AGILE: GIScience Series, 6, 29, 2025. https://doi.org/10.5194/agile-giss-6-29-2025
Proceedings of the 28th AGILE Conference on Geographic Information Science, 10–13 June 2025.
Eds.: Auriol Degbelo, Serena Coetzee, Carsten Keßler, Monika Sester, Sabine Timpf, Lars Bernard
This contribution underwent peer review based on a full paper submission.

Uncovering Regional Structures and Collaboration Hubs in Baden-Württemberg through Clustering a Company Hyperlink Network

Umut Nefta Kanilmaz [1,*], Thomas Schneidergruber [1,*], Sebastian Schmidt [1,2], Robert Dehghan [2,3], and Johannes Scholz [1]

[1]Department of Geoinformatics—Z GIS, University of Salzburg, Salzburg, Austria
[2]ISTARI.AI, Mannheim, Germany
[3]Institute for SME Research and Entrepreneurship, University of Mannheim, Mannheim, Germany
[*]These authors contributed equally to this work.

Correspondence: Umut Nefta Kanilmaz (umutnefta.kanilmaz@plus.ac.at)

Abstract.

The hyperlink reference networks of company websites offer a promising approach to modeling inter-firm collaboration and provide valuable insights into e.g. technology diffusion. However, most existing studies analyze these networks without considering the geographic embedding of company locations, disregarding potentially significant spatial factors. To address this gap, we investigate spatial patterns and structures in collaboration networks within the German federal state of Baden-Württemberg. Comparing results from network community detection and unsupervised clustering approaches, we examine firstly, how the collaboration networks are structured in geographic space and secondly, assess whether companies with similar characteristics are also geographically close. Our findings reveal three distinct classes, namely regional, cross-regional, and hub-centered spatial network embeddings, even though a clustering of node attributes did not reveal substantial spatial similarities. These results highlight the important interaction between the virtual hyperlink references of companies and their embedding in geographic space.

Submission Type. Analysis.

BoK Concepts. [AM11-6] Other classic network problems [IP3-4-10], Classification features and feature space [GD] Geospatial Data.

Keywords. Spatial Clustering, Network Communities, Hyperlink Network, Company Attributes.

1 Introduction

Inter-firm collaboration has been identified as a key driver of regional innovation (Hervás-Oliver et al., 2021). Since traditional innovation research primarily relies on the analysis of patents or surveys, insights about such collaborations are often limited to rather small firm samples (Kinne and Axenbeck, 2020). An alternative is the analysis of firm activities through corporate websites. While the textual content reflects a firm's technological expertise, a hyperlink between two firm websites suggests ongoing collaboration, based on the assumption that hyperlinks are deliberately set (De Maeyer, 2013). While individual hyperlinks may have been created for reasons unrelated to collaboration, the overall hyperlink network can thus serve as a proxy for inter-firm relationships, reflecting broader social and cultural structures (Halavais, 2008). As many companies maintain a website, this approach enables large-scale analyses of collaboration between companies (Bailey et al., 2018). Most companies act at a geographic location indicated by their firm address. Assuming collaboration does not only occur in virtual spaces through internet technologies, the hyperlink network also provides a way to analyze inter-firm collaboration and knowledge flow between geographic regions (Abbasiharofteh et al., 2023). Most studies focus on the virtual hyperlink network structure and do not explore the geographical environment nor the geographic embedding of collaboration networks in conjunction with the virtual network structure. Thus, we identified a research gap that we address with this research work.

Therefore, this study uses hyperlink network data of firms in the German state of Baden-Württemberg (BW) to identify how inter-firm collaboration is structured in geo-

graphic space. BW was chosen due to its multicenter structure, long-term economic growth (Glückler et al., 2020), high-tech clusters (Schlossstein and Yun, 2008), and inherent diversity regarding both industrial sectors and population density (Hoffmann et al., 2024). We intentionally chose a state-level rather than a nation-level study area: The firm-level network of Germany is dominated by companies in urban centers with many companies that have an extensive collaboration networks (Schmidt et al., 2025). These firms exert a strong "pull" on the network, therefore overshadowing interesting regional structures.

Our study addresses the following research questions (RQ):

- **RQ1**: How does the hyperlink network of firms in BW manifest in geographic space?
- **RQ2**: Based on firm characteristics derived from the company website text, how do similar companies cluster in geographic spaces?

For RQ1, we hypothesize that the spatial proximity between firms dominates the network structures, rather than trans-regional collaborations facilitated by virtual communication technologies. For RQ2, we expect to find spatial clusters that correspond to technology clusters inspired by Markusen (1996).

To answer these research questions, we modeled the hyperlink network between firms in BW and additionally retrieved information derived from website text characterizing each firm. For RQ1, we intend to detect clusters in network structure only, i.e. disregarding the firm-level information. For RQ2, the network structure is disregarded and instead, unsupervised Machine Learning (ML) algorithms are trained on the firm data to detect structures based on the firm's similarities. Finally, both clustering results are spatially visualized, analyzed for spatial patterns, and compared to understand the underlying collaboration patterns.

Understanding the spatial clustering of point features in geographic space typically involves using measures of spatial autocorrelation (Boots and Tiefelsdorf, 2000) (Getis and Ord, 1992). These methods conceptualize spatial neighborhoods through a spatial weight matrix and identify local clusters of specific features. They are, however, less suitable for analyzing spatial clusters in higher-dimensional data. We intend to answer RQ2 by identifying clusters based on features derived from firm websites and assessing their spatial embedding. Unlike the mentioned traditional methods, our approach leverages the capabilities of ML algorithms and does not require the conceptualization of a predefined geographic neighborhood. Similarly, we analyze the company network for clusters independently of any geographical method and expect the emergence of spatial structures. This approach has been applied to social networks, where studies have demonstrated that virtual communication networks exhibit structures aligned with spatial boundaries (Arthur and Williams, 2019), (Ratti et al., 2010), (Scellato et al., 2011).

2 Data Retrieval

We retrieved all firms in BW from the ORBIS database (Bureau van Dijk, 2023), which contains information about the firm's domain, founding year, the number of employees and its address. We geocoded the latter using the Nominatim API. Each corporate website was then scraped with a depth of 25 subpages in April 2023, following the general workflow introduced by Kinne and Lenz (2021). Based on the extracted texts, we calculated various firm-level indicators, representing the relative importance of each topic for the respective firm. For this, we applied classification models on innovation, sustainability, artificial intelligence, and 3D printing (see table 1).

