Related variety and state-sponsored R&D collaboration: a geographical and industrial analysis in Czechia

Petr Horák[1,2,*]

[1] Department of Social Geography and Regional Development, Faculty of Science, Charles University, Czechia
[2] Technology Agency of the Czech Republic, Czechia
* Corresponding author: horakpe3@natur.cuni.cz

ABSTRACT

This paper aims to explore the influence of related variety on direct state-supported R&D cooperation across various geographical levels to understand regional performance differentiation and economic base restructuring in Czechia by employing Frenken et al.'s (2007) methodological approach to calculate a related and unrelated variety for all NACE and NACE C-Manufacturing. Findings indicate that the city of Prague has the highest unrelated and related variety, followed by the cities of Brno, Ostrava, and Pilsen. Calculation just for C-Manufacturing changes the ordering significantly. Furthermore, intra-regional and extra-regional pairwise R&D cooperation in joint projects is calculated. The cluster analysis of Czech microregional data (SO ORP) reveals patterns such as emerging collaborators and collaboration powerhouses. Linear regression analyses established a strong positive association between R&D collaboration intensity and related variety, while a negative link was observed with unrelated variety. Similar relationships were observed in the manufacturing sector (NACE-C).

1. Introduction

The differentiation of regional performance and the different restructuring of the economic base of the region began to be explained in the last 15 years through the concept of related variety developed by the Dutch school of evolutionary economic geography (Frenken and Boschma 2007; Frenken et al. 2007). The concept follows and to a certain extent overcomes the traditional dual concept of the process of spillover of knowledge between companies and institutions, where on the one hand the advantages resulting from the concentration of a certain industry in space (so-called Marshall-Arrow-Romer externality) and on the other the second benefits resulting from the creation of knowledge spillovers within a diversified economic structure (the so-called J. Jacobs externality). The work of Nooteboom (2000) highlighting the role of cognitive proximity in various spheres of communication and interaction and indirectly in the production process became an equally important source of inspiration.

The related variety allows us to analytically capture the potential for cooperation and knowledge transfer in various geographical units. Moreover, the contribution of related variety to the overall growth and economic development of the region has been documented in existing studies (Frenken et al. 2007). This approach works with the diversity of industries within the region, which are cognitively connected and can maximize the potential of opportunities, growth of existing industries, and the potential of local resources for new industries.

In Czechia, research analysis mapping related / unrelated varieties is very limited. An exception is the research by Květoň and Šafr (2019), who measured unrelated variety and regional embeddedness of interregional and intersectoral relations in Czechia. Blažek et al. (2016) tried to clarify different methods of calculating related variety using the example of Czech R&D projects. Furthermore, geographical and cognitive proximity was clarified in the example of R&D collaborative projects (Květoň et al. 2022), however, these authors did not directly apply the concept of related variety. Therefore, this paper is the first attempt to calculate the unrelated and related variety for the whole of Czechia on regional, district, and municipalities with extended competence levels. The article also provides a partial reflection on the critique of the methodological approach to related variety as presented by Bathelt and Storper (2023).

The overarching goal of the article is to investigate the factors influencing the intensity of R&D collaborations in state-supported projects, focusing specifically on the role of related and unrelated variety across sectors and regions. By conducting multiple linear regression analyses, the article tests a series of hypotheses to elucidate how the diversity and specificity of industries within a region impact the propensity for research and development collaborations.

The article is designed as follows. First, the concept of related and unrelated variety is described, and current knowledge from the Czech environment is also emphasized. Subsequently, the research question is presented. Next, the methodical approach to the measurement of related variety and also to R&D cooperation is described in detail. In the following section, the hypotheses are tested and empirical results are interpreted.

2. Conceptual departures for related variety assessment

Related variety refers to the co-location of different sectors sharing commonalities and complementary competencies, which is conducive to knowledge spillovers underpinning regional growth and innovation (Corradini and Vanino 2022).

Research on related and unrelated variety has been ongoing for several years, with many studies exploring the effects of these concepts on regional growth, innovation, and entrepreneurship. Regarding the related variety and innovation process, previous research suggests that related variety can enhance regional innovation. When industries in a region are cognitively similar and have inter-industry knowledge spillovers, it becomes easier for innovation to occur (Martynovich and Taalbi 2022; Ejdemo and Örtqvist 2022). Moreover, Innocenti et al. (2021) emphasized that local related variety enhances the overall innovation rate and can contribute to recombination or incremental innovation. In terms of technological breakthroughs, related variety would raise the likelihood of innovations in general, but unrelated variety would raise the likelihood of breakthrough innovations, which in themselves are rare (Castaldi et al. 2015).

Furthermore, researchers have explored the role of related variety in regional diversification and path development (Yeung 2020), the relevance of relatedness research in economic diversification and regional competitiveness (Ferraz et al. 2021), the integration of related variety and strategic coupling in understanding regional industrial diversification and economic resilience (Yeon et al. 2022) or the relationship between relatedness, growth, and industry clustering (Bond-Smith and McCann 2019).

These studies have contributed to a deeper understanding of the role of related variety in economic geography research. They have highlighted the importance of relatedness and diversity in economic activities for regional competitiveness, growth, employment, and resilience.

Researchers have also emphasized the need to consider the social, cultural, and institutional dimensions of economic activities and the importance of context sensitivity in economic-geographic theorizing. The current knowledge in the related variety research

provides insights into the complexities and dynamics of economic systems within different spatial contexts.

Researchers use various methods to measure related and unrelated variety in their studies. One methodology used to compute related and unrelated variety is based on entropy measures (e.g. Frenken et al. 2010) and this method is also applied in this paper. However, it is necessary to point out that the whole calculation of kin diversity is to some extent an "ex-ante approach to the evaluation of cooperation and knowledge transfer" (see Blažek et al. 2017), and a higher kin diversity does not guarantee a more effective transfer of knowledge and information but expresses a certain assumption for such cooperation. This methodological approach presupposes that relatively diverse firms are cognitively close enough to understand each other and cooperate, but at the same time far enough away not to compete with each other. Companies and their representatives can therefore "understand" each other and have something to offer, but at the same time, they do not threaten each other on the market.

3. Research question

The central inquiry of this study aims to unravel the complex interplay between regional R&D collaborations within state-supported projects and the potential influence of the related variety within the SO ORP. Utilizing cluster analysis, the research seeks to delineate distinct patterns or clusters of regions based on their R&D collaborative dynamics. The central research question driving this investigation is: "How does R&D collaboration, as manifested by collaborating firms and research organizations in state-supported projects, relate to the related variety in Czech microregions (SO ORP)?"

Following this primary research question and based on the current state of related variety knowledge (e.g. Bathelt and Storper 2023), several hypotheses have been developed to provide a structured approach to addressing the research question:

Hypothesis 1: Based on Květoň and Horák (2018), who clarified the differentiation of R&D capacities at the regional NUTS 3 level in Czechia when subjected to k-means clustering based on relative joint R&D projects, related variety and unrelated variety, the SO ORP (Czech microregions) will yield more than two distinct clusters.

Hypothesis 2: Based on the current state of knowledge about related variety in different countries (Wise and Anderson 2017, Boschma and Iammarino 2009) we expect that the intensity of R&D collaboration in state-supported projects will be positively associated with the related variety.

Hypothesis 3: The intensity of R&D collaboration of state-supported projects in the manufacturing sector (NACE-C) is positively associated with the related variety specific to manufacturing.

