Title

1. Introduction

Researchers have conceptualised innovation as a key economic force throughout the discipline's history. It is most prominent as a central component of economic growth, where technological change drives sustained increases in living standards. However, measurement and methodological barriers hampered research that directly evaluates the empirical aspects of innovation. While some innovation inputs, such as R&D spending, are directly observed, knowledge is not easily quantified, and the complexity of the networks governing its creation and dissemination poses a significant identification challenge.

The literature has attempted to overcome these empirical challenges in two primary ways: with measures of investment in innovation and through patent data. The second, which I focus on here, comprises a complete census of all patenting activity under a patent office and contains extensive information about inventions, inventors, and intellectual property rights holders. Although patents are only intermediate products in the innovation process, they offer a direct insight into how knowledge spreads since each patent cites prior art to demonstrate how it stands above it.[1]

One focal point for research using citation data has been testing whether knowledge spillovers are geographically localised. In the case of patent citations, this is akin to asking if inventors who cite each other also tend to be near each other once we control for other drivers of spatial concentration. To test this hypothesis, I use data from the United States Patent and Trademark Office (USPTO) on patents and citations. My primary contribution is to adopt tools

[1] For example, see the Patent and Trademark Law Amendments Act (1980), which codifies conditions for the citation of prior art in the US.

from natural language processing to improve matching approaches developed in prior research.

After reviewing relevant literature on the localisation of knowledge spillovers, I discuss my conceptual framework in Section 2, data and methods in Section 3, and results in Section 4. Finally, I conclude with a summary and potential extension in Section 5.

1.1. Patent citations and knowledge spillovers

Jaffe, Trajtenberg and Henderson (1993) introduced many key ideas underlying the analysis of knowledge spillovers through patent data. They argue that patent citations can accurately reflect these spillovers once we exclude commercial relations between inventors and assignees. The key argument is that inventors do not include citations that do not reflect spillovers since it would unnecessarily restrict the scope of an invention. Conversely, not citing a patent reflecting a knowledge flow would be challenged by the patent examiner, an expert over the relevant technologies.

However, findings could still reflect other agglomerative forces even if we exclude commercial transactions. Firms and inventors can benefit from sharing inputs to production and improve matching in labour markets when they co-locate (Carlino and Kerr, 2015, Section 6.4). Jaffe, Trajtenberg and Henderson (1993)'s insight was to create a control group of patents that mimic the technological and temporal characteristics of the citing patents. For each cited-citing patent pair, they find a control patent that does not cite the cited patent and has the same application year and three-digit United States Patent Classification (USPC) class as the citing patent.

They find localisation at the Standard Metropolitan Statistical Area (SMSA), state, and national levels. However, these results depended on how well the three-digit classes proxied endogenous factors. This issue led Thompson and Fox-Kean (2005a) to use the six-digit classes for matching controls to reassess earlier results. They find no evidence of intranational localisation. They argued that three-digit controls hid significant intra-class heterogeneity, but Henderson, Jaffe and Trajtenberg (2005) commented that boundaries between six-digit classes were arbitrary. Thompson and Fox-Kean (2005b) wrote another reply, but the issue remained unsettled.

Murata *et al.* (2014) followed in the spirit of these earlier papers but proposed important methodological advances. Its primary contribution was to adapt the localisation test of Duranton and Overman (2005) for the context of patent citations. Jaffe, Trajtenberg and Henderson (1993) and Thompson and Fox-Kean (2005a) used a discrete localisation measure. They compared the frequency at which cited and citing patents originated in the same discrete spatial unit (i.e., SMSA, state, and country) to that of cited and control. Not only did two patents in neighbouring units have the same impact as those across the country in the final results, also implied that the results were sensitive to the modifiable areal unit problem (Wong, 2009).

In contrast, the Duranton and Overman (2005) test, which I describe in more detail in Section 3, treats observations as points in continuous space. This approach addresses the aforementioned issues and incorporates a richer information set in estimated parameters. Murata *et al.* (2014) find evidence supporting localisation in 70% of all three-digit classes when using three-digit controls and in a third when using six-digit controls. Additionally, more than 10% of classes showed dispersion when using six-digit controls.

The latter point implies that aggregate results might fail to show localisation as the opposing forces would cancel out, which explains the different results in the original papers. However, it does not address the quality of either control. To do so, Murata *et al.* (2014) conducted a sensitivity analysis which generalises the controls to include three and six-digit as limiting cases. Their simulations show that most classes will still show localisation unless the matching procedure introduces extremely high selection bias.

