

The geography of innovation as reflected by social media

Article

Published Version

Creative Commons: Attribution-Noncommercial 4.0

Open Access

Corradini, C. ORCID: <https://orcid.org/0000-0002-1980-7993>
(2021) The geography of innovation as reflected by social media. *Environment and Planning A: Economy and Space*, 53 (2). pp. 227-229. ISSN 1472-3409 doi: <https://doi.org/10.1177/0308518X20961107> Available at <https://centaur.reading.ac.uk/108036/>

It is advisable to refer to the publisher's version if you intend to cite from the work. See [Guidance on citing](#).

To link to this article DOI: <http://dx.doi.org/10.1177/0308518X20961107>

Publisher: SAGE Publications

All outputs in CentAUR are protected by Intellectual Property Rights law, including copyright law. Copyright and IPR is retained by the creators or other copyright holders. Terms and conditions for use of this material are defined in the [End User Agreement](#).

www.reading.ac.uk/centaur

CentAUR

Central Archive at the University of Reading

Reading's research outputs online



The geography of innovation as reflected by social media

EPA: Economy and Space

2021, Vol. 53(2) 227–229

© The Author(s) 2020



Article reuse guidelines:

sagepub.com/journals-permissions

DOI: 10.1177/0308518X20961107

journals.sagepub.com/home/eprn**Carlo Corradini** 

University of Birmingham, UK

Abstract

Capturing whether mysteries of trade are ‘in the air’ remains a central issue for understanding the spatial nature of innovative activities. We offer new insights on the intangible dimension of the geography of innovation comparing the spatial distribution of patenting rates and a novel measure of ‘innovative buzz’ from a 9-billion-word corpus based on 900 million geolocated tweets across US counties. We find the use of words connected to innovation and technology to be characterised by significant spatial concentration. The results show the presence of strong bivariate correlation in the spatial clustering between the two elements, pointing to stimulating opportunities in the use of big data from social media for economic geography research.

Keywords

Social Media, geotagged Twitter, geography of innovation, patents, big data

Researchers have started leveraging new sources of geo-localised information available through Twitter to offer insights on a diverse range of spatial perspectives including localised personality traits and geographical variation of heart-disease (Eichstaedt et al., 2015; Obschonka et al., 2019). Similarly, linguistics analysis has utilised big data from social media to reveal regional linguistic variation (Grieve et al., 2018).

This study applies big data analysis to explore intangible elements in the geography of innovation. We compare the spatial clustering of patents per capita (Pat_Cap), collected from the USPTO, to a novel variable reflecting social media discussion and ‘buzz’ around topics related to technological innovation. This variable, labelled InnoTech_Tw, is based on a corpus of around 8.9 billion words from 890 million geocoded Tweets across US counties in 2014 (for more information on the dataset, see Grieve et al., 2018). It is defined as the sum of the relative frequency of all words in each US county weighted by their cosine similarity to both terms *innovation* and *technology* obtained by applying the word2vec algorithm on a dataset of 300 dimensional vectors for 3 million words trained on a corpus of roughly 100

Corresponding author:

Carlo Corradini, University of Birmingham, Birmingham, UK.

Email: c.corradini@bham.ac.uk

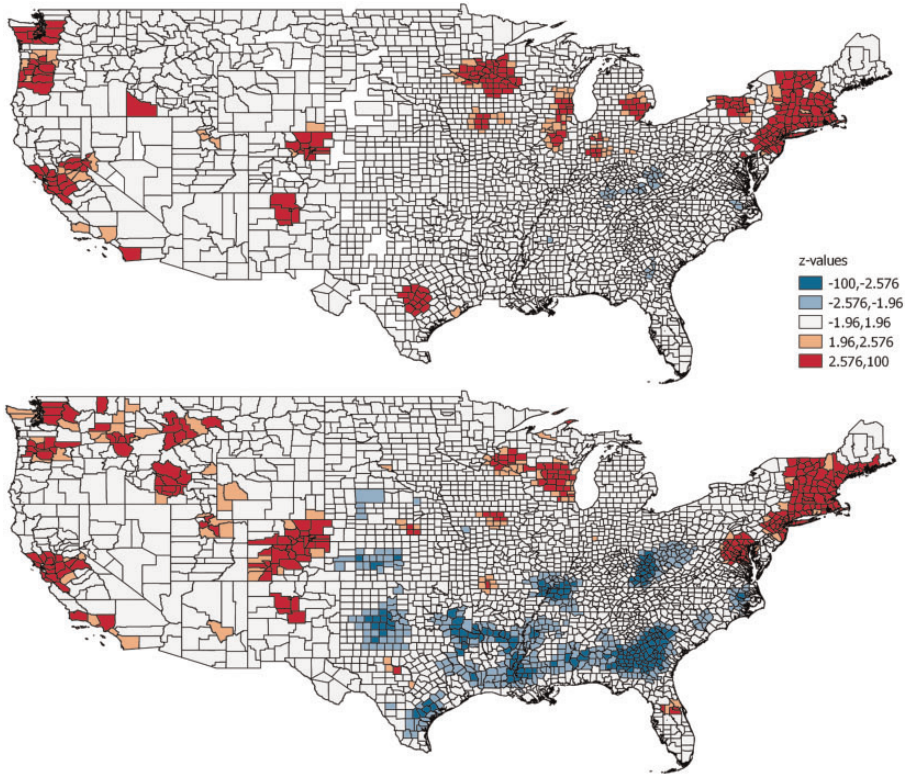


Figure 1. Getis-Ord G_i^* z-scores for Pat_Cap (Top panel) and for InnoTech_Tw (Bottom panel).