Furthermore, hyperlinks were extracted from the HTML content to analyze connections to other corporate websites in BW, following Abbasiharofteh et al. (2023). Based on these hyperlinks, a collaboration network of companies was modeled: Each company was represented as a network node with a unique company identifier, while a hyperlink from company A to company B was modeled as a directed edge between their respective nodes. Additionally, we removed 152 company nodes with very high number of in- or outgoing links, which had little informative value in the context of collaboration analyses such as domain service providers. For the node attribute clustering described in 3, all 147,269 company nodes were analyzed. The clustering of the network considered only the fully connected graph, i.e. network isolates with no other hyperlink reference were removed. Accordingly, the resulting company collaboration network in BW contained 63,542 nodes and 142,581 hyperlink edges.

3 Methods

The overall goal of this research is to study the spatial manifestation of **a) firm clusters due to the hyperlink network structure**, disregarding the node attributes, and **b) clusters due to companies attribute similarity**, disregarding the network structure.

For case a), we employed the Louvain community detection method (Blondel et al., 2008). This algorithm clusters the network into components by optimizing the modularity score, i.e. nodes in the identified component have stronger internal connections than external ones. As the Louvain algorithm is not deterministic, we used a sampling method and ran the community detection $n = 100$ times. For each run, we calculated the modularity score of the communities and selected the run with the highest modularity score. Since each company node has a position in geographic space, the resulting network clusters can be mapped and visually examined for geographic patterns.

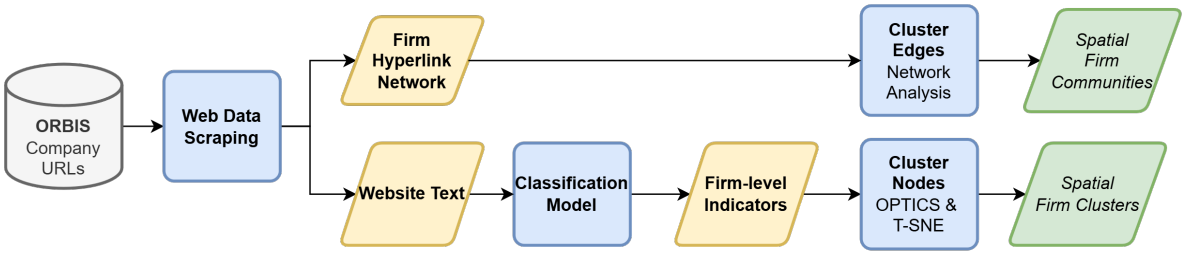

Figure 1. The workflow followed in this analysis.

Table 1. Firm-level indicators used as node-level features. The intensities and *innoprob* were calculated based on the website texts.

Variable	Description	Source	Mean	Std	Min	Max
employees	Number of employees	Bureau van Dijk (2023)	70.83	2,333.3	1	402,614
age_years	Firm age (in years)	Bureau van Dijk (2023)	27.9	26.6	1	749
sustainability_intensity	Relative importance of sustainability-related topics	Schmidt et al. (2022)	0.09	0.32	0	5.82
ai_intensity	Relative importance of artificial intelligence	Dahlke et al. (2024)	0.009	0.08	0	2.96
3d_printing_intensity	Relative importance of 3D printing	Schwierzy et al. (2022)	0.005	0.07	0	4.21
innoprob	Predicted innovativeness of a company	Kinne and Lenz (2021)	0.3	0.19	0.03	0.93
indegree	Number of links by other firms	Website text	0.96	8.23	0	2,069
outdegree	Number of links to other firms	Website text	0.96	4.85	0	1,376

For case b), we clustered the company-related features listed in table 1 with the density-based clustering approach OPTICS (Ordering Points To Identify the Clustering Structure) (Ankerst et al., 1999), an extension of DBSCAN (Ester et al., 1996). OPTICS creates an ordered representation of data points based on their density connectivity. Two key concepts are the *reachability distance*, which identifies how close points are to each other in terms of density, and the *core distance*, which determines the minimum distance for a point to be considered a core point. Through analyzing these distances, a reachability plot is generated. This plot visually represents the structure of the data and reveals clusters. We used OPTICS and the reachability outcome to identify clusters for the node features listed in table 1.

To further improve the results of the node-level clustering, we performed a dimensionality reduction from $d_{in} = 8$ to $d_{out} = 2$ using the t-distributed Stochastic Neighbor Embedding (t-SNE) algorithm (Van der Maaten and Hinton, 2008). The t-SNE algorithm begins by computing pairwise similarities in the high-dimensional space using a Gaussian distribution. In the lower dimensional space, a Student's t-distribution is utilized to model similarities. The distribution's shape leads to more distinct clusters, pushing dissimilar points apart and grouping similar points closer together. The alignment between the high-dimensional and low-dimensional similarity distributions is quantified using the Kullback-Leibler (KL) divergence (Kullback and Leibler, 1951) as the objective function, which is then iteratively minimized through a gradient descent optimization process.

Both OPTICS and t-SNE require hyperparameter optimization to achieve optimal performance. For OPTICS, we evaluated the *min_points* parameter within a range of 50 to 450 in increments of 100. For the *metric* parameter, we used the *cosine* and *Minkowski* distance with $p = 2$, as preliminary experiments with $p = 1$ did not produce promising results. For t-SNE, we optimized *perplexity*, the maximum number of iterations, and the distance metric. The *perplexity* parameter was selected in the range 80 to 200 in increments of 20, while the maximum number of iterations was chosen from 1,000, 1,500, and 2,000. We again considered the *cosine* and *Minkowski* distance with $p = 2$. The choice of *perplexity* range was guided by the need to balance local and global structure preservation, as lower values focus on fine-grained local relationships, while higher values capture broader global patterns.

4 Results

The dataset utilized in this work consists of 147,260 unique firm data points and their features are analyzed in the following section with Figure 2 indicating their correlation. Notably, the age of the company does only correlate very slightly with the innovativeness of a company. The other noteworthy interaction is the weak correlation between the innovativeness and the features describing sustainability, as well as AI.

