Gabriel Lobo de Oliveira, 202421832, EC423

Knowledge plays a critical role in many economic models, most apparent in endogenous human capital formation theories. Evaluating the assumptions of these models has proved challenging due to the difficulty of measuring knowledge, but the increased availability of patent data has addressed some of these issues. One particularly fruitful application for this data has been analysing to what extent knowledge spillovers are localised.

Jaffe, Trajtenberg and Henderson (1993)'s seminal paper largely initiated this literature by recognising that patent citations (of other patents) could proxy for these spillovers. However, they argued that spatial concentration could reflect other factors, such as the location of industries that created these patents. They addressed this issue by finding a "control" patent to originating-citing patent pairs. Matched controls did not cite the originating patents but belonged to a similar technological classification as the citing one.

They found strong evidence of localisation in knowledge spillovers, but subsequent papers have questioned the results. Most prominent, Thompson and Fox-Kean (2005a) argued that the classification system used to match citing patents with their controls was too broad, so they reproduced the analysis with a more granular approach. The paper did not find significant effects except at the country level, which led to further discussion (Henderson, Jaffe and Trajtenberg, 2005; Thompson and Fox-Kean, 2005b).

Building on this discussion, I propose investigating patterns in the geographic characteristics of knowledge spillovers over time. I intend to contribute to this literature in three main ways:

1. The USPTO has significantly simplified access to its data, making performing analyses on a larger scale (e.g., multiple reference years for originating patents) more feasible[1]. Their database applies a consistent inventor disambiguation strategy and retroactively classifies patents.
2. Researchers have significantly advanced the computational tools and methods for analysing spatial data. One noteworthy example is Murata *et al.* (2014), which looks at the distances between inventors rather than a binary classification for whether they are in the same statistical region.
3. Similarly, natural language processing has made massive strides over recent years, with pre-trained state-of-the-art models becoming freely available and computationally efficient. I intend to use semantic similarity models to match citing patents to their controls.

The third point is the main contribution of my essay. The core idea is to use vector embeddings, transforming text (e.g., words or sentences) into mathematical vectors. The location of these vectors captures semantic relationships between the vectorised documents, and I can compute some measure of distances between texts (e.g., cosine similarity).

[1] I already received API access from the USPTO and developed some data retrieval functions. The raw data is organised and relatively easy to clean.

References

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.

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.

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.

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.