billion words from Google News¹ (Mikolov et al., 2013). In line with text analysis methods, the values are adjusted by the inverse word frequency in order to confer a smaller weight to words that are more common in the corpus.

We compare the spatial distribution patterns of the two variables analysing their Getis-Ord $G_i^*(d)$ statistic as measure of local clustering (Getis and Ord, 1992). The threshold distance for the spatial weight matrix is set at 100 km.² We map the z-scores in Figure 1. Patents per capita reflect the well-known uneven distribution of patenting intensity across US counties. Similarly, z-scores for InnoTech_Tw also indicate marked spatial concentration in ‘innovative buzz’ within tweets. More importantly, spatial clusters for the two variables broadly overlap.

We illustrate this relationship in Figure 2 across four quantiles of college education levels. The overall Pearson’s correlation coefficient for the Getis-Ord G_i^* z-scores is 0.68 ($p < .001$). Considering only significant z-scores beyond the 1.96 significance level this increases to 0.75. The strength in this relationship is confirmed by the strong Pearson’s correlation between log of patents per capita and InnoTech_Tw of 0.69 ($p < .001$) shown in the right panel of Figure 2.

These results highlight the significant opportunities available in big data and language analysis from social media for economic geography. We note this may be particularly important in unpacking the intangible dimension of the innovative character of places.

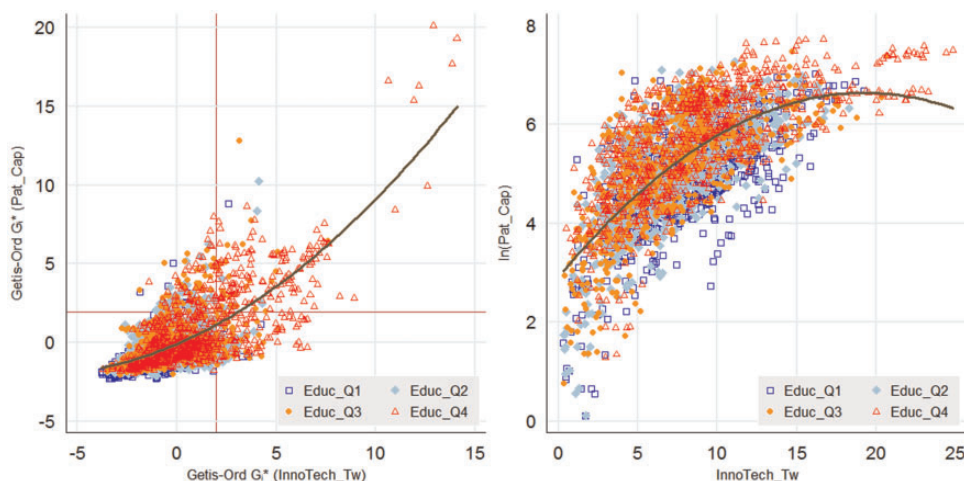


Figure 2. Correlation between Getis-Ord G_i^* z-scores for Pat_Cap and InnoTech_Tw (Left panel); correlation between log of Pat_Cap and InnoTech_Tw (Right panel).

Declaration of conflicting interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship and/or publication of this article: The research was funded by a Leverhulme Trust research fellowship (Grant Number: RF-2020–505\7).

ORCID iD

Carlo Corradini  <https://orcid.org/0000-0002-1980-7993>

Notes

1. Available here: <https://code.google.com/archive/p/word2vec/>
2. Results are robust at 50, 150 and 200km thresholds.

References

- Eichstaedt JC, Schwartz HA, Kern ML, et al. (2015) Psychological language on twitter predicts county-level heart disease mortality. *Psychological Science* 26(2): 159–169.
- Getis A and Ord JK (1992) The analysis of spatial association by use of distance statistics. *Geographical Analysis* 24(3): 189–206.
- Grieve J, Nini A and Guo D (2018) Mapping lexical innovation on American social media. *Journal of English Linguistics* 46(4): 293–319.
- Mikolov T, Chen K, Corrado G, et al. (2013) *Efficient estimation of word representations in vector space*. Proceedings of International Conference on Learning Representations (ICLR).
- Obschonka M, Lee N, Rodríguez-Pose A, et al. (2019) Big data methods, social media, and the psychology of entrepreneurial regions: capturing cross-county personality traits and their impact on entrepreneurship in the USA. *Small Business Economics* : 1–22.