Organ. Sci. 23, 1409–1427. https://doi.org/10.1287/orsc.1110.0678.

Branstetter, L.G., Drev, M., Kwon, N., 2019. Get with the program: software-driven innovation in traditional manufacturing. Manag. Sci. 65, 541–558. https://doi.org/10.1287/mnsc.2017.2960.

Brynjolfsson, E., Hitt, L.M., 1998. Beyond the productivity paradox. Commun. ACM 41, 49–55. https://doi.org/10.1145/280324.280332.

Brynjolfsson, E., Hitt, L.M., 2003. Computing productivity: firm-level evidence. Rev. Econ. Stat. 85, 793–808. https://doi.org/10.1162/003465303772815736.

Brynjolfsson, E., Saunders, A., 2010. Wired for Innovation: How Information Technology Is Reshaping the Economy. MIT Press, Cambridge, MA.
Bustos, S., Yıldırım, M.A., 2022. Production ability and economic growth. Res. Policy 51, 104153. https://doi.org/10.1016/j.respol.2020.104153.
Bustos, S., Gomez, C., Hausmann, R., Hidalgo, C.A., 2012. The dynamics of nestedness predicts the evolution of industrial ecosystems. PLoS One 7, e49393. https://doi.org/10.1371/journal.pone.0049393.
Can, M., Gozgor, G., 2017. The impact of economic complexity on carbon emissions: evidence from France. Environ. Sci. Pollut. Res. 24, 16364–16370. https://doi.org/10.1007/s11356-017-9219-7.
Cennamo, C., Santaló, J., 2019. Generativity tension and value creation in platform ecosystems. Organ. Sci. 30, 617–641. https://doi.org/10.1287/orsc.2018.1270.
Chattergoon, B., Kerr, W.R., 2022. Winner takes all? Tech clusters, population centers, and the spatial transformation of US invention. Res. Policy 51, 104418. https://doi.org/10.1016/j.respol.2021.104418.
Chávez, J.C., Mosqueda, M.T., Gómez-Zaldívar, M., 2017. Economic complexity and regional growth performance: evidence from the Mexican economy. Rev. Reg. Stud. 47, 201–219. https://doi.org/10.52324/001c.8023.
Chinazzi, M., Gonçalves, B., Zhang, Q., Vespignani, A., 2019. Mapping the physics research space: a machine learning approach. EPJ Data Sci. 8, 1–18. https://doi.org/10.1140/epjds/s13688-019-0210-z.
Chu, L.K., Hoang, D.P., 2020. How does economic complexity influence income inequality? New evidence from international data. Economic Analysis and Policy 68, 44–57. https://doi.org/10.1016/j.eap.2020.08.004.
Ciuriak, D., Ptashkina, M., 2020. Towards a Robust Architecture for the Regulation of Data and Digital Trade (CIGI Paper No. 240). Centre for International Governance Innovation, Waterloo. https://doi.org/10.2139/ssrn.3423394.
Corrado, C., Hulten, C., Sichel, D., 2005. Measuring capital and technology: An expanded framework. In: Corrado, C., Haltiwanger, J., Sichel, D. (Eds.), Measuring Capital in the New Economy. University of Chicago Press, Chicago, IL, pp. 11–46.
Daniotti, S., Wachs, J., Feng, X., Neffke, F., 2025. Who is using AI to code? Global diffusion and impact of generative AI. arXiv preprint. https://doi.org/10.48550/arXiv.2506.08945 arXiv:2506.08945.
Del Rio-Chanona, R.M., Laurentsyeva, N., Wachs, J., 2024. Large language models reduce public knowledge sharing on online Q&a platforms. Proc. Natl. Acad. Sci. Nexus 3, pgae400. https://doi.org/10.1093/pnasnexus/pgae400.
Doğan, B., Driha, O.M., Balsalobre Lorente, D., Shahzad, U., 2021. The mitigating effects of economic complexity and renewable energy on carbon emissions in developed countries. Sustain. Dev. 29, 1–12. https://doi.org/10.1002/sd.2125.
Domini, G., 2022. Patterns of specialization and economic complexity through the lens of universal exhibitions, 1855-1900. Explor. Econ. Hist. 83, 101421. https://doi.org/10.1016/j.eeh.2021.101421.
Eghbal, N., 2020. Working in Public: The Making and Maintenance of Open Source Software. Stripe Press, San Francisco, CA.
Farinha, T., Balland, P.-A., Morrison, A., Boschma, R., 2019. What drives the geography of jobs in the US? Unpacking relatedness. Ind. Innov. 26, 988–1022. https://doi.org/10.1080/13662716.2019.1591940.
Fritz, B.S., Manduca, R.A., 2021. The economic complexity of US metropolitan areas. Reg. Stud. 55, 1299–1310. https://doi.org/10.1080/00343404.2021.1884215.
Gao, J., Zhou, T., 2018. Quantifying China's regional economic complexity. Phys. A 492, 1591–1603. https://doi.org/10.1016/j.physa.2017.11.084.