Table 1 provides statistical key information about the distribution of the entire study data, i.e., including isolates. The standard deviation (2342.34) of the variable *employ-*

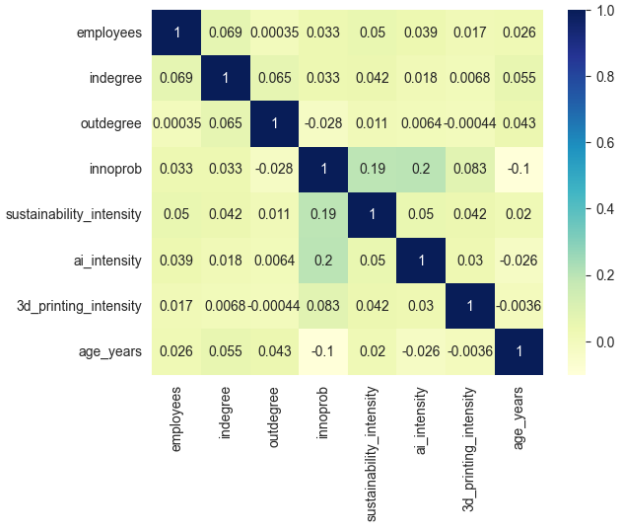

Figure 2. Correlations of the features in this dataset.

ees suggests high variability in the employee count of companies. The mean *indegree* is 0.96, with a standard deviation of 8.23. While most firms have low incoming links, some experience high connectivity, possibly indicating hubs in a network or highly interconnected companies. The same conclusion can be drawn for the variable *outdegree*. The values for the variables *innoprob*, *sustainability intensity*, *ai intensity*, *3d_printing_intensity*, and *age_years* all exhibit substantial variation. This demonstrates severe data skewness, thus, a log transformation was applied as a preprocessing step. Additionally, the data was standardized using z-scores.

Several columns contained a large proportion of missing data. To enhance data quality, we applied feature imputation using linear regression models. However, this approach did not yield meaningful improvements in both predictive accuracy and consecutive clustering performance. As a result, missing values were instead replaced with either zero or the median. For instance, the columns *employees* and *age_years* exhibited 37% and 29% missing data points, respectively, which were replaced with the corresponding median. Similarly, the columns *innoprob*, *sustainability_intensity*, *ai_intensity* and *3d_printing_intensity* had a high amount of missing data ranging between 16% and 19%. Those values were set to zero.

4.1 Results of Network Analysis

Selecting the sample with the highest modularity from the Louvain community detection algorithm outcome, as described in 3, resulted in 2,157 communities with a modularity score of 0.68. Focusing on the 20 largest communities, we conducted a visual analysis and identified three primary structural types: **cross-regional**, **regional**, and **hub-centered** communities.

Figure 3 illustrates three representative examples of these categories. The node size reflects the number of incoming links, with larger nodes representing companies with a higher amount of references by other companies.

- **The cross-regional community (A)** has a considerable concentration of nodes in the Stuttgart metropolitan area. However, the overall distribution of company nodes is geographically wide and connects urban areas across the whole state of BW.

- **The hub-centered community (B)** is characterized by the dominant cities Karlsruhe and, to a lesser degree, Stuttgart. These hubs dominate the overall network structure. Despite a broad geographic distribution, the community is shaped around a single influential company.

- **The regional community (C)** has a densely connected cluster of nodes primarily in the north of Ulm and the west of Stuttgart. Unlike the cross-regional structure (A), the geographic spread is narrower and more localized.

In addition, a small amount of communities shared features of both the cross-regional and hub-centered categories and were classified as a mixed-form. Appendix A summarizes all community visualizations for completeness.

4.2 Results of OPTICS

Figure 4 presents one example of our approach to clustering the data using only node level attributes. The reachability plot originates from a run of OPTICS with $min\_samples = 350$ and *metric* = Minkowski, as this configuration depicts a clean reachability graph. The cutoff was set to 1.1, resulting in four distinct clusters.

As a next step, we examined if the detected clusters also exhibit spatial structuring and visualized them, see Figure 5. However, the data points of the individual OPTICS clusters do not show any discernible spatial organization or localized aggregation (see individual plots in Appendix B1), suggesting a lack of inherent spatial clustering in the dataset.

To further assess the presence of any meaningful separations, Table 2 presents the mean values of each feature, stratified by cluster-ID. The mean number of employees in Cluster 1 is approximately 50% higher than across the other clusters, suggesting a difference in organizational size or workforce distribution. The standard deviation values reflect the variability in the number of employees within each cluster, with Cluster 2 exhibiting the highest dispersion (2,382) and Cluster 3 the lowest (1,331).

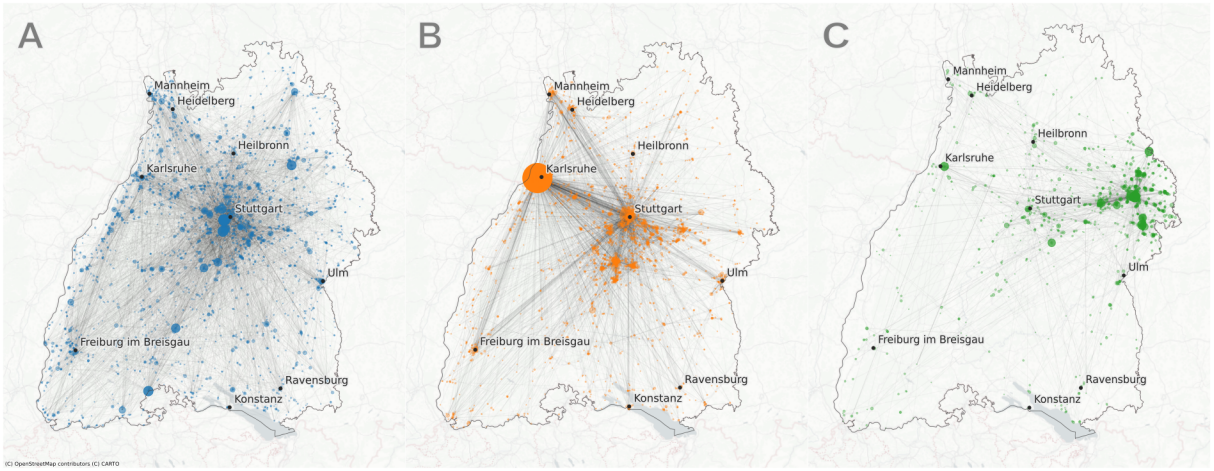

Figure 3. Exemplary results of the community detection: A cross-regional (A), hub-centered (B), and regional (C) community structure. Node size indicates the indegree.