4. Methodology

This study draws upon the methodological approach of "Related variety" presented by Frenken et al. (2007). This approach allows us to analytically capture the potential for cooperation and knowledge transfer in the geographical unit. Moreover, the contribution of related variety to the overall growth and economic development of the region has been documented in existing studies (Frenken et al. 2007). This approach works with the diversity of industries within the geographical unit, which is cognitively connected and can maximize the potential of opportunities, growth of existing industries, and the potential of local resources for new industries.

4.1 Data

Underlying data for the assessment of related and unrelated variety are drawn from the Register of Economic Subjects (RES). This source provides information on all economically active entities in Czechia. From this data, it is possible to filter out the legal persons engaged in business, i.e. firms. In this paper, all legal forms of business are selected. From the MagnusWeb database, information is secured on the number of employees of individual firms. Data for cooperation in R&D are drawn from the IS VAVAI database. The related and unrelated variety is calculated for:
1) the full breadth of NACE 2-digit industries
2) for NACE C-Manufacturing.

The results and underlying calculation are published on GitHub to enable further research: https://github.com/ph1559/related-variety/.

4.2 Data limitation

The sources used, despite being the best publicly available, have serious limitations of which the analysis in this paper is aware. First, the number of employees in MagnusWeb may not be available for all firms listed in RES. For this reason, the available sample of data is listed below. Tab. 1 presents all firms with more than one employee by RES compared to the number of available employees from MagnusWeb. The version of RES and number of employees is relative to the year 2021. In the absence of information on the number of employees for 2021, the nearest available value is used. Firms with available data on the number of employees were further used to calculate the unrelated/related variety.

Second, it should also be noted that some larger concerns do not split the number of employees by production facilities. The employers that stand out the most are Škoda Auto, Siemens, Bosch, Honeywell.

Škoda Auto a.s. is classified in the location Mladá Boleslav (CZ020), with the listed number of employees as 35,063. The listed production plants of Škoda Auto a.s. are located in Mladá Boleslav, Kvasiny

Region	NUTS 3 region name	Firms with information about the number of employees (MagnusWeb)	Total firms with one or more employees (RES)	Proportion of available employee data
CZ010	Prague	31,863	69,448	45.9%
CZ020	Central Bohemian Region	10,227	17,445	58.6%
CZ031	South Bohemian Region	5,968	9,262	64.4%
CZ032	Pilsen Region	5,579	8,242	67.7%
CZ041	Karlovy Vary Region	2,308	3,645	63.3%
CZ042	Ústí nad Labem Region	6,087	8,797	69.2%
CZ051	Liberec Region	3,645	5,893	61.9%
CZ052	Hradec Králové Region	5,567	7,566	73.6%
CZ053	Pardubice Region	4,834	7,132	67.8%
CZ063	Vysočina Region	4,058	5,914	68.6%
CZ064	South Moravian Region	14,919	25,424	58.7%
CZ071	Olomouc Region	5,543	8,742	63.4%
CZ072	Zlín Region	6,534	9,021	72.4%
CZ080	Moravian-Silesian Region	12,134	17,328	70.0%
Total	All regions	119,266	203,859	58.5%

Source: Own calculations based on data drawn from RES and MAGNUS.

(CZ041) and Vrchlabí (CZ041). The division has approximately 20,000 employees in Mladá Boleslav, 9,000 in Kvasiny and 6,000 in Vrchlabí. Siemens, s.r.o. has 9,691 employees according to MagnusWeb and is divided into seven legal entities. Five of these entities are listed in Prague (CZ010). However, it also has production plants in Brno (CZ064), Drásov (CZ064), Frenštát pod Radhoštěm (CZ072), Trutnov (CZ041), Letohrad (CZ053) and Mohelnice (CZ071). Brno and Drásov has its legal entity and therefore the employee number corresponds to the correct NUTS 3 region.

Next concern is Robert Bosch. There are Robert Bosch subsidiaries in eight cities in Czechia. They have a total of four production plants, one service centre and one logistics warehouse. In total, they are divided into five legal entities. These partly reflect the territorial division of the Group. The Honeywell Group is divided into 4 legal entities in the RES. The biggest shortcoming in this case is that MagnusWeb does not provide the number of employees for its largest entity Honeywell Aerospace s.r.o. However, its spin-off plant in Olomouc (CZ071) is a separate entity. The activities of this concern are still concentrated in Prague (CZ010) and Brno (CZ064).

The data of other larger companies and foreign concerns might be subject to similar problems with the difference between the location of the legal entity and the location of the production facilities. The data are also sensitive to the reporting of agricultural production, which will play a lesser role in the following calculations. These limitations need to be reflected in the interpretation of the results obtained.

Third, information about the CZ-NACE sector is important for the following unrelated and related variety calculations. The main CZ-NACE code is used for the calculation of five and two places. This indicator is assigned by the Czech Statistical Office (CZSO) and is based on the largest reported sales volume of own sales of goods and services, change in inventories of own operations and capitalization. These three items are grouped in the CZSO accounts under one heading of output. The CZ-NACE classification is therefore not an answer to the general classification of an enterprise, but rather a description of its economic activity. For example, Honeywell Aerospace Olomouc s.r.o. has listed the main CZ-NACE 30900 equivalent to C30.9 – Manufacture of transport equipment n.e.c. but is considered for CZ-NACE 30300 with the equivalent of C30.3 – Manufacture of air and spacecraft and related machinery. As the CZ-NACE performance reporting methodology is uniform, it can be expected that similar nuances will be evenly spread across the national sample and thus partially cancel each other out. However, it is imperative that this shortcoming is taken into account when interpreting the results.

Fourth, the data used from IS VaVaI contain only R&D collaborations under direct public support (in the form of collaborative projects). Unfortunately, data for private R&D collaboration are not available and therefore the dataset used is not exhaustive. These limitations will be taken into account when interpreting the results.

4.3 Unrelated and Related Variety Calculation

In the first step, an unrelated diversity index was calculated using the provided formulas by Frenken et al. (2007). The following calculations, P_S represents the

share of employees in industry S (section) compared to the total number of employees Z in the territorial unit i in period t. For P_S NACE industries are used in two places.

$$P_s = (\frac{Z_S}{Z})$$

S takes the values of all five-digit NACE sector codes.

$$UVar_i = \sum_{s=1}^{S} P_s (\frac{1}{P_s})$$

UVar is the resulting value of unrelated variety for the geographical unit. The temporal aspect (the year to which the unrelated diversity value relates) is not considered in this view. The latest available employment data is used (the most common year is 2021).

In the next step, the related variety index was calculated according to the formulas below:

$$RVar_i = \sum_{s=1}^{S} P_s H_s$$

and

$$H_s = \sum_{d \in S_S} \frac{p_i}{P_s} (\frac{1}{\frac{p_i}{P_S}})$$

Where p_i represents the proportion of NACE employees per five locations relative to the total number of employees in a geographical unit. *RVar* equations are used for obtaining related variety indexes for regions, districts and SO ORP. P_S and H_S are calculated separately for each geographic level.

The subsequent results of related and unrelated diversity are comparable only at the same level of the territorial unit due to the nature of the calculation. Subsequently, the related and unrelated variety is calculated for three samples of enterprises according to the main NACE sector indicated: firstly, the calculation was carried out for the whole range of NACE sectors, from Section A – Agriculture, forestry, and fishing to Section S – Other activities. Furthermore, the calculation was performed on the NACE range falling within NACE section C – Manufacturing.