1.2. Developments in patent text analysis

Although results from Murata *et al.* (2014) show that the evidence for localisation is robust, developments in patent analysis tools have created an opportunity to re-examine the matching approach. Natural language processing (NLP) has progressed incredibly in the past few years, and patents contain rich textual data in their abstracts, claims, and invention descriptions. Economists have already begun incorporating textual data sources in research, but only in a limited capacity. For an introduction to text data in economics, see Gentzkow, Kelly and Taddy (2019).

Arts, Cassiman and Gomez (2017) use the Jaccard similarity coefficient, the size of the intersection of words divided by the size of the union of words in two documents, to identify similar patents using their titles and abstracts. They found that patents matched with this method were likelier to have the same assignees and inventors, technological classification, and cite one another. The results were also validated by a panel of experts, highlighting that the index had weak matching power for patents with little text. As an application, they conduct a discrete space version of the matching approach, finding evidence for localisation.

However, Arts, Cassiman and Gomez (2017) and other uses of textual data within innovation economics (Kelly *et al.*, 2021; Kalyani *et al.*, 2025, for example) have focused on simple text-based statistics, far behind the current state-of-the-art. Since Vaswani *et al.* (2017) introduced the transformer architecture, deep learning has dominated much of the natural language domain. A leading use of the architecture has been for contextual embeddings, which are vector representations of the meaning of documents. These have since outperformed previous approaches in tasks like clustering, similarity, and information retrieval (Reimers and Gurevych, 2019), and researchers can easily access pre-trained models through libraries like Sentence Transformers in Python.

Once we encode a patent's text as an embedding, we can obtain a similarity measure between patents by calculating the vectors' angle (i.e., the cosine). More similar patents would have a smaller angle between their encodings, and we can explore the similarity space using nearest neighbours algorithms (Kelly *et al.*, 2021, Section 6.4). Sveva Ascione and Sterzi (2024) evaluate the performance of embedding models for patent similarity and find that transformer-based models outperform static measures. They use patent interferences, the case of distinct inventors submitting nearly identical claims simultaneously, as the benchmark for these tests.

I use a state-of-the-art contextual embedding model to match control to citing patents. However, due to time constraints, I have not fine-tuned the model. Despite this limitation, the model I use performs significantly better on benchmarks than the baseline contextual embedding model used by Sveva Ascione and Sterzi (2024). Feng (2020) is the only application of contextual embeddings in a matching localisation estimator that I could find. However, because it uses a discrete measure of space, it also suffers from spatial aggregation problems. They generally find weaker localisation evidence than Jaffe, Trajtenberg and Henderson (1993).

1.3. Alternative approaches and issues

Many authors have used alternative approaches to identify various aspects of agglomeration, including the localisation of knowledge spillovers. Buzard *et al.* (2017) find that R&D labs are highly concentrated. Buzard *et al.* (2020) match patents from these R&D clusters to show that these citations are also highly localised. They find that discrete agglomeration measures are likely to understate the degree of localisation and that knowledge spillovers operate over short distances.

The latter point matches with evidence from Arzaghi and Henderson (2008), which uses granular location data of advertising agencies in Manhattan and find that spillovers decay quickly. Kerr and Kominers (2015), further discussed in Section 3, provide a theoretical model showing that small interaction distances form larger clusters than these distances. They find that patent citation data generally supports their model's predictions. Other research supports localisation in the introduction of new words to trademarks (Graevenitz, Graham and Myers, 2021), patent interferences (Ganguli, Lin and Reynolds, 2020), and patent and research citations amongst universities (Belenzon and Schankerman, 2013).

Although inventors generally add patent citations, examiners and third parties might include them. Thompson (2006) compared variation within a patent in matching rates between inventor-added and examiner-added citations, finding that inventor-added ones are more localised. However, he uses a discrete space estimator. Buzard *et al.* (2020) re-tested their hypothesis with and without examiner-added citations, finding that the exclusion has no significant impact on the final results. Interestingly, Chen (2017) finds that examiner-added citing patents' texts are more similar to their cited patents than inventor-added.