Goldbeck, M., 2025. Bit by bit: colocation and the death of distance in software developer networks. J. Econ. Geogr. 25, 569–583. https://doi.org/10.1093/jeg/lbaf002.
Gortmaker, J., 2025. Open source software policy in industry equilibrium. Working Paper, Harvard University. https://jeffgortmaker.com/files/Open_Source_Software_Policy_in_Industry_Equilibrium.pdf.
Greenstein, S., Nagle, F., 2014. Digital dark matter and the economic contribution of apache. Res. Policy 43, 623–631. https://doi.org/10.1016/j.respol.2014.01.003.
Guevara, M.R., Hartmann, D., Aristarán, M., Mendoza, M., Hidalgo, C.A., 2016. The research space: using career paths to predict the evolution of the research output of individuals, institutions, and nations. Scientometrics 109, 1695–1709. https://doi.org/10.1007/s11192-016-2125-9.
Haberl, H., Wiedenhofer, D., Virág, D., Kalt, G., Plank, B., Brockway, P., Fishman, T., Hausknost, D., Krausmann, F., Leon-Gruchalski, B., 2020. A systematic review of the evidence on decoupling of GDP, resource use and GHG emissions, part II: synthesizing the insights. Environ. Res. Lett. 15, 065003. https://doi.org/10.1088/1748-9326/ab842a.
Hartmann, D., Guevara, M.R., Jara-Figueroa, C., Aristarán, M., Hidalgo, C.A., 2017. Linking economic complexity, institutions, and income inequality. World Dev. 93, 75–93. https://doi.org/10.1016/j.worlddev.2016.12.020.
Hausmann, R., Hidalgo, C.A., Bustos, S., Coscia, M., Simoes, A., Yildirim, M.A., 2014. The Atlas of Economic Complexity: Mapping Paths to Prosperity. MIT Press, Cambridge, MA.
Hecht, B., Hong, L., Suh, B., Chi, E.H., 2011. Tweets from Justin Bieber's Heart: The Dynamics of the Location Field in User Profiles, in: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. ACM, New York, NY, pp. 237–246. https://doi.org/10.1145/1978942.1978976.
Hidalgo, C.A., 2021. Economic complexity theory and applications. Nat. Rev. Phys. 3, 92–113. https://doi.org/10.1038/s42254-020-00275-1.
Hidalgo, C.A., 2023. The policy implications of economic complexity. Res. Policy 52, 104863. https://doi.org/10.1016/j.respol.2023.104863.
Hidalgo, C.A., Hausmann, R., 2009. The building blocks of economic complexity. Proc. Natl. Acad. Sci. 106, 10570–10575. https://doi.org/10.1073/pnas.0900943106.
Hidalgo, C.A., Stojkoski, V., 2025. The theory of economic complexity. arXiv preprint. https://doi.org/10.48550/arXiv.2506.18829 arXiv:2506.18829.
Hidalgo, C.A., Klinger, B., Barabási, A.-L., Hausmann, R., 2007. The product space conditions the development of nations. Science 317, 482–487. https://doi.org/10.1126/science.1144581.
Hidalgo, Balland, P.-A., Boschma, R., Delgado, M., Feldman, M., Frenken, K., Glaeser, E., He, C., Kogler, D.F., Morrison, A., Neffke, F., Rigby, D., Stern, S., Zheng, S., Zhu, S., 2018. The principle of relatedness. In: Morales, A.J., Gershenson, C., Braha, D., Minai, A.A., Bar-Yam, Y. (Eds.), Unifying Themes in Complex Systems IX, Springer Proceedings in Complexity. Springer International Publishing, pp. 451–457. https://doi.org/10.1007/978-3-319-96661-8_46.
Hubacek, K., Chen, X., Feng, K., Wiedmann, T., Shan, Y., 2021. Evidence of decoupling consumption-based CO2 emissions from economic growth. Adv. Appl. Energy 4, 100074. https://doi.org/10.1016/j.adapen.2021.100074.
Inoua, S., 2023. A simple measure of economic complexity. Res. Policy 52, 104793. https://doi.org/10.1016/j.respol.2023.104793.
Jaffe, A.B., 1986. Technological opportunity and spillovers of R&D: evidence from firms' patents, profits and market value. Am. Econ. Rev. 76, 984–1001. https://doi.org/10.3386/w1815.
Jara-Figueroa, C., Jun, B., Glaeser, E.L., Hidalgo, C.A., 2018. The role of industry-specific, occupation-specific, and location-specific knowledge in the growth and survival of new firms. Proc. Natl. Acad. Sci. 115, 12646–12653. https://doi.org/10.1073/pnas.1800475115.
Koch, P., 2021. Economic complexity and growth: Can value-added exports better explain the link? Econ. Lett. 198, 109682. https://doi.org/10.1016/j.econlet.2020.109682.
Kogler, D.F., Rigby, D.L., Tucker, I., 2013. Mapping knowledge space and technological relatedness in US cities. Eur. Plan. Stud. 21, 1374–1391. https://doi.org/10.1080/09654313.2012.755832.