4.4 R&D cooperation

To measure the amount of R&D cooperation in a geographic context four indicators are calculated:
1) Internal R&D cooperation within the region.
2) External R&D cooperation outside of the region.
3) Internal R&D cooperation within the region taking into consideration only firms.
4) External R&D cooperation outside of the region taking into consideration only firms.

The focus on firms is done by subsetting the dataset only for firms. Cooperation between organizations

taken into account is within the years 2006–2021. Internal R&D cooperation is calculated as the number of two or more firms in the same region in a collaborative research project. External R&D cooperation is calculated as the number of projects with cooperation outside of the region. The point for the project is granted to all participating regions.

4.5 Cluster analysis

Before initiating the clustering process, the data was subject to a preliminary examination to ensure it was suitable for cluster analysis. Any missing values were addressed, and potential outliers were either rectified or justified. The variables were also normalized to ensure equal weightage during clustering. Normalization was achieved using the min-max scaling method, which transforms the data into a range between 0 and 1, ensuring that each variable contributes equally to the clustering process (Virmani, Taneja, Malhotra, 2015). The formula for min-max scaling is given by:

$$normalized(x) = \frac{x - \min(x)}{\max(x) - \min(x)}$$

In the article's cluster analysis methodology, three fundamental metrics are used for clustering: Joint state-supported projects, related variety, and unrelated variety. It was observed that there's a pronounced correlation between related variety, unrelated variety, and the number of project collaborations to the number of employees. Therefore, to avoid potential biases, these variables are adjusted by dividing them by the latter metric. Without such a modification, the clustering could inadvertently emphasize primarily the population size of SO ORPs, rather than the intended nuances of the regions.

To address the related and unrelated diversity bias mentioned by Bathelt and Storper (2023), the derived metrics are divided by a number of employees. This reduces the importance of large cities and towns in favor of microregions with higher related and unrelated diversity per employee.

Three primary metrics were chosen for clustering on the level of SO ORP:

$$\frac{joint\ projects_r}{number\ of\ employees_r} \tag{1}$$

$$\frac{related\ variety_r}{number\ of\ employees_r} \tag{2}$$

$$\frac{unrelated\ variety_r}{number\ of\ employees_r} \tag{3}$$

The underscored r signifies the Czech microregion (SO ORP). The appropriate number of clusters determined using the Elbow Method was 3. This involved running the k-means clustering on the dataset for a

Tab. 2 Regional unrelated and related variety for Czechia.

Region code	NUTS 3 region name	Unrelated variety	Related variety	Unrelated variety only NACE-C	Related variety only NACE-C
CZ010	Prague	2.028	0.535	0.674	0.245
CZ020	Central Bohemian Region	0.819	0.191	0.778	0.252
CZ031	South Bohemian Region	0.451	0.090	0.471	0.142
CZ032	Pilsen Region	0.461	0.090	0.508	0.145
CZ041	Karlovy Vary Region	0.183	0.027	0.181	0.040
CZ042	Ústí nad Labem Region	0.490	0.109	0.493	0.165
CZ051	Liberec Region	0.317	0.058	0.401	0.094
CZ052	Hradec Králové Region	0.404	0.081	0.454	0.137
CZ053	Pardubice Region	0.400	0.076	0.442	0.125
CZ063	Vysočina Region	0.354	0.075	0.453	0.137
CZ064	South Moravian Region	0.869	0.221	0.707	0.274
CZ071	Olomouc Region	0.400	0.082	0.443	0.141
CZ072	Zlín Region	0.492	0.121	0.539	0.201
CZ080	Moravian-Silesian Region	0.812	0.181	0.771	0.268

Source: Own calculations based on data drawn from RES and MAGNUS.

range of values of k (e.g., k from 1 to 10), and then for each value of k computing the sum of squared distances from each point to its assigned centre. The 'elbow' of the curve represents an optimal value for k (a balance between precision and computational cost) (MacQueen 1967). In the article, the final number of clusters is selected to be 5 to better represent the granularity of Czech microregions (SO ORP).

With the selected value of k, k-means clustering was applied to the dataset using the chosen metrics. The k-means algorithm seeks to minimize the squared sum of Euclidean distances from the mean of each cluster (Ismkhan 2017). The iterative algorithm divides the microregions into k clusters based on the similarity in their R&D collaborative dynamics. The k-means clustering was executed using the following command in R:

> kmeans(data_normalized, centers = k_optimal, nstart = 25)
> kmeans_result=kmeans(data_normalized, centers = k_optimal, nstart = 25)

The nstart parameter ensures that the algorithm is initialized multiple times to avoid local optima (Hartigan and Wong, 1979).

5. Initial results

The results of calculations in the previous part are structured as follows: First unrelated and related variety for NUTS 3 regions, districts, and microregions, and second internal and external cooperation for the same geographical units are presented.

5.1 Unrelated and Related Variety in NUTS 3 regions, districts, and microregions

The main NUTS 3 region that dominates in unrelated and related variety is Prague, the capital of Czechia. Because Prague is a capital city many firms have domicile there even though most of their employees and production is located elsewhere. The results demonstrate its greatest general diversity in the nomenclature of economic activities in both related and unrelated fields. The second place is usually occupied by the second largest city, Brno, and its surrounding NUTS 3 region, the South Moravian Region. Significant differences in ranking become apparent when evaluating related varieties and unrelated varieties for the NACE-C manufacturing only.

5.2 Regional unrelated and related variety

Tab. 2 shows the related and unrelated variety dominance of the three NUTS 3 regions where the three largest cities in Czechia are located: Prague, South Moravian Region, and Moravian-Silesian Region. Following this table further data are visualized as cartograms. The change in order can be seen when only manufacturing (NACE-C) is taken into consideration.

For related variety only in manufacturing (NACE-C), Prague lags behind the South Moravian Region, Moravian-Silesian Region, and even the Central Bohemian Region, but it is not surprising. These regions show higher nomenclature specialization. It shows that even the economically most important NUTS 3 region in Czechia (Prague) may be most diverse in terms of broad nomenclature (NACE) but not in terms of industry (Fig. 1).
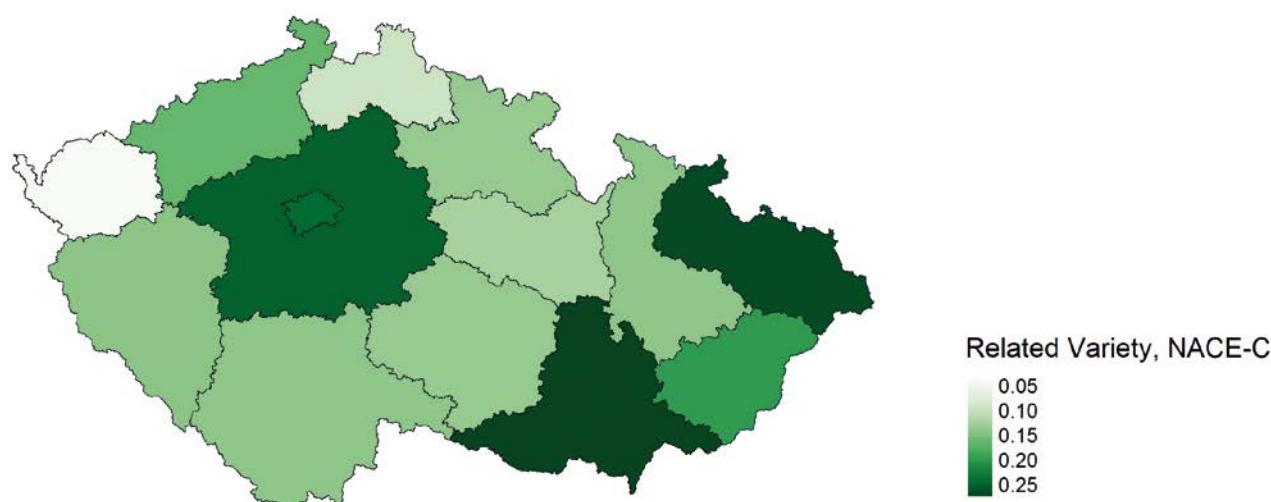
Fig. 1 Related variety only for manufacturing (NACE-C) in Czechia.