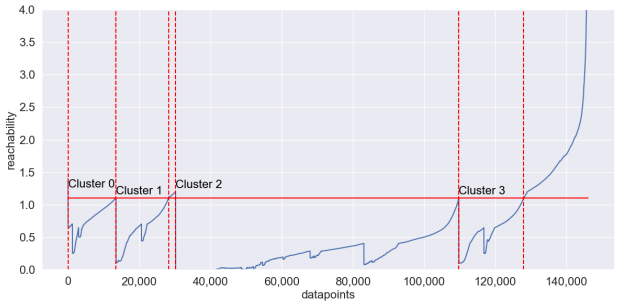

Figure 4. OPTICS reachability plot and detected clusters.

Cluster	0	1	2	3
Feature				
employees	47.8	63.1	41.4	43.6
indegree	0.8	0.84	0.83	0.84
outdegree	0.93	0.94	0.92	1.02
sustainability_intensity	0.089	0.091	0.08	0.086
ai_intensity	0.006	0.007	0.006	0.007
3d_printing_intensity	0.004	0.003	0.004	0.003
age_years	28.2	28.8	27.4	27.3
innoprob	0.304	0.306	0.307	0.305

Table 2. Mean feature values per OPTICS cluster result.

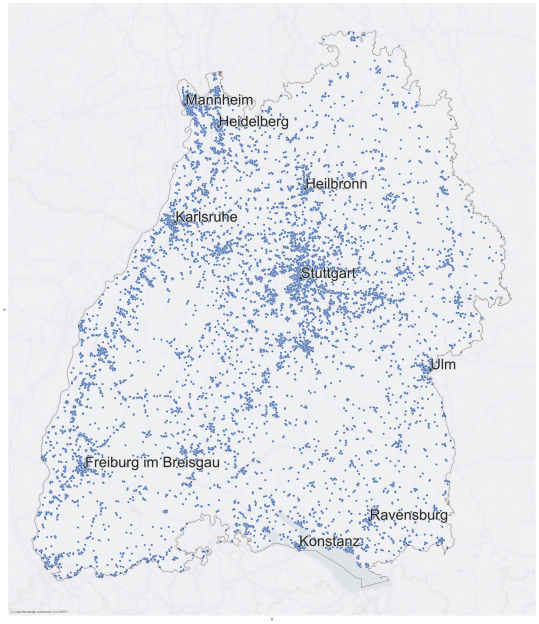

Figure 5. Exemplary visualization of OPTICS cluster 1. No discernible spatial patterns have been identified across clusters.

4.3 Results of t-SNE

Figure 6 visualizes the clustered t-SNE embeddings, illustrating the data structure. Following hyperparameter optimization, t-SNE was configured with a perplexity of 140, 2,000 iterations, and a cosine distance metric. Clustering was subsequently performed using the DBSCAN algorithm (Ester et al., 1996), setting the parameters *eps* to 3 and *min_samples* to 75 To refine results, clusters with fewer than 300 data points were reclassified as noise. This adjustment was deemed necessary due to the proliferation of small clusters, particularly in the lower-left region (low emb_x and emb_y values) of the plot.

The t-SNE embeddings generated with the cosine metric exhibited an improved separation of clusters compared to those computed using the Euclidean metric. Notably, a distinct, comparatively large, circular cluster is visible at the midpoint of Figure 6. This cluster was observed across all t-SNE embeddings derived with the cosine metric, irrespective of other parameter settings. However, when the Euclidean metric was used, this structure was significantly less pronounced.

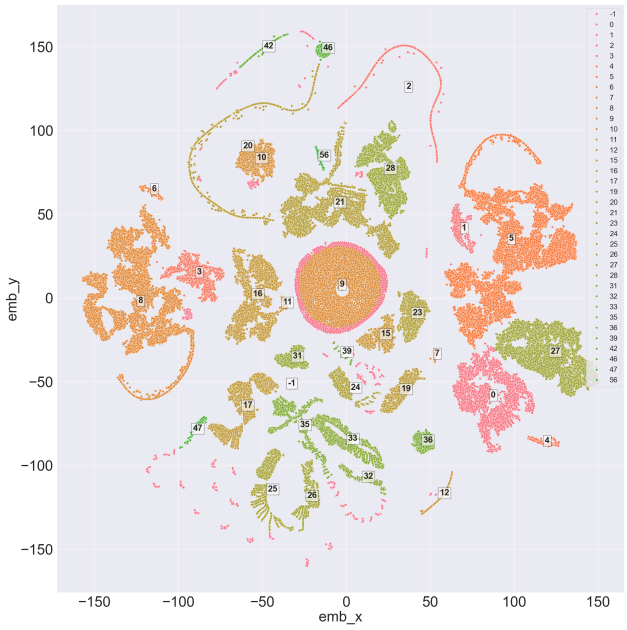

Figure 6. Depiction of the clustering of the t-SNE embedding.

To assess whether the clustering results exhibited any spatial relationships, the clustered t-SNE embeddings were visualized in Figure 7. However, no clear spatial relationships were present, suggesting that the clustering was not driven by spatial proximity.