Chen (2017) also finds that either case is more similar than between non-citing patents, but some have raised concerns about citation data in economics. Roach and Cohen (2013) compare patent citations with a survey of research reports from firm's R&D lab managers. They find that patents tend not to cite basic research important to their development, pointing towards understanding citations as capturing specific elements of knowledge spillovers. Kuhn, Younge and Marco (2020) show that the nature of patent citations has also been changing over time using measures of textual similarity. Hence, economists should consider citations alongside other measures and be careful when comparing across periods.

Researchers have also made significant methodological advances. Buzard *et al.* (2020) use coarsened exact matching (Iacus, King and Porro, 2017), which improves the balance of groups compared to my approach. This estimator and other balancing methods, such as the Covariate Balancing Propensity Score (Imai and Ratkovic, 2014), could efficiently incorporate even more information into creating balanced patent analysis samples.

Continuous space estimators have also been more prevalent in economics since Duranton and Overman (2005). Marcon and Puech (2017) have introduced a typology of these measures. Some measures, such as Lang, Marcon and Puech (2019), might provide more interpretable estimators with better properties, but economists have not generally adopted them.

2. Conceptual framework

3. Data and methods

3.1. Sample construction

I used data from the USPTO's PatentsView platform, which helps disseminate intellectual property data. It contains quarterly updated datasets created from raw patent information. The datasets undergo an entity (e.g., inventors, assignees, and locations) disambiguation process. PatentsView also standardises locations with places from the OSMNames project[2]. Finally, it contains data on five different patent classification systems, including the USPC and its successor, the Cooperative Patent Classification (CPC).

The platform serves data through an API and bulk downloads files. I used the API to fetch all utility patents granted between 2005 and 2025 with at least one US-based inventor and one US-based corporate assignee. I then excluded all patents with multiple assignees, any missing information, or citations not added by inventors or examiners. Finally, I removed those patents with abstracts in the bottom and top 0.01 per cent of the character count. These conditions were satisfied by around 2 million unique patents and 11,500 unique inventors. I collected citations, locations, and CPC class information from the bulk download files.

I chose utility patents and corporate assignees to match the sample construction choices of Jaffe, Trajtenberg and Henderson (1993) and Thompson and Fox-Kean (2005a). Previous papers generally included patents granted between the mid-70s and 1990 or 2000, so my sample covers a different period.

[2]The OSMNames database contains place names from OpenStreetMap and geographic information. It is available at https://osmnames.org/.

This fact likely would imply weaker localisation effects if we consider the increased importance of the internet. I restricted the number of assignees for two primary reasons: it simplified my coding, and multiple assignees could reflect complex commercial arrangements with inventors.

I chose utility patents and corporate assignees to match the sample construction choices of Jaffe, Trajtenberg and Henderson (1993) and Thompson and Fox-Kean (2005a). Previous papers generally included patents granted between the mid-70s and 1990 or 2000, so my sample covers a different period. This fact likely would imply weaker localisation effects if we consider the increased importance of the internet. For example, Kerr and Kerr (2018) highlight the increase in cross-country inventor teams. I restricted the number of assignees for two primary reasons: it simplified my coding, and multiple assignees could reflect complex commercial arrangements with inventors.

Missing observations are likely missing randomly and amount to less than a thousand observations. I observed that summary tables describing missingness varied between requests for the same data, likely indicating API issues. The applicant, the examiner, and third parties can include citations. I only kept the first two since the third, which comprises a small group, might not reflect spillovers. Although we could use the same argument to favour removing examiner-added citations, previous research has not used this restriction. An interesting extension would have been to examine both samples, but given time restrictions, I opted for the larger one.

Following the results of Arts, Cassiman and Gomez (2017), I restricted the abstract size to improve the matching quality. Short abstracts, the shortest being three characters, likely do not contain enough information to generate an appropriate embedding. Long abstracts would either require trimming prior to encoding or a model with a longer sequence length (i.e., the length of the text it can encode). Models that satisfy the latter condition require more compute, but given that the right tail of character sizes is thin, the impact of including them is likely nil.

3.2. Matching strategy

The general matching procedure is as follows. First, we select a period from which we define a set of cited patents. We then find all patents that cite any of the patents in the originating set. The cited-citing pairs then correspond to our treatment group. For each cited-citing pair, we define a set of multiple control patents that do not cite the cited patent of the pair but have characteristics similar to those of the control. The cited-control pairs correspond to the control group.