Korkmaz, G., Santiago Calderón, J.B., Kramer, B.L., Guci, L., Robbins, C.A., 2024. From GitHub to GDP: a framework for measuring open source software innovation. Res. Policy 53, 104954. https://doi.org/10.1016/j.respol.2024.104954.
Lapatinas, A., Garas, A., Boleti, E., Kyriakou, A., 2019. Economic Complexity And Environmental Performance: Evidence From A World Sample (MPRA Woring Paper No. 92833). https://ideas.repec.org/p/pra/mprapa/92833.html.
Lee, K.-K., Vu, T.V., 2019. Economic complexity, human capital and income inequality: a cross-country analysis. Jpn. Econ. Rev. 71, 695–718. https://doi.org/10.1007/s42973-019-00026-7.
Li, Y., Neffke, F.M.H., 2024. Evaluating the principle of relatedness: estimation, drivers and implications for policy. Res. Policy 53, 104952. https://doi.org/10.1016/j.respol.2024.104952.
Liang, S., Tan, Q., 2024. Can the digital economy accelerates China's export technology upgrading? Based on the perspective of export technology complexity. Technol. Forecast. Soc. Change 199, 123052. https://doi.org/10.1016/j.techfore.2023.123052.
Liu, H., Wang, L., Shen, Y., 2023. Can digital technology reduce carbon emissions? Evidence from Chinese cities. Front. Ecol. Evol. 11, 1205634. https://doi.org/10.3389/fevo.2023.1205634.
Mariani, M.S., Ren, Z.-M., Bascompte, J., Tessone, C.J., 2019. Nestedness in complex networks: observation, emergence, and implications. Phys. Rep. 813, 1–19. https://doi.org/10.1016/j.physrep.2019.04.001.
Mealy, P., Teytelboym, A., 2020. Economic complexity and the green economy. Res. Policy 51, 103948. https://doi.org/10.1016/j.respol.2020.103948.
Mealy, P., Farmer, J.D., Teytelboym, A., 2019. Interpreting economic complexity. Sci. Adv. 5, eaau1705. https://doi.org/10.1126/sciadv.aau1705.
Meyerovich, L.A., Rabkin, A.S., 2013. Empirical Analysis of Programming Language Adoption, in: Proceedings of the 2013 ACM SIGPLAN International Conference on Object Oriented Programming Systems Languages & Applications. ACM, New York, NY, pp. 1–18. https://doi.org/10.1145/2509136.2509515.
Muneepeerakul, R., Lobo, J., Shutters, S.T., Goméz-Liévano, A., Qubbaj, M.R., 2013. Urban economies and occupation space: Can they get "there" from "Here"? PLoS One 8, e73676. https://doi.org/10.1371/journal.pone.0073676.
Nagle, F., 2018. Learning by contributing: gaining competitive advantage through contribution to crowdsourced public goods. Organ. Sci. 29, 569–587. https://doi.org/10.1287/orsc.2018.1202.
Nagle, F., 2019. Open source software and firm productivity. Manag. Sci. 65, 1191–1215. https://doi.org/10.1287/mnsc.2017.2977.
Natera, J.M., Castellacci, F., 2021. Transformational complexity, systemic complexity and economic development. Res. Policy 50, e104275. https://doi.org/10.1016/j.respol.2021.104275.
Neffke, F., Henning, M., 2013. Skill relatedness and firm diversification. Strateg. Manag. J. 34, 297–316. https://doi.org/10.1002/smj.2014.
Neffke, F., Henning, M., Boschma, R., 2011. How do regions diversify over time? Industry relatedness and the development of new growth paths in regions. Econ. Geogr. 87, 237–265. https://doi.org/10.1111/j.1944-8287.2011.01121.x.
OECD, 2023. Handbook on Measuring Digital Trade, Second edition. Organisation for Economic Co-operation and Development, Paris. https://doi.org/10.1787/ac99e6d3-en.
Pérez-Balsalobre, S., Llano Verduras, C., Díaz-Lanchas, J., 2019. Measuring Subnational Economic Complexity: An Application With Spanish Data (JRC Working Paper No. 05/2019). Seville, European Commission, Joint Research Centre.
Poncet, S., de Waldemar, F.S., 2013. Economic Complexity and Growth. Rev. écon. 64, 495–503. https://doi.org/10.3917/reco.643.0495.
Poncet, S., de Waldemar, F.S., 2015. Product relatedness and firm exports in China. World Bank Econ. Rev. 29, 579–605. https://doi.org/10.1596/27688.
Rahmati, P., Tafti, A., Westland, J.C., Hidalgo, C., 2021. When all products are digital: complexity and intangible value in the ecosystem of digitizing firms. Manag. Inf. Syst. Q. 45, 1025–1058. https://doi.org/10.25300/MISQ/2021/15384.