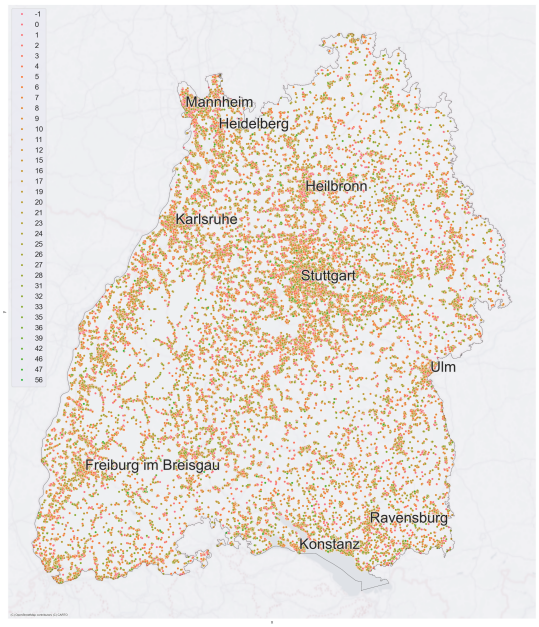

Figure 7. Spatial visualization of the t-SNE clustering result.

5 Discussion and Conclusion

This study analyzed how the collaboration networks in the German federal state of BW cluster based on hyperlink network structure and node properties, and how these clusters manifest geographically. Addressing RQ1 and studying the network edge based clustering, we identified three primary types of communities: cross-regional, regional, and hub-centered, akin to suggested structures in (Markusen, 1996). In contrast to our initial hypothesis, these structures were equally present in the hyperlink network of BW. We further observed that firms in larger urban centers, such as Stuttgart and Karlsruhe, form hub-centered communities, which is likely due to their size and reach in a region. The regional, spatially confined network structures suggest different collaboration patterns, potentially due to local specialized industry or a regional tech cluster. To address RQ2, we applied node-attribute-based clustering. Although OPTICS identified four clusters, substantial within-cluster variability limited clear differentiation. Similarly, t-SNE embeddings suggested distinct clusters, though their structure was largely driven by parameter settings rather than intrinsic data separation. Density-based clustering revealed no clear geographic patterns, indicating that cluster membership was not location-dependent. The absence of clear spatial structures in the node clustering may result from features lacking sufficient information for the study. Many node attributes came from website texts describing firms in specific technologies like 3D printing or AI, thus representing only a subset of firms. In addition, the ORBIS dataset contained a substantial amount of missing data. While this issue was addressed through feature imputation, the results proved unsatisfactory. Non-linear models or multiple imputation might have improved data quality but risk amplifying minor effects and introducing bias into the clustering. More comprehensive data with detailed firm-characterizing features are needed to identify meaningful clusters. The presented results reveal distinct spatial structures within the collaboration network of BW through the analysis of hyperlink data — an approach that can be readily extended to international contexts, where such analyses remain largely unexplored. Examining these global collaboration patterns can provide valuable insights into the structure of international partnerships, which, in turn, may support their strengthening and contribute to the advancement of the United Nation's Sustainable Development Goal 17.

6 Data and Software Availability Section

The code used for this analysis is fully available. Due to commercial issues with the data, an anonymized sample is provided together with the full code in this repository. The full dataset is available upon request from Sebastian Schmidt (sebastian.schmidt@istari.ai).

Declaration of Generative AI in writing

The authors declares that they have used Generative AI tools in the preparation of this manuscript. Specifically, the AI tools were utilized for language editing, improving grammar, word usage and sentence structure, but not for

generating scientific content, research data, or substantive conclusions. All intellectual and creative work, including the analysis and interpretation of data, is original and has been conducted by the authors without AI assistance.

Acknowledgments

We thank Johannes Dahlke for providing expert knowledge on the outlier filtering of the firm data.

This paper was funded through the Austrian Research Promotion Agency (FFG) and the DigitalInnovationLayer project (FO999898902).

Appendix A: Network Communities

Here, we provide a detailed overview over the communities detected as described in section 3. The three exemplary results in section 4.1 illustrated the detected classes, namely **regional** (Figure A1), **cross-regional** (Figure A2) and **hub-centered** (Figure A3) communities. Some communities featured traits of two classes and were therefore categorized as mixed form (Figure A4).

Appendix B: Geospatial Visualization of Node Clustering Results

B1 OPTICS Clustering Results

In Figure B1, we visualized the resulting clusters of the OPTICS clustering as described in section 4.2. Although the reachability plot indicated distinctly separated clusters, they do not show any spatially distinct features.

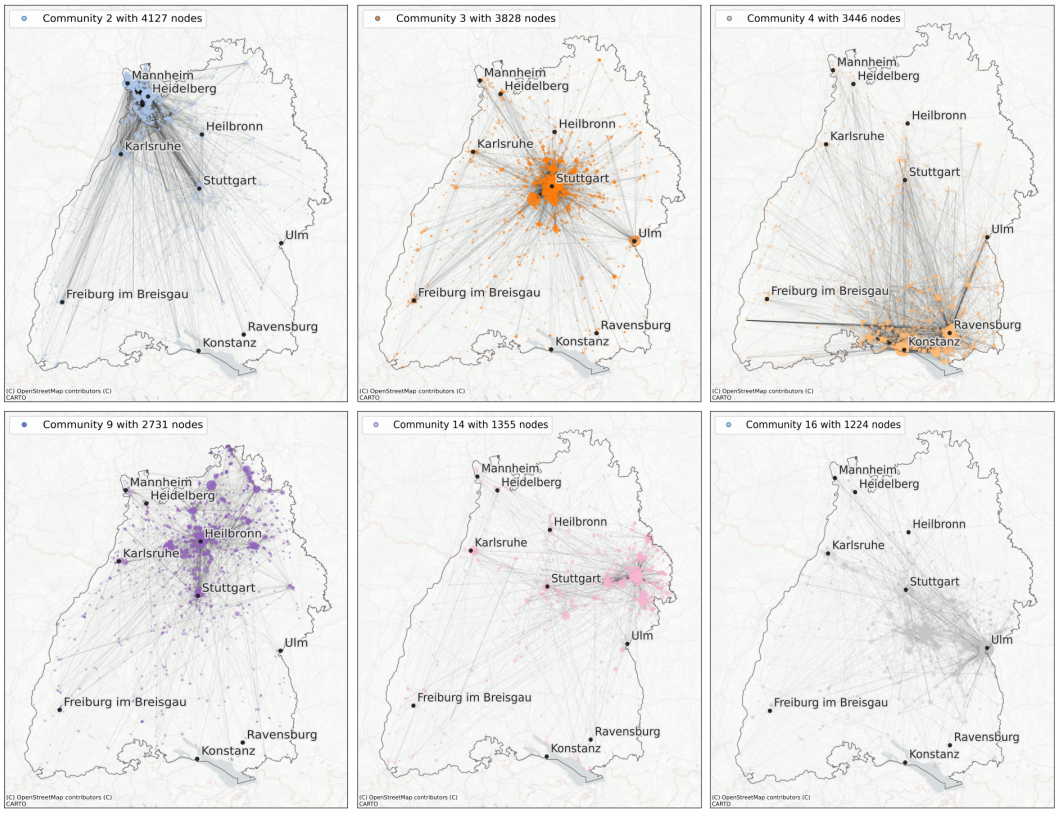

Figure A1. Regional network communities.