5.3 Unrelated and related variety in districts of Czechia

On the level of districts (okres) in Czechia, Prague shows the highest variety in all measured aspects. In the Tab. 3, the first 10 districts by related variety of manufacturing (NACE-C) are shown. The highest-scoring districts are cities and towns.

Importantly the industrial districts also show higher unrelated variety when only NACE-C is considered. The outlier that wasn't particularly visible in other measurements is the district of Mladá Boleslav which hosts large car manufacturing capacities. With a general unrelated variety of 0.171, the unrelated variety only for NACE-C is 0.245, contrary to this the related variety is 0.012 and if only manufacturing is taken into consideration it is 0.019 (Fig. 2).

5.4 Unrelated and related variety on the Czech microregional level

The lowest presented geographic level of Czechia in this paper is a municipality with extended powers (SO ORP). There are 206 such units in Czechia. The unrelated and related variety follows the expected trend. The capital Prague scores highest and Brno. Ostrava and Pilsen follow it. In unrelated variety of manufacturing Mladá Boleslav holds 2nd place. That doesn't correspond to its size in the population (19th). Fig. 3 also shows that in terms of industry, the spots of nearly zero related variety are not located along borders. This unexpected phenomenon may lead one to think about the location of the inner peripheries. Furthermore, strong SO ORPs in terms of related variety in manufacturing are often bordered by SO ORPs that

Tab. 3 Related and unrelated variety of Czech districts, first 10 ordered by Related variety NACE-C.

District code	District name	Unrelated variety	Related variety	Unrelated variety NACE-C	Related variety NACE-C
CZ0100	Praha	2.028	0.535	0.674	0.245
CZ0642	Brno-město	0.518	0.099	0.292	0.082
CZ0724	Zlín	0.222	0.044	0.229	0.069
CZ0806	Ostrava-město	0.363	0.055	0.238	0.049
CZ0534	Ústí nad Orlicí	0.132	0.020	0.197	0.045
CZ0323	Plzeň-město	0.242	0.036	0.236	0.044
CZ0804	Nový Jičín	0.136	0.021	0.215	0.043
CZ0723	Vsetín	0.142	0.021	0.185	0.041
CZ0643	Brno-venkov	0.152	0.025	0.174	0.038
CZ0722	Uherské Hradiště	0.137	0.022	0.162	0.037
CZ0207	Mladá Boleslav	0.171	0.012	0.245	0.019

Source: Own calculations based on data drawn from RES and MAGNUS.
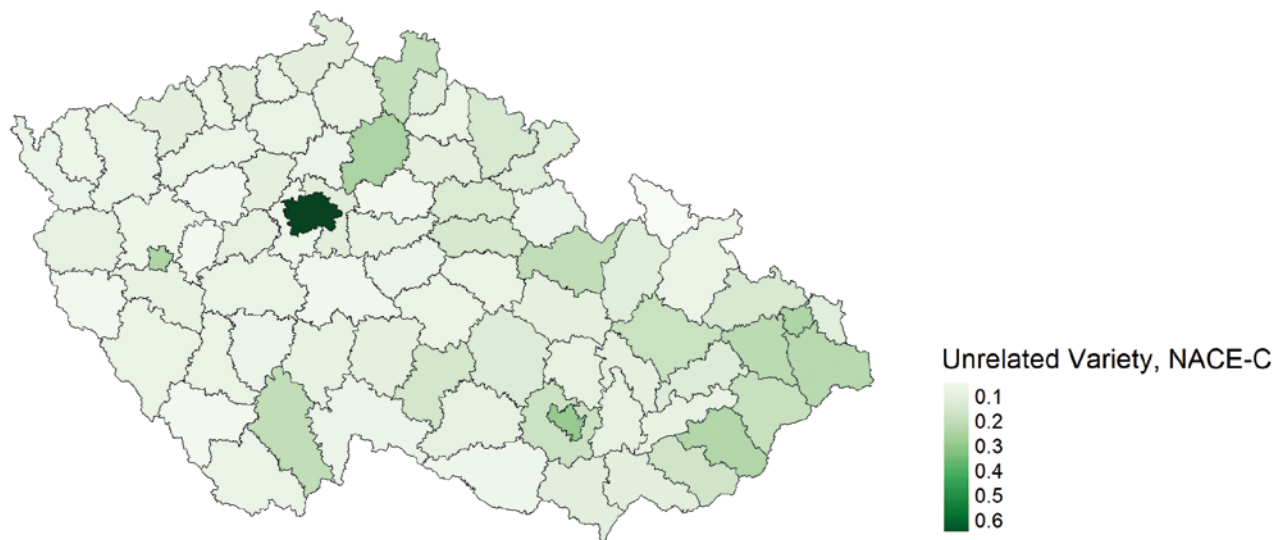
Fig. 2 Unrelated variety only for manufacturing (NACE-C).

have no related variety in manufacturing. This offers room for further research and discussion of inner peripheries.

5.5 R&D Cooperation in NUTS 3 regions, districts and municipalities

To measure R&D and collaboration in the NUTS 3 regions the CEP database of collaborative projects is used. The CEP data contains partially state-supported R&D projects. Some of them are joint projects of R&D collaboration. Project data for supported projects with collaboration that started in the years 2006–2022 are used. These are 12,577 unique projects with collaborations with a total of 3,852 unique organizations, including 3,006 unique firms. Together these projects represent support from the state budget of 155.65 billion CZK, which is 56.27% of all projects for the same period in CEP. It is 28.11% of the public budget dedicated to R&D in 2006–2022 and about 11.3% of total R&D expenditure in Czechia (GERD). Thus, the scope of the analysis of R&D cooperation is limited to this slice of approximately 11.3% of the total R&D expenses. In the article, R&D cooperation is examined using the CEP dataset and a pairwise collaboration methodology. This approach identifies every possible two-region combination in which organizations are jointly engaged in a project. Every organization, including each faculty as a distinct entity, is taken into account. Collaborations are classified into those occurring within the same region (intra-regional) and those bridging different regions (extra-regional). This
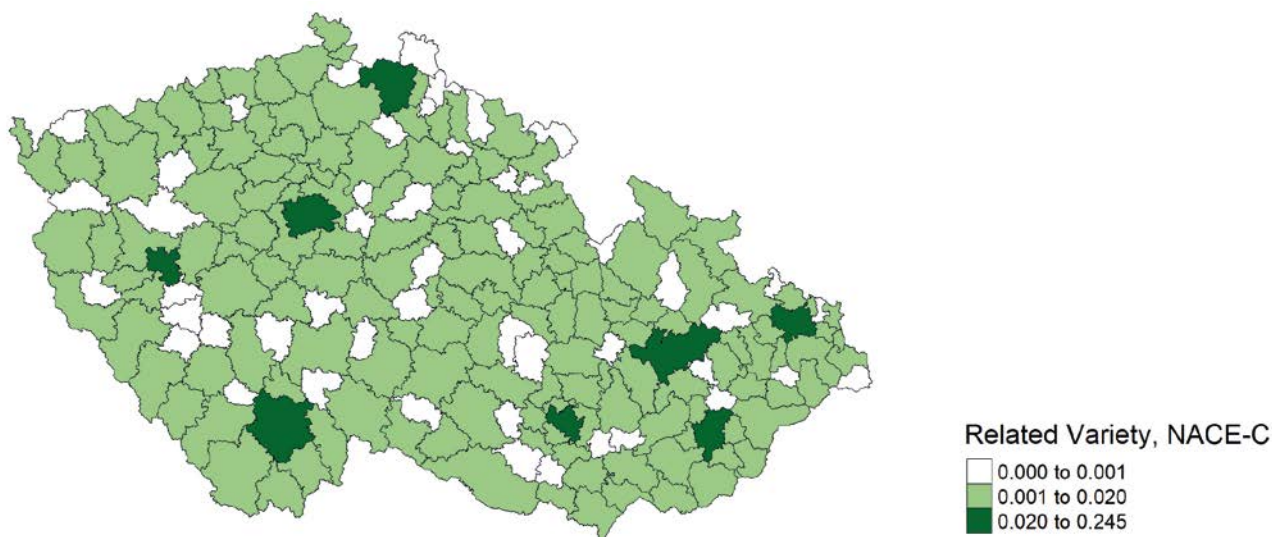