Rock, D., 2019. Engineering Value: The Returns To Technological Talent And Investments In Artificial Intelligence (SSRN Working Paper No. 3427412). https://doi.org/10.2139/ssrn.3427412.

Romero, J.P., Gramkow, C., 2021. Economic complexity and greenhouse gas emissions. World Dev. 139, 105317. https://doi.org/10.1016/j.worlddev.2020.105317.

Salinas, M.G., 2021. Proximity and Horizontal Policies: The Backbone of Export Diversification (Working Paper No. 2021/064). International Monetary Fund. https://doi.org/10.5089/9781513571614.001.

Sbardella, A., Pugliese, E., Pietronero, L., 2017. Economic development and wage inequality: a complex system analysis. PLoS One 12, e0182774. https://doi.org/10.1371/journal.pone.0182774.

Servedio, V.D., Bellina, A., Calò, E., De Marzo, G., 2024. Economic complexity in mono-partite networks. arXiv preprint. https://doi.org/10.48550/arXiv.2405.04158 arXiv:2405.04158.

Shapiro, C., Varian, H.R., 1999. Information Rules: A Strategic Guide to the Network Economy. Press, Boston, MA, Harvard Bus.

Shrestha, N., Botta, C., Barik, T., Parnin, C., 2022. Here we go again: why is it difficult for developers to learn another programming language? Commun. ACM 65, 91–99. https://doi.org/10.1145/3511062.

Stojkoski, V., Utkovski, Z., Kocarev, L., 2016. The impact of services on economic complexity: service sophistication as route for economic growth. PLoS One 11, e0161633. https://doi.org/10.1371/journal.pone.0161633.