I intended to define the originating set as patents granted between 2005 and 2010 and consider all citations in my raw data. However, the matching process became too computationally expensive, so I had to create a reduced set to satisfy my time constraints. I defined all patents granted in the first month of 2005 with at least one citation in my data over the next five years as belonging to the cited set. I then joined each citing patent in these five years to their respective citation. As in previous papers, I exclude any cited-citing pairs with the same assignee or any inventor in common. These citations likely reflect commercial arrangements between assignees and inventors or continued work, so they are not true externalities.

To select the control patents, I proceeded in two steps. First, for each cited-citing pair, I select all patents granted between 2005 and 2010 with an application date within 180 days of the citing patent's application date, removing those with the same assignee or inventors as their respective cited patent.

Then, I encoded all unique citing and potential control patents as embeddings, and for each citing patent, I found the patents in the 99.9th similarity quantile. The intersection between patents within the date range and the similarity quantile corresponds to the set of admissible controls. Table 1 shows the count of unique patents before and after the embeddings — only a single cited patent had no citing patent with an appropriate control patent.

Table 1

Sample sizes before and after similarity matching

		Unique patents		
	Observations	Cited	Citing	Control
Pre-similarity matching	214,886,441	1,199	2,758	306,966
Post-similarity matching	147,088	1,198	2,754	70,231

Notes: The pre-similarity matching data includes controls with an application date within 180 days of the citing patent's application date that do not cite the cited patent or have any inventors or assignees in common. The post-similarity matching data consists of the intersection of the first group with the 308 nearest neighbours over all citing and control patents.

To encode the patent abstracts, I used the "nomic-embed-text-v2-moe" embedding model (Nussbaum and Duderstadt, 2025). This model has open-sourced data, weights (analogous to regression coefficients), and code and is available on HuggingFace. Then, I queried the nearest neighbours, as determined by the cosine of the embeddings, with an approximate nearest neighbour algorithm. The similarity quantile corresponded to the 308 nearest neighbours.

I also matched the final set of cited patents to their respective CPC sections. There are nine sections, A to H and Y, and patents might have more than one section. Table 2 lists the number of patents in each section alongside their description. Although the number of cited patents is proportional to the number of cite-citing-control triples, it is imperfect. This fact might indicate class heterogeneity in citations and matching rates. Coupled with the points

I make in Section 1.3, we must be careful when interpreting differences in localisation across classes.

Table 2

Counts of cited patents and patent triples by CPC section

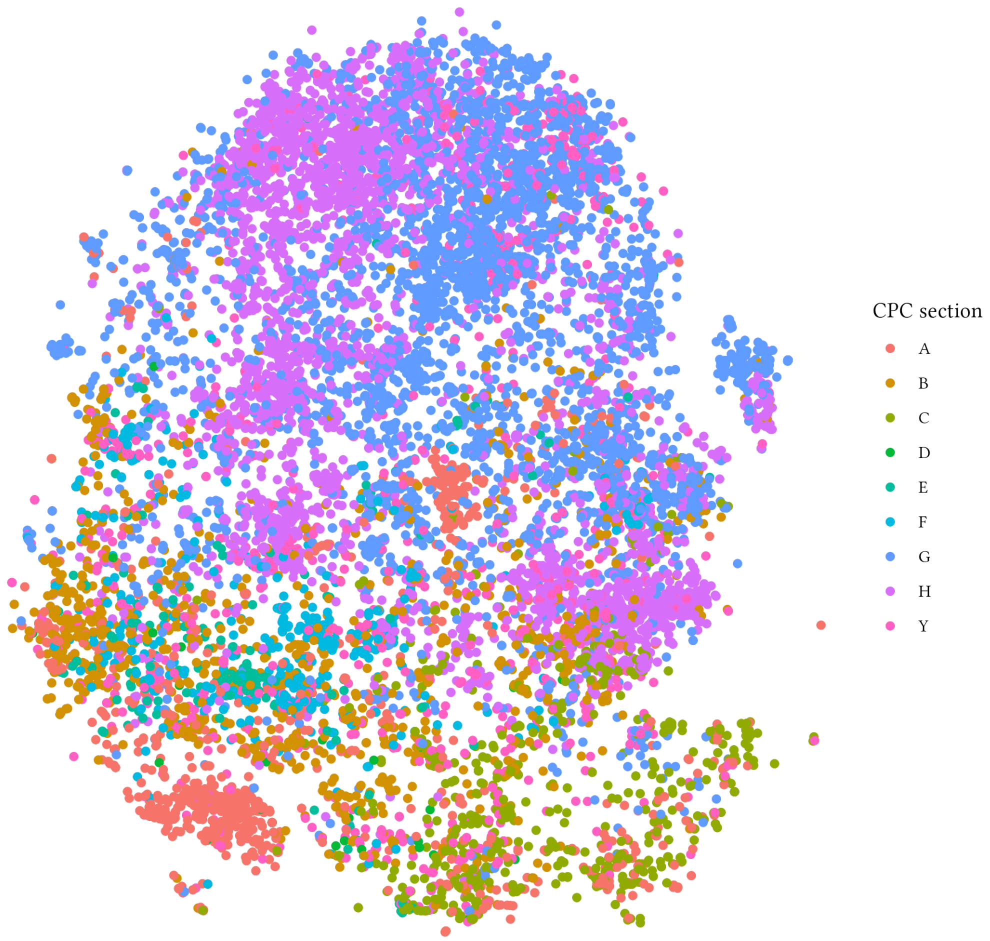
CPC section	Description	Counts: Cited patents	Counts: Patent triples
A	Human necessities	184	29,284
B	Performing operations; transporting	220	25,055
C	Chemistry; metallurgy	139	17,491
D	Textiles; paper	10	501
E	Fixed constructions	33	6,495
F	Mechanical engineering; lighting; heating; weapons; blasting engines or pumps	76	7,879
G	Physics	492	76,343
H	Electricity	448	68,480
Y	General tagging of new technological developments; general tagging of cross-sectional technologies spanning over several sections of the IPC; technical subjects covered by former USPC cross-reference art collections [XRACs] and digests	286	35,987