Fig. 3 Related variety only for manufacturing (NACE-C) in Czech municipalities with extended powers.
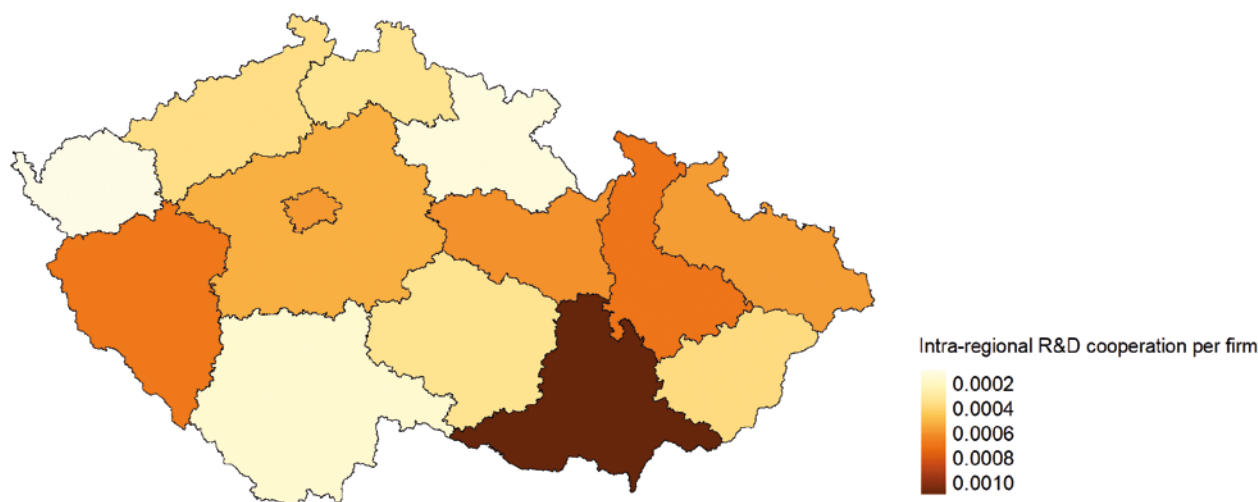
Fig. 4 R&D cooperation of all organizations.

method offers a thorough insight into the landscape of regional R&D collaborations. The values are calculated at three geographic levels: a) NUTS 3 region b) district c) Czech microregions (SO ORP). 75,774 participations with collaboration is observed. Of these, 28,474 (37.58%) are intra-regional and 47,300 (62.42%) are extra-regional. The collaborations that did not leave the Prague borders (Prague-Prague) are 23.46% of the direct collaborations. In the case of only firm-firm cooperations, we observe 15,200 direct links between firms within projects. Of these, 2,964 (19.50%) are intra-regional and 12,236 (80.50%) are extra-regional. The collaborations that did not leave the Prague (Prague-Prague) border account for 10.11% of the direct collaborations. Next, we conduct pairwise connections for organizations that are not in the set of firms and are in RVVI's list of research organizations. For these research organizations only, we observe 11,688 direct connections of which 5,138 (43.95%) are intra-regional and 6,550 (56.04%) extra-regional. Direct Prague-Prague connections accounted for 35.06% of the connections (Fig. 4).

6. Hypothesis testing

6.1 Hypothesis 1

To test the first hypothesis: "Based on Květoň and Horák (2018), who clarified the differentiation of R&D capacities at the regional NUTS 3 level in Czechia when subjected to k-means clustering based on relative joint R&D projects, related variety and unrelated variety, the SO ORP (Czech microregions) will yield more than two distinct clusters." cluster analysis was employed on the provided data. Cluster analysis groups data points into clusters so that data points in the same cluster are more similar to each other than to those in other clusters.

In the article, k-means clustering was applied to segment SO ORP regions using three primary metrics: shared state-funded projects, unrelated variety and related variety.

Given the significant variance between clusters in terms of the three metrics, it can be concluded that distinctive patterns indeed emerge among SO ORP regions when characterized by their R&D collaborative dynamics in terms of shared state-funded projects, related, and unrelated variety. The appropriate number of clusters determined using the k-means (Elbow Method) was 3. Therefore, based on this cluster analysis, hypothesis H1 fails to be rejected. The results indicate that distinctive patterns (clusters) emerge, aligning with the hypothesis's premise. The results indicate that distinctive patterns (clusters) emerge, aligning with the hypothesis's premise, with the chosen five clusters providing more granularity and detail in understanding the distinctive patterns in the R&D collaborative dynamics within the Czech microregions:

Cluster 1: Microregions classified under "Emerging Collaborators'' present a promising picture. They show a higher related variety over unrelated variety when metrics are divided by the number of employees. The average related variety divided by number of employees can compete with larger and sophisticated microregions such as Prague. These microregions are also involved in a commendable number of state-funded projects. They cater to a moderate population and have a substantial employment base, positioning them as areas that are budding and showing promise in their collaborative endeavors.