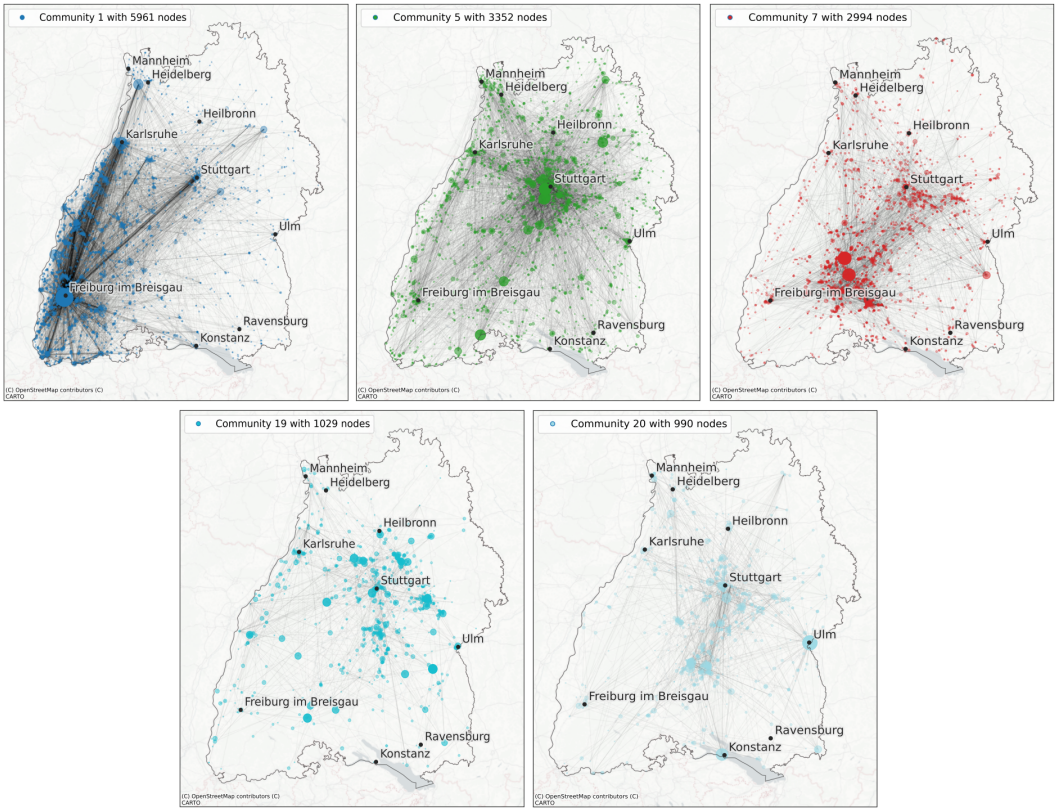

Figure A2. Cross-regional network communities.

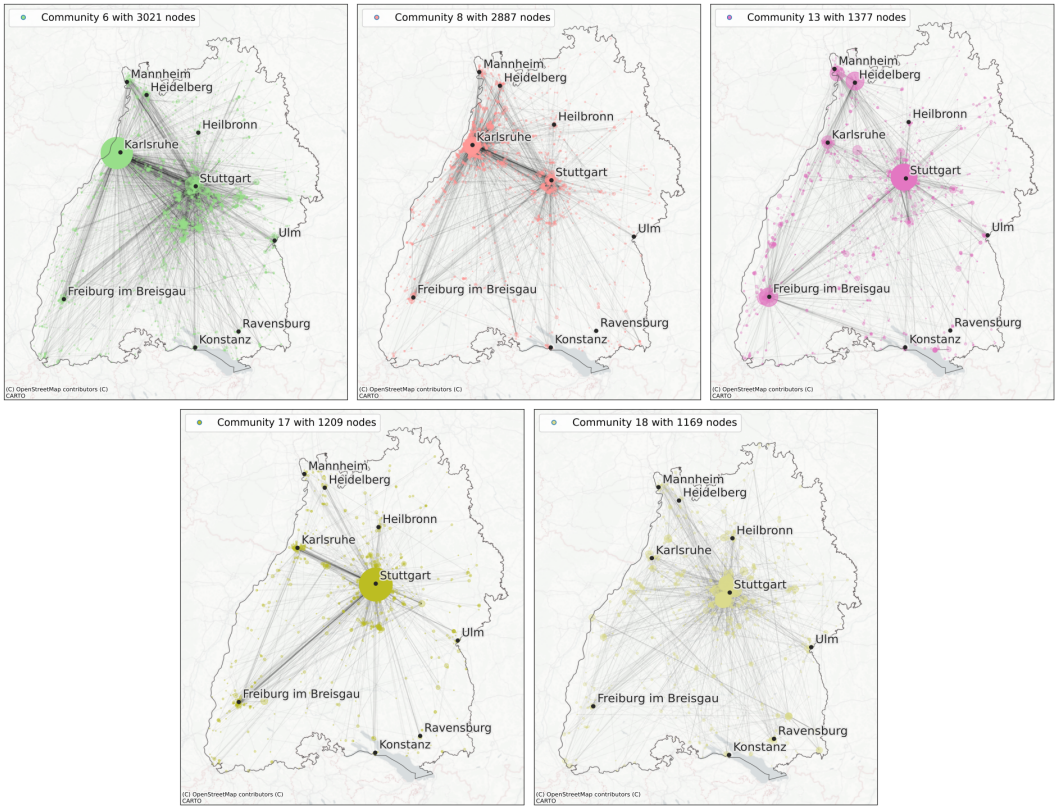

Figure A3. Hub-centered network communities.