Notes: Cited patent counts refer to the count of unique cited patents in the final sample. A triple is a cited, citing, and control patent. Hence, multiple triples exist for each cited patent and each cited-citing pair. The definitions of the CPC section are from the European Patent Office website.

Figure 1 shows a random sample of embeddings represented in 2 dimensions using the t-SNE dimensionality reduction algorithm from the original 256 dimensions. Hence, proximity denotes semantic similarity. Colours denote the patent class, and we see evidence of clustering based on these colours. These patterns highlight how the text and the classes of the patents are connected, which is captured by the embeddings.

Figure 1

Patent embeddings and CPC classes



Notes: The figure shows a two-dimensional visualisation of 50,000 citing and control patent embeddings from a random sample. I keep only a unique CPC section selected at random per patent. I reduce dimensions using the t-SNE algorithm with a perplexity of 30 and theta of 0.5. Closer observations have similar meanings.

3.3. Localisation test

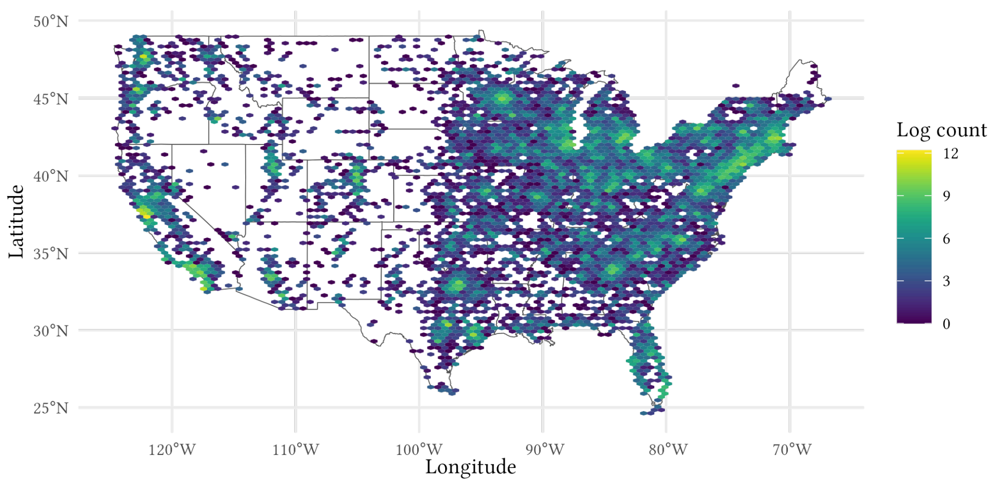
I calculated the geodesic distance between all cited-citing and cited-control pairs using their latitudes and longitudes for the localisation test. However, patents do not have a location, so most previous research constructed one from inventor locations. Since patents have multiple inventors who may work in different parts of the country, I followed Kerr and Kominers (2015)'s procedure to construct them.