Cluster 2: "Collaboration Powerhouses". The microregions classified under this cluster are characterized by a harmonised interplay of related and unrelated variety metrics, especially when contextualised against the number of employees. What
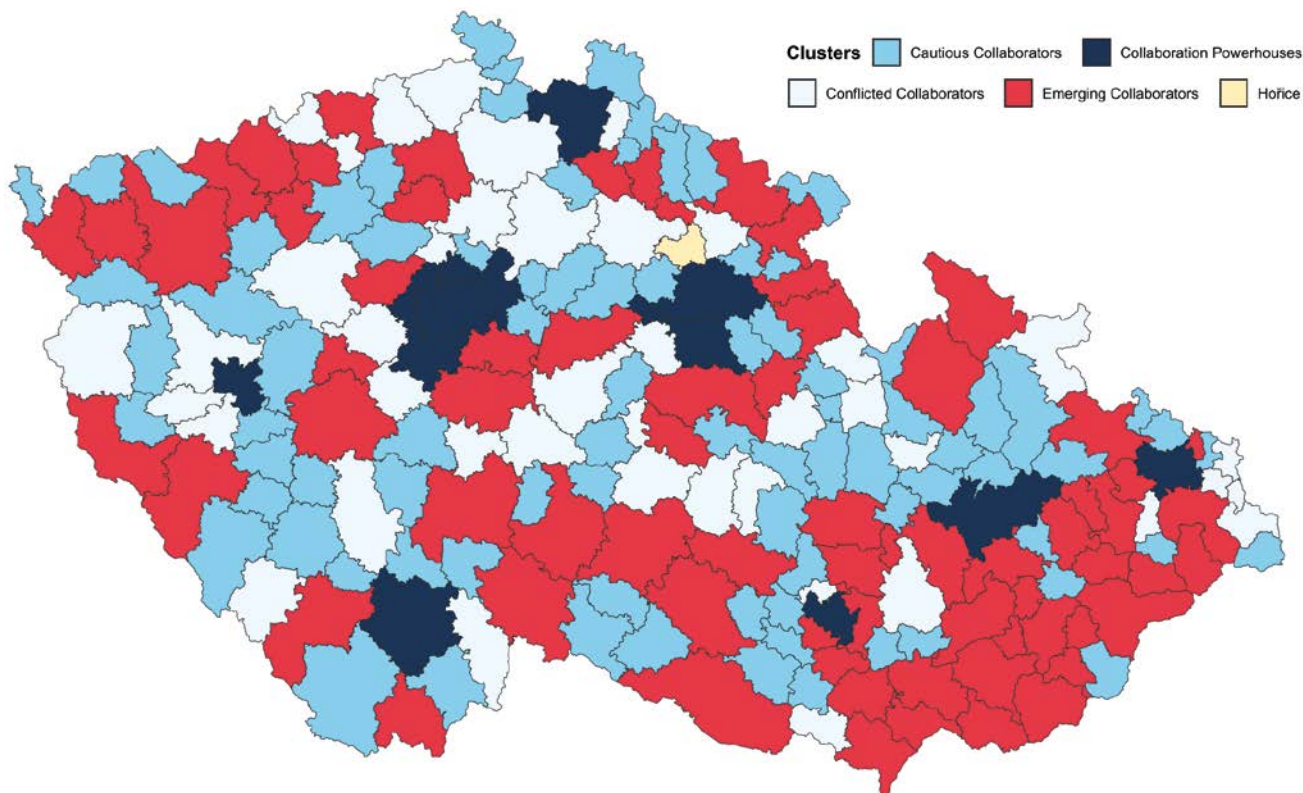
Fig. 5 Clustering of Czech microregions (SO ORP) based on related variety and direct state-supported R&D cooperation in Czechia.

sets them apart is the impressive level of collaboration per employee. With a pronounced related variety, these microregions have participated in a substantial number of state-funded projects. They hold a significant population and are supported by a large workforce. These microregions can aptly be described as the nexus of collaborative activities, making them true collaboration epicenters.

Cluster 3: "Hořice". This microregion did fall into one unique cluster as there is a research organization "Výzkumný a šlechtitelský ústav ovocnářský Holovousy s.r.o." that was supported by 128 projects and within this project created 218 extra-regional and 26 intra-regional connections.

Cluster 4: "Cautious Collaborators". Microregions in this cluster are low in related variety and collaborations, particularly when metrics are proportioned against the number of employees. Their absolute related variety and involvement in state-funded projects are also very low. The SO ORPs are inhabited by a smaller population and supported by a moderate workforce. They have established a modest footprint in the collaboration arena and have room to explore further synergies.

Cluster 5: "Conflicted Collaborators". Microregions in this cluster present an intriguing dichotomy. Despite their relatively low related variety when adjusted for the number of employees, they exhibit a heightened engagement in state-funded R&D projects. This suggests a distinct focus on select,

specialized areas of expertise or perhaps a concentration of knowledge within certain domains. The participation rate in collaborations remains consistently high, indicating an active pursuit of partnerships and shared initiatives. The SO ORPs in question have a moderate population. These microregions are navigating a path that, while conflicted between specialized knowledge and broad collaboration, holds the potential for unique growth trajectories (Fig. 5).

6.2 Hypothesis 2

"Based on the current state of knowledge about related variety in different countries (Wise and Anderson 2017, Boschma and Iammarino 2009) we expect that the intensity of R&D collaboration in state-supported projects will be positively associated with the Related Variety."

In the article, a linear regression analysis was conducted to examine the relationship between the intensity of R&D collaborations and the related variety among SO ORP. The linear model results suggest a strong positive association between related variety and the intensity of R&D collaborations. In contrast, unrelated variety exhibited a significant negative relationship with R&D collaboration intensity. Additionally, certain clusters, population density, and the total number of employees in a microregion also influenced R&D collaborations, though not all were

Tab. 4 Linear regression of R&D collaboration and related variety.

Variable	Estimate	Standard Error	p-value	Significance
(Intercept)	198.900	51.78160	0.00017	***
Related variety	98338.300	11719.52476	0	***
Unrelated variety	−17679.100	1619.86162	0	***
Employees	0.029	0.00784	0.00031	***
cluster 1	−139.200	52.73475	0.00897	**
cluster 2	503.700	127.00239	0.00010	***
cluster 3	90.100	233.80022	0.70026	
cluster 4	−104.100	49.48022	0.03673	*
popdens	0.130	0.13156	0.31688	

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
Residual standard error: 229.3 on 197 degrees of freedom
Multiple R-squared: 0.9955, Adjusted R-squared: 0.9953
F-statistic: 5407 on 8 and 197 DF, p-value: < 2.2e−16
Source: Author.

Tab. 5 Linear regression of R&D collaboration and related variety in manufacturing.

Variable	Estimate	Standard Error	p-value	Significance
(Intercept)	16.5000	4.750	0.00064	***
Related variety NACE-C	5171.2000	248.200	< 2e−16	***
Unrelated variety NACE-C	−1097.5000	185.500	1.43e−08	***
Employees in NACE-C	0.0055	0.001	0.00001	***
cluster1	−13.7000	3.900	0.00062	***
cluster2	−22.2000	8.700	0.01144	*
cluster3	−11.9000	19.200	0.53702	
cluster4	−6.6000	4.100	0.10548	
popdens	−0.0180	0.010	0.07771	.

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
Residual standard error: 18.79 on 197 degrees of freedom
Multiple R-squared: 0.9414, Adjusted R-squared: 0.939
F-statistic: 395.5 on 8 and 197 DF, p-value: < 2.2e−16
Source: Author.

statistically significant. The model captures approximately 99.55% of the variation in total cooperations. Given these findings, the article concludes that the data supports the hypothesis, emphasizing the role of related industries in fostering R&D collaborative dynamics across microregions. Based on these results, the article fails to reject hypothesis 2, affirming that microregions with higher related variety tend to have intensified R&D collaborations (Tab.4).

Multiple Linear Regression Equation 1:
R&D pairwise cooperation
$= \beta_0 + \beta_1 RV + \beta_2 UV + \beta_3 EMPL + \beta_4 C1 + \beta_5 C2 + \beta_6 C3 + \beta_7 C4 + \beta_8 C5 + \beta_9 Popdens + \in$

6.3 Hypothesis 3

"The intensity of R&D collaboration of state-supported projects in the manufacturing sector (NACE-C) is positively associated with the related variety specific to manufacturing."