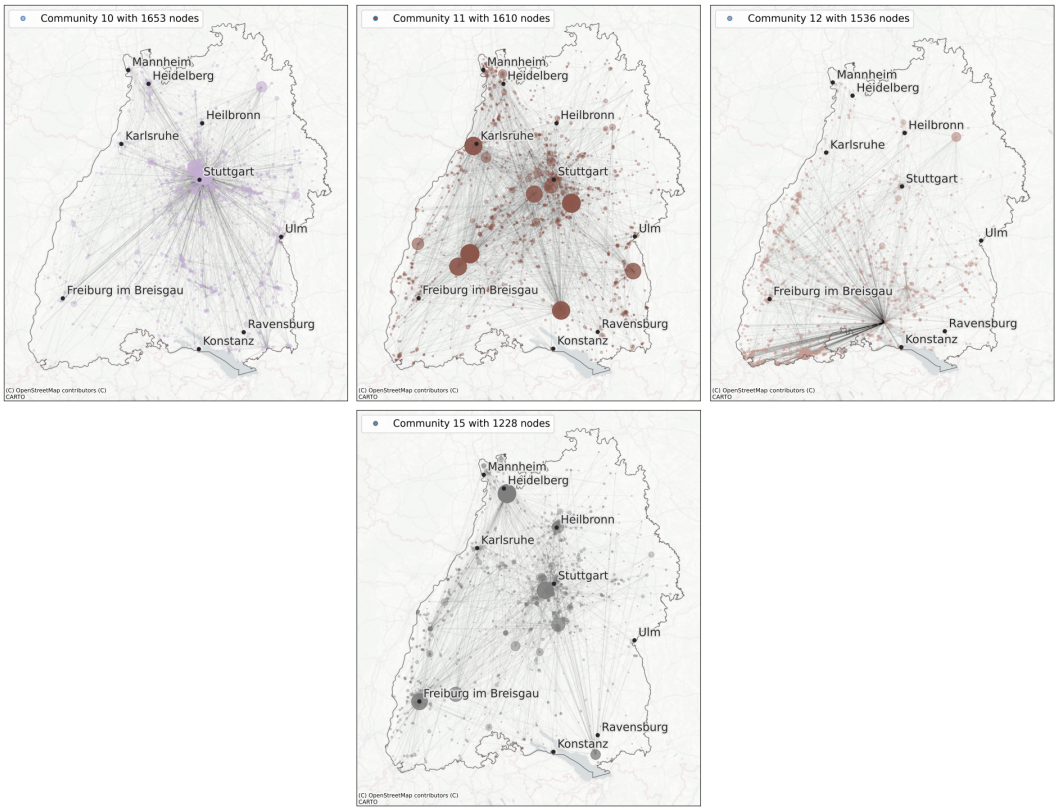

Figure A4. Mixed-form network communities.

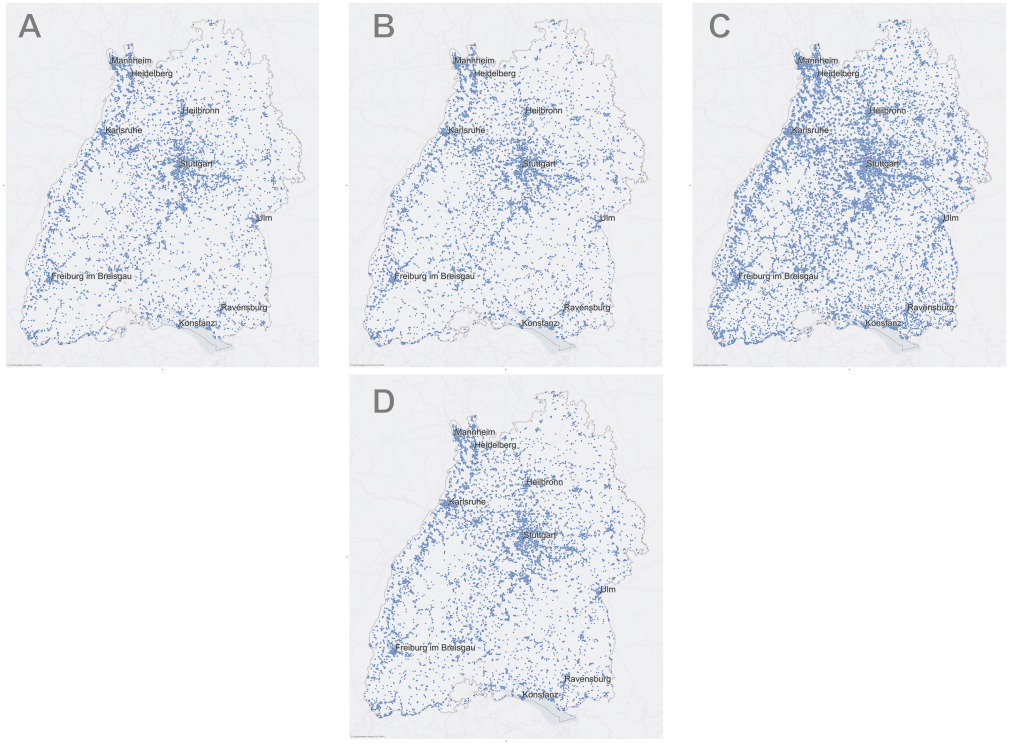

Figure B1. Geo-spatial visualization of OPTICS clusters.

References

Abbasiharofteh, M., Krüger, M., Kinne, J., Lenz, D., and Resch, B.: The Digital Layer: Alternative Data for Regional and Innovation Studies, Spatial Economic Analysis, 18, 507–529, https://doi.org/10.1080/17421772.2023.2193222, 2023.

Ankerst, M., Breunig, M. M., Kriegel, H.-P., and Sander, J.: OPTICS: Ordering points to identify the clustering structure, ACM Sigmod record, 28, 49–60, https://doi.org/10.1145/304181.304187, 1999.

Arthur, R. and Williams, H. T. P.: The human geography of Twitter: Quantifying regional identity and inter-region communication in England and Wales, PLOS ONE, 14, e0214 466, https://doi.org/10.1371/journal.pone.0214466, publisher: Public Library of Science, 2019.