First, I selected areas where the most inventors live for each patent. If this location was not unique, I selected the one corresponding to the lowest inventor sequence within the restricted set. The inventor sequence variable lists the order in which the patent lists its inventors. Generally, the first inventor has contributed the most to the invention, and the order between the others matters less. Therefore, the second step should capture the location that contributed the most or at least be random.

Figure 2 shows the matched sample's binned log count of all cited, citing, and control patents. As expected, the unconditional spatial distribution of patents shows geographic concentration. A clear example is Silicon Valley in California. This factor is one of the primary motivators for a matching strategy.

Figure 2

Binned patent distribution



Notes: The figure shows the spatial distribution of all patents in the matched sample in 100 hexagonal bins. The colours are in the log scale.

4. Results

5. Conclusion

Bibliography

Arts, S., Cassiman, B. and Gomez, J.C. (2017) "Text matching to measure patent similarity," *Strategic Management Journal*, 39(1), pp. 62–84. Available at: https://doi.org/10.1002/smj.2699.

Arzaghi, M. and Henderson, J.V. (2008) "Networking off Madison Avenue," *Review of Economic Studies*, 75(4), pp. 1011–1038. Available at: https://doi.org/10.1111/j.1467-937X.2008.00499.x.

Belenzon, S. and Schankerman, M. (2013) "Spreading the Word: Geography, Policy, and Knowledge Spillovers," *Review of Economics and Statistics*, 95(3), pp. 884–903. Available at: https://doi.org/10.1162/REST_a_00334.

Buzard, K. *et al.* (2017) "The agglomeration of American R&D labs," *Journal of Urban Economics*, 101, pp. 14–26. Available at: https://doi.org/10.1016/j.jue.2017.05.007.

Buzard, K. *et al.* (2020) "Localized knowledge spillovers: Evidence from the spatial clustering of R&D labs and patent citations," *Regional Science and Urban Economics*, 81. Available at: https://doi.org/10.1016/j.regsciurbeco.2019.103490.

Carlino, G. and Kerr, W.R. (2015) "Chapter 6 - Agglomeration and Innovation," in G. Duranton, J.V. Henderson, and W.C. Strange (eds.) *Handbook of Regional and Urban Economics*. Elsevier, pp. 349–404. Available at: https://doi.org/https://doi.org/10.1016/B978-0-444-59517-1.00006-4.

Chen, L. (2017) "Do patent citations indicate knowledge linkage? The evidence from text similarities between patents and their citations," *Journal of Informetrics*, 11(1), pp. 63–79. Available at: https://doi.org/10.1016/j.joi.2016.04.018.

Duranton, G. and Overman, H.G. (2005) "Testing for localization using micro-geographic data," *Review of Economic Studies*, 72(4), pp. 1077–1106. Available at: https://doi.org/Doi 10.1111/0034-6527.00362.

Feng, S. (2020) "The proximity of ideas: An analysis of patent text using machine learning," *PLoS One*, 15(7), p. e234880. Available at: https://doi.org/10.1371/journal.pone.0234880.

Ganguli, I., Lin, J. and Reynolds, N. (2020) "The Paper Trail of Knowledge Spillovers: Evidence from Patent Interferences," *American Economic Journal: Applied Economics*, 12(2), pp. 278–302. Available at: https://doi.org/10.1257/app.20180017.

Gentzkow, M., Kelly, B. and Taddy, M. (2019) "Text as Data," *Journal of Economic Literature*, 57(3), pp. 535–574. Available at: https://doi.org/10.1257/jel.20181020.

Graevenitz, G. von, Graham, S.J.H. and Myers, A.F. (2021) "Distance (still) hampers diffusion of innovations," *Regional Studies*, 56(2), pp. 227–241. Available at: https://doi.org/10.1080/00343404.2021.1918334.

Henderson, R., Jaffe, A. and Trajtenberg, M. (2005) "Patent Citations and the Geography of Knowledge Spillovers: A Reassessment: Comment," *American Economic Review*, 95(1), pp. 461–464. Available at: https://doi.org/10.1257/0002828053828644.

Iacus, S.M., King, G. and Porro, G. (2017) "Causal Inference without Balance Checking: Coarsened Exact Matching," *Political Analysis*, 20(1), pp. 1–24. Available at: https://doi.org/10.1093/pan/mpr013.

Imai, K. and Ratkovic, M. (2014) "Covariate Balancing Propensity Score," *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 76(1), pp. 243–263. Available at: https://doi.org/10.1111/rssb.12027.