In the article, a linear model is employed to test the second hypothesis that investigates the relationship between the intensity of R&D collaboration in manufacturing (NACE-C) and the related variety specific to manufacturing. This model takes into account various control variables, integrating factors such as unrelated variety, clusters, population density, and the total number of employees in the manufacturing sector. Through linear regression, the model provides a robust statistical framework to determine how the diversity of manufacturing activities, both related and unrelated, along with microregional characteristics, influences collaborative R&D efforts in the sector (Tab. 5).

Multiple Linear Regression Equation 2:
R&D pairwise cooperation NACE-C
$= \beta_0 + \beta_1 RV_{NACE-C} + \beta_2 UV_{NACE-C} + \beta_3 EMPL_{NACE-C} + \beta_4 C1 + \beta_5 C2 + \beta_6 C3 + \beta_7 C4 + \beta_8 C5 + \beta_9 Popdens + \in$

This assertion is underlined by the positive and highly significant coefficient for the variable related variety for NACE-C (manufacturing). As the related variety specific to manufacturing increases, the intensity of R&D collaboration in the sector also witnesses a marked increase.

Notably, while the related variety presents a positive relationship with collaboration intensity, the unrelated variety for NACE-C displays a negative and significant relationship. This suggests that a higher unrelated variety in the manufacturing sector could act as a detriment to the intensity of R&D collaborations. The observed negative association might indicate that when activities are too diversified or unrelated in a microregion, it becomes challenging to find common ground or mutual benefits, thereby reducing collaborative endeavors even in partially state-funded R&D projects.

Considering control variables, it's evident that certain clusters, notably cluster1 (Emerging Collaborator) and cluster2 (Collaboration Powerhouses), exhibit a negative statistical relationship with R&D collaborations in manufacturing. This suggests that SO ORPs belonging to these clusters might have some inherent characteristics or challenges impeding collaboration. Conversely, the total number of employees in the NACE-C shows a positive and significant relationship with collaboration intensity, pointing to the fact that microregions with a larger workforce in manufacturing have heightened collaborative activities.

Lastly, the population density presents a marginally negative influence on collaborations. This might imply that in densely populated areas, the nature of industrial activities could be more fragmented or diverse, possibly diluting the intensity of focused R&D collaborations in manufacturing.

7. Discussion of empirical findings

In this section, the article delves into the influence of related variety on regional development, particularly within the context of collaboration in state-supported R&D projects. This analysis is contrasted with established research, highlighting both similarities and differences in approach and findings. Notably, Frenken et al. (2007), Boschma and Iammarino (2009), and Boschma, Minondo, and Navarro (2013) all underscored that regions with a pronounced related variety tend to witness enhanced employment growth. This observation, while aligning with the broader theme of this article, diverges in its primary approach and objectives. While these studies primarily focused on employment growth as a direct outcome of related variety, this article pivots towards understanding the dynamics of collaboration within the context of related variety. In this topic, Ebersberger, Herstad, and Koller (2014) further explored the connection between regional knowledge bases, collaborations, regional technological specialization and related variety. The specialization reduced domestic collaborations, while related technological variety bolstered international innovation ties.

The sector-specific effects of related variety, as highlighted by Bishop and Gripaios (2010) and Hartog, Boschma, and Sotarauta (2012), offer another dimension of comparison. These studies suggest that the influence of related variety can vary significantly across different sectors. In the context of this article, the use of NACE-C (manufacturing) classifications provides a broader lens to dissect the specific and industrial part of the economy. Bishop and Gripaios (2010) emphasized the potential oversimplification of broadly categorizing sectors into manufacturing and services. They argue that these sectors, in their essence, are heterogeneous, leading to varied mechanisms and extents of spillovers between them. Driven by this perspective, they employed a more granular approach, examining employment growth in individual 2-digit sectors. Given this critique by Bishop and Gripaios, it seems prudent for future research to delve deeper into the manufacturing NACE-C classification, breaking it down further into specific 2-digit sectors for a more nuanced understanding.

The results of the hypothesis testing, when viewed through the lens of economic geography and regional development literature, offer a nuanced understanding of R&D collaborative dynamics in SO ORP microregions. The summarising work by Content and Frenken (2016) underscores the importance of related variety in economic development such as employment growth. Their comprehensive literature review suggests that regions with a diverse yet related set of industries tend to exhibit higher levels of innovation (observed through labour productivity) and economic growth. The critique by Bathelt and Storper (2023) on measuring related and unrelated variety as entropy which leads to a strong statistical link between related variety, unrelated variety, and the population of a (micro)region, is addressed by controlling for these factors and using the number of employees as a denominator.

On the other note, the observed negative relationship between unrelated variety and R&D collaboration intensity in our results resonates with the foundational principles of the related variety literature. As highlighted by Marek and Blažek (2016) and Květoň, Novotný, Blažek, Marek (2022), a region with technologically related industries often benefits from enhanced knowledge spillovers, learning, and growth. However, when activities become overly diversified or unrelated, it can pose challenges in finding synergies, potentially reducing collaborative endeavors. This perspective aligns with the argument that spatial externalities are most potent among firms with related but distinct knowledge. Yet, it's crucial to consider the insights from Grillitsch et al. (2018), who emphasize the potential of unrelated diversification in fostering radical innovations, especially in regions with strong human capital.

In light of the findings from the article's hypothesis testing and the insights from Květoň et al. (2022), it becomes evident that the dynamics of R&D collaboration in SO ORP microregions are multifaceted. The observed negative relationship between unrelated variety and R&D collaboration intensity underscores the challenges of excessive diversification in hindering synergies. Květoň et al. (2022) further illuminate this by revealing that while R&D collaborations often span large cognitive distances, they are not arbitrary. Firms, in their pursuit of innovation, tend to collaborate with partners that, although unrelated, share closer cognitive proximity than other potential collaborators. This intricate balance between diversification and the quest for synergies is further complicated by the geographical dynamics, as seen in the predominant inter-regional linkages in Czech regional innovation systems.

8. Conclusions

In this paper, the concepts of unrelated variety and related variety were introduced and empirically analyzed, first separately and then for manufacturing (NACE-C). Furthermore, concepts of intra-regional and extra-regional R&D cooperation within state-supported joint projects were introduced and assessed.

They are divided into R&D cooperation of firms only, research organizations only and then all pairwise connections for all types of organizations. The mentioned indicators were calculated at three geographical levels, NUTS 3 regional, district, and microregional.

The indicators yield the following results: Prague and the South Moravian Region ranked first in unrelated and related variety. In the case of the related variety of the NUTS 3 region in industry (NACE-C) Prague drops out of the first place. Analysis of unrelated variety at the district level for industry (NACE-C) shows that districts with a large industrial presence tend to have higher unrelated variety rather than related variety. Microregional level analysis shows that municipalities with the highest related variety for industry (NACE-C) are often adjacent to municipalities with almost zero related variety. Such results can be further explored in detail and built on the research of the inner peripheries of Czechia. In the case of the number of collaborations counted by joint R&D projects, it is evident that although Prague has the highest number of absolute collaborations, in a relative view (number of collaborations divided by the number of companies in the region) Prague falls into the background and the highest values are reported by the Pilsen Region, South Moravian Region and Pardubice Region.