Stojkoski, V., Koch, P., Coll, E., Hidalgo, C.A., 2023a. The growth, geography, and implications of trade in digital products. arXiv preprint. https://doi.org/10.48550/arXiv.2310.02253 arXiv:2310.02253.

Stojkoski, V., Koch, P., Hidalgo, C.A., 2023b. Multidimensional economic complexity and inclusive green growth. Commun. Earth Environ. 4, 130. https://doi.org/10.1038/s43247-023-00770-0.

Stojkoski, V., Koch, P., Coll, E., Hidalgo, C.A., 2024. Estimating digital product trade through corporate revenue data. Nat. Commun. 15. https://doi.org/10.1038/s41467-024-49141-z.

Tacchella, A., Cristelli, M., Caldarelli, G., Gabrielli, A., Pietronero, L., 2012. A new metrics for countries' fitness and products' complexity. Sci. Rep. 2. https://doi.org/10.1038/srep00723.

Tian, S., Wu, Y., Nanjing Agricultural University, 2025. Southwestern University. In: Zhou, W. (Ed.), Digitalization And Income Inequality: Evidence From Households (Working Paper No. 250008–2). Asian Development Bank. https://doi.org/10.22617/WPS250008-2.

Trujillo, M.Z., Hébert-Dufresne, L., Bagrow, J., 2022. The penumbra of open source: projects outside of centralized platforms are longer maintained, more academic and more collaborative. EPJ Data Sci. 11. https://doi.org/10.1140/epjds/s13688-022-00345-7.

Valverde, S., Solé, R.V., 2015. Punctuated equilibrium in the large-scale evolution of programming languages. J. R. Soc. Interface 12, 20150249. https://doi.org/10.1098/rsif.2015.0249.

Valverde, Sole, 2015. A cultural diffusion model for the rise and fall of programming languages. Hum. Biol. 87, 224. https://doi.org/10.13110/humanbiology.87.3.0224.

Wachs, J., 2023. Digital traces of brain drain: developers during the Russian invasion of Ukraine. EPJ Data Sci. 12. https://doi.org/10.1140/epjds/s13688-023-00389-3.

Wachs, J., Nitecki, M., Schueller, W., Polleres, A., 2022. The geography of open source software: evidence from GitHub. Technol. Forecast. Soc. Change 176, 121478. https://doi.org/10.1016/j.techfore.2022.121478.

Wang, Q., Zhang, F., 2021. The effects of trade openness on decoupling carbon emissions from economic growth–evidence from 182 countries. J. Clean. Prod. 279, 123838. https://doi.org/10.1016/j.jclepro.2020.123838.

Weber, I., Semieniuk, G., Westland, T., Liang, J., 2021. What you Exported Matters: Persistence in Productive Capabilities across Two Eras of Globalization (Working Paper No. 2021-02). University of Massachusetts, Department of Economics, Amherst, MA. https://doi.org/10.7275/21780201.

Wiedenhofer, D., Virág, D., Kalt, G., Plank, B., Streeck, J., Pichler, M., Mayer, A., Krausmann, F., Brockway, P., Schaffartzik, A., 2020. A systematic review of the evidence on decoupling of GDP, resource use and GHG emissions, part I: bibliometric and conceptual mapping. Environ. Res. Lett. 15, 063002. https://doi.org/10.1088/1748-9326/ab8429.

Wolf, T., Debut, L., Sanh, V., Chaumond, J., Delangue, C., Moi, A., Cistac, P., Rault, T., Louf, R., Funtowicz, M., et al., 2019. Huggingface's transformers: state-of-the-art natural language processing. arXiv preprint. https://doi.org/10.48550/arXiv.1910.03771 arXiv:1910.03771.

Wright, N.L., Nagle, F., Greenstein, S., 2023. Open source software and global entrepreneurship. Res. Policy 52, 104846. https://doi.org/10.1016/j.respol.2023.104846.

Zhang, Z., Chen, L., Li, J., Ding, S., 2024. Digital economy development and carbon emission intensity—mechanisms and evidence from 72 countries. Sci. Rep. 14, 28459. https://doi.org/10.1038/s41598-024-78831-3.