Bailey, M., Cao, R., Kuchler, T., Stroebel, J., and Wong, A.: Social Connectedness: Measurement, Determinants, and Effects, Journal of Economic Perspectives, 32, 259–280, https://doi.org/10.1257/jep.32.3.259, 2018.

Blondel, V. D., Guillaume, J.-L., Lambiotte, R., and Lefebvre, E.: Fast Unfolding of Communities in Large Networks, Journal of Statistical Mechanics: Theory and Experiment, 2008, P10 008, https://doi.org/10.1088/1742-5468/2008/10/P10008, 2008.

Boots, B. and Tiefelsdorf, M.: Global and local spatial autocorrelation in bounded regular tessellations, Journal of Geographical Systems, 2, 319–348, https://doi.org/10.1007/PL00011461, 2000.

Bureau van Dijk: ORBIS - Data, https://orbis.bvdinfo.com/, 2023.

Dahlke, J., Beck, M., Kinne, J., Lenz, D., Dehghan, R., Wörter, M., and Ebersberger, B.: Epidemic Effects in the Diffusion of Emerging Digital Technologies: Evidence from Artificial Intelligence Adoption, Research Policy, 53, 104 917, https://doi.org/10.1016/j.respol.2023.104917, 2024.

De Maeyer, J.: Towards a Hyperlinked Society: A Critical Review of Link Studies, New Media & Society, 15, 737–751, https://doi.org/10.1177/1461444812462851, 2013.

Ester, M., Kriegel, H.-P., Sander, J., Xu, X., et al.: A density-based algorithm for discovering clusters in large spatial databases with noise, in: kdd, vol. 96, pp. 226–231, 1996.

Getis, A. and Ord, J. K.: The Analysis of Spatial Association by Use of Distance Statistics, Geographical Analysis, 24, 189–206, https://doi.org/10.1111/j.1538-4632.1992.tb00261.x, _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1538-4632.1992.tb00261.x, 1992.

Glückler, J., Punstein, A. M., Wuttke, C., and Kirchner, P.: The 'Hourglass' Model: An Institutional Morphology of Rural Industrialism in Baden-Württemberg, European Planning Studies, 28, 1554–1574, https://doi.org/10.1080/09654313.2019.1693981, 2020.

Halavais, A.: The Hyperlink as Organizing Principle, in: The Hyperlinked Society: Questioning Connections in the Digital Age, edited by Turow, J. and Tsui, L., The New Media World, pp. 39–55, Univ. of Michigan Press [u.a.], 2008.

Hervás-Oliver, J.-L., Parrilli, M. D., Rodríguez-Pose, A., and Sempere-Ripoll, F.: The Drivers of SME Innovation in the Regions of the EU, Research Policy, 50, 104 316, https://doi.org/10.1016/j.respol.2021.104316, 2021.

Hoffmann, L., Gilardi, L., Schmitz, M.-T., Erbertseder, T., Bittner, M., Wüst, S., Schmid, M., and Rittweger, J.: Investigating the Spatiotemporal Associations between Meteorological Conditions and Air Pollution in the Federal State Baden-Württemberg (Germany), Scientific Reports, 14, 5997, https://doi.org/10.1038/s41598-024-56513-4, 2024.

Kinne, J. and Axenbeck, J.: Web Mining for Innovation Ecosystem Mapping: A Framework and a Large-Scale Pilot Study, Scientometrics, 125, 2011–2041, https://doi.org/10.1007/s11192-020-03726-9, 2020.

Kinne, J. and Lenz, D.: Predicting Innovative Firms Using Web Mining and Deep Learning, PLOS ONE, 16, https://doi.org/10.1371/journal.pone.0249071, 2021.

Kullback, S. and Leibler, R. A.: On Information and Sufficiency, The Annals of Mathematical Statistics, 22, 79–86, https://www.jstor.org/stable/2236703, publisher: Institute of Mathematical Statistics, 1951.

Markusen, A.: Sticky Places in Slippery Space: A Typology of Industrial Districts, Economic Geography, 72, 293, https://doi.org/10.2307/144402, 1996.

Ratti, C., Sobolevsky, S., Calabrese, F., Andris, C., Reades, J., Martino, M., Claxton, R., and Strogatz, S. H.: Redrawing the Map of Great Britain from a Network of Human Interactions, PLOS ONE, 5, e14 248, https://doi.org/10.1371/journal.pone.0014248, publisher: Public Library of Science, 2010.

Scellato, S., Noulas, A., Lambiotte, R., and Mascolo, C.: Socio-Spatial Properties of Online Location-Based Social Networks, Proceedings of the International AAAI Conference on Web and Social Media, 5, 329–336, https://doi.org/10.1609/icwsm.v5i1.14094, number: 1, 2011.

Schlossstein, D. F. and Yun, J. J.: Innovation Cluster Characteristics of Baden-Wuerttemberg and Gyeongi-Do, Asian Journal of Technology Innovation, 16, 83–112, https://doi.org/10.1080/19761597.2008.9668658, 2008.

Schmidt, S., Kinne, J., Lautenbach, S., Blaschke, T., Lenz, D., and Resch, B.: Greenwashing in the US Metal Industry? A Novel Approach Combining SO_2 Concentrations from Satellite Data, a Plant-Level Firm Database and Web Text Mining, Science of the Total Environment, 835, 155 512, https://doi.org/10.1016/j.scitotenv.2022.155512, 2022.

Schmidt, S., Kanilmaz, U. N., Abbasiharofteh, M., and Resch, B.: Analysing the spatial manifestation of hyperlink networks of sustainability-engaged firms.A case study for Germany, Austria and Switzerland (under review)., Spatial Economic Analysis, 2025.

Schwierzy, J., Dehghan, R., Schmidt, S., Rodepeter, E., Stömmer, A., Uctum, K., Kinne, J., Lenz, D., and Hottenrott, H.: Technology Mapping Using WebAI: The Case of 3D Printing, https://arxiv.org/pdf/2201.01125, 2022.

Van der Maaten, L. and Hinton, G.: Visualizing data using t-SNE., Journal of machine learning research, 9, http://jmlr.org/papers/v9/vandermaaten08a.html, 2008.