A final comparison of these indicators with each other at the regional level shows a high degree of correlation between R&D cooperation and the general unrelated and related variety, but lower only in the case of the unrelated and related variety for industry (NACE-C). The research question of the article is elaborated in three hypotheses. The first hypothesis posited that SO ORP regions, when characterized by their R&D collaborative dynamics in terms of shared state-funded projects and related and unrelated variety, would show distinctive patterns. To investigate this, k-means clustering was applied to group the SO ORP regions based on shared state-funded projects and related variety. After identifying five distinct clusters, the analysis revealed considerable differences between these clusters based on the given metrics. Consequently, the hypothesis was not rejected, suggesting that distinctive patterns were indeed evident among the regions.

Among the clusters identified, a standout group was labeled "Collaboration Powerhouses". These SO ORP microregions were distinctive due to their dominant related variety, especially when adjusted for the number of employees. Their collaboration intensity per employee was also noteworthy. They possess a significant population and are backed by a substantial workforce, these microregions stood out as central hubs of collaborative activities, solidifying their reputation as true epicenters of collaboration. Next "Emerging Collaborators" microregions showed promise with a commendable number of state-funded projects and a higher related variety, hinting at their

potential growth. In contrast, "Hořice" was a singular microregion due to a unique research organization significantly supported by 128 R&D state-supported projects. The "Cautious Collaborators" with both related variety and collaborations being modest, suggesting they might be lacking their footing in the R&D landscape. Lastly, "Conflicted Collaborators" displayed an interesting dichotomy, showing potential for unique growth trajectories while navigating a balance between specialized knowledge and broad collaboration.

The second hypothesis suggested a positive association between the intensity of R&D collaboration in state-supported projects and the related variety. A linear regression analysis was conducted to explore this relationship. The outcome showcased a robust positive correlation between related variety and the intensity of R&D collaborations. Conversely, unrelated variety had a significant negative relationship with R&D collaboration intensity. Some other variables, like certain clusters, population density, and the total number of employees in a microregion, also influenced the collaborations, though not all significantly. Overall, the data provided strong support for the hypothesis.

For the third hypothesis, it was proposed that the intensity of R&D collaboration in state-supported projects in the manufacturing sector (NACE-C) would be positively correlated with the related variety specific to manufacturing. A linear regression model was employed, factoring in several control variables. The results underscored a significant positive relationship between the related variety in manufacturing and the intensity of R&D collaborations. However, a notable discovery was that a higher unrelated variety in manufacturing is negatively associated with the intensity of R&D collaborations.

Above all, the study identifies clear clusters within the SO ORP microregions based on their collaborative tendencies and sector closeness. Furthermore, a distinct positive relationship emerges between the intensity of related variety, also in the manufacturing sector, and the extent of cooperation within state-supported joint projects. Interestingly the unrelated variety relates to the extent of cooperation negatively.

References

Bathelt, H., Storper, M. (2023): Related Variety and Regional Development: A Critique, Economic Geography, https://doi.org/10.1080/00130095.2023.2235050.

Bishop, P., Gripaios, P. (2010): Spatial Externalities, Relatedness and Sector Employment Growth in Great Britain. Regional Studies 44(4), 443–454, https://doi.org/10.1080/00343400802508810.

Blažek, J., Marek, D., Květoň, V. (2016): The Variety of Related Variety Studies: Opening the Black Box of Technological Relatedness via Analysis of Inter-firm R&D

Cooperative Projects. Papers in Evolutionary Economic Geography (online), http://econ.geo.uu.nl/peeg /peeg1611.pdf.

Bond-Smith, S., McCann, P. (2019): A Multi-sector Model of Relatedness, Growth and Industry Clustering. Journal of Economic Geography 5(20), 1145–1163, https://doi.org /10.1093/jeg/lbz031.

Boschma, R., Iammarino, S. (2009): Related Variety, Trade Linkages, and Regional Growth in Italy. Economic Geography 85(3), 289–311, https://doi.org/10.1111 /j.1944-8287.2009.01034.x.

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. Economic Geography 89(1), 29–51, https:// doi.org/10.1111/j.1944-8287.2012. 01170.x.

Castaldi, C., Frenken, K., Los, B. (2015): Related Variety, Unrelated Variety and Technological Breakthroughs: An analysis of US State-Level Patenting. Regional Studies 49(5), 767–781, https://doi.org/10.1080/00343404 .2014.940305.

Corradini, C., Vanino, C. (2022): Path dependency, regional variety and the dynamics of new firm creation in rooted and pioneering industries, Journal of Economic Geography 22(3), 631–651, https://doi.org/10.1093 /jeg/lbab021.

Ebersberger, B., Herstad, S. J., Koller, C. (2014): Does the composition of regional knowledge bases influence extra-regional collaboration for innovation? Applied Economics Letters 21(3), 201–204, https://doi.org /10.1080/13504851.2013.848019.

Frenken, K., Boschma, R. (2007): A theoretical framework for Evolutionary Economic Geography: Industrial dynamics and urban growth as a branching process. Journal of Economic Geography 7(5), 635–649, https://doi.org/10.1093/jeg/lbm018.

Frenken, K., van Oort, F., Verburg, T. (2007): Related variety, unrelated variety and regional economic growth. Regional Studies 41(5), 685–697, https://doi.org /10.1080/00343400601120296.

Grillitsch, M., Asheim, B., Trippl, M. (2018): Unrelated knowledge combinations: the unexplored potential for regional industrial path development. Cambridge Journal of Regions, Economy and Society 11(2), 257–274, https://doi.org/10.1093/cjres/rsy012.

Hartigan, J. A., Wong, M. A. (1979): Algorithm AS 136: A K-Means Clustering Algorithm. Journal of the Royal Statistical Society. Series C (Applied Statistics) 28(1), 100–108, https://doi.org/10.2307/2346830.

Hartog, M., Boschma, R., Sotarauta, M. (2012): The Impact of Related Variety on Regional Employment Growth in Finland 1993–2006: High-Tech versus Medium/ Low-Tech. Industry and Innovation 19 (6), 459–476, https://doi.org/10.1080/13662716.2012.718874.

Ismkhan, H. (2017): An initialization method for the k-means using the concept of useful nearest centers. Journal of Advanced Engineering Research and Science, https://doi.org/10.48550/arXiv.1705.03613.

Květoň, V., Novnotný, J., Blažek, J., Marek, D. (2022): The role of geographic and cognitive proximity in knowledge networks: The case of joint R&D projects. Papers in Regional Science 101(2), 351–372, https://doi.org /10.1111/pirs.12656.

Květoň, V., Šafr, K. (2019): Regional embeddedness, relatedness and inter-regional linkages among less developed regions in Central Europe. European Planning Studies 27(5), 862–884, https://doi.org/10.1080 /09654313.2019.1576591.

MacQueen, J. (1967): Some methods for classification and analysis of multivariate observations. In Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability, eds L. M. Le Cam, J. Neyman 1, 281–297. Berkeley, CA: University of California Press.

Nooteboom, B. (2000): Learning by Interaction: Absorptive Capacity, Cognitive Distance and Governance. Journal of Management and Governance 4, 69–92, https://doi.org /10.1023/A:1009941416749.

Virmani, D., Taneja, S., Malhotra, G. (2015): Normalization based K means Clustering Algorithm. Journal of Advanced Engineering Research and Science, https:// doi.org/10.48550/arXiv.1503.00900.

Yeon, J., Jun, B. (2022). The spillover effect of neighboring port on regional industrial diversification and regional economic resilience. General Economics, https://doi.org /10.48550/arxiv.2204.00189.

Yeung, H. (2020): Regional worlds: from related variety in regional diversification to strategic coupling in global production networks. Regional Studies 55(6), 989–1010, https://doi.org/10.1080/00343404.2020 .1857719.