Jaffe, A.B., Trajtenberg, M. and Henderson, R. (1993) "Geographic Localization of Knowledge Spillovers as Evidenced by Patent Citations," *The Quarterly Journal of Economics*, 108(3), pp. 577–598. Available at: https://doi.org/10.2307/2118401.

Kalyani, A. *et al.* (2025) "The Diffusion of New Technologies," *The Quarterly Journal of Economics*, 140(2), pp. 1299–1365. Available at: https://doi.org/10.1093/qje/qjaf002.

Kelly, B. *et al.* (2021) "Measuring Technological Innovation over the Long Run," *American Economic Review: Insights*, 3(3), pp. 303–320. Available at: https://doi.org/10.1257/aeri.20190499.

Kerr, S.P. and Kerr, W.R. (2018) "Global Collaborative Patents," *The Economic Journal*, 128(612), pp. F235–F272. Available at: https://doi.org/10.1111/ecoj.12369.

Kerr, W.R. and Kominers, S.D. (2015) "Agglomerative Forces and Cluster Shapes," *Review of Economics and Statistics*, 97(4), pp. 877–899. Available at: https://doi.org/10.1162/REST_a_00471.

Kuhn, J., Younge, K. and Marco, A. (2020) "Patent citations reexamined," *The RAND Journal of Economics*, 51(1), pp. 109–132. Available at: https://doi.org/10.1111/1756-2171.12307.

Lang, G., Marcon, E. and Puech, F. (2019) "Distance-based measures of spatial concentration: introducing a relative density function," *The Annals of Regional Science*, 64(2), pp. 243–265. Available at: https://doi.org/10.1007/s00168-019-00946-7.

Marcon, E. and Puech, F. (2017) "A typology of distance-based measures of spatial concentration," *Regional Science and Urban Economics*, 62, pp. 56–67. Available at: https://doi.org/10.1016/j.regsciurbeco.2016.10.004.

Murata, Y. *et al.* (2014) "Localized Knowledge Spillovers and Patent Citations: A Distance-Based Approach," *Review of Economics and Statistics*, 96(5), pp. 967–985. Available at: https://doi.org/10.1162/REST_a_00422.

Nussbaum, Z. and Duderstadt, B. (2025) *Training Sparse Mixture Of Experts Text Embedding Models*. Available at: https://doi.org/10.48550/arXiv.2502.07972.

Reimers, N. and Gurevych, I. (2019) "Sentence-BERT: Sentence Embeddings using Siamese BERT-Networks," *2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (Emnlp-Ijcnlp 2019)*, pp. 3982–3992.

Roach, M. and Cohen, W.M. (2013) "Lens or Prism? Patent Citations as a Measure of Knowledge Flows from Public Research," *Manage Sci*, 59(2), pp. 504–525. Available at: https://doi.org/10.1287/mnsc.1120.1644.

Sveva Ascione, G. and Sterzi, V. (2024) *A comparative analysis of embedding models for patent similarity*. Available at: https://doi.org/10.48550/arXiv.2403.16630.

Thompson, P. (2006) "Patent Citations and the Geography of Knowledge Spillovers: Evidence from Inventor- and Examiner-added Citations," *Review of Economics and Statistics*, 88(2), pp. 383–388. Available at: https://doi.org/10.1162/rest.88.2.383.

Thompson, P. and Fox-Kean, M. (2005a) "Patent Citations and the Geography of Knowledge Spillovers: A Reassessment," *American Economic Review*, 95(1), pp. 450–460. Available at: https://doi.org/10.1257/0002828053828509.

Thompson, P. and Fox-Kean, M. (2005b) "Patent Citations and the Geography of Knowledge Spillovers: A Reassessment: Reply," *American Economic Review*, 95(1), pp. 465–466. Available at: https://doi.org/10.1257/0002828053828617.

Vaswani, A. *et al.* (2017) "Attention Is All You Need," *Advances in Neural Information Processing Systems 30 (Nips 2017)*, 30.

Wong, D.W. (2009) "Modifiable Areal Unit Problem," in R. Kitchin and N. Thrift (eds.) *International Encyclopedia of Human Geography*. Oxford: Elsevier, pp. 169–174. Available at: https://doi.org/https://doi.org/10.1016/B978-008044910-4.00475-2.

"Patent and Trademark Law Amendments Act" (1980).