

Listening to the buzz: exploring the link between firm creation and regional innovative atmosphere 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>,
Folmer, E. and Rebmann, A. (2022) Listening to the buzz:
exploring the link between firm creation and regional
innovative atmosphere as reflected by social media.
Environment and Planning A: Economy and Space, 54 (2). pp.
347-369. ISSN 1472-3409 doi:
<https://doi.org/10.1177/0308518X211056653> Available at
<https://centaur.reading.ac.uk/111522/>

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/10.1177/0308518X211056653>

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

Listening to the buzz: Exploring the link between firm creation and regional innovative atmosphere as reflected by social media

EPA: Economy and Space

1–23

© The Author(s) 2021



Article reuse guidelines:

sagepub.com/journals-permissions

DOI: 10.1177/0308518X211056653

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

University of Birmingham, UK

Emma Folmer

University of Groningen, The Netherlands

Anna Rebmann

King's College London, UK

Abstract

This paper presents a novel approach to capture ‘buzz’, the vibrancy and knowledge exchange propensity of localised informal communication flows. Building on a conceptual framework based on relational economic geography, we argue the content of buzz may allow to probe into the character of places and investigate what is ‘in the air’ within regional entrepreneurial milieux. In particular, we analyse big data to listen for the presence of buzz about innovation – defined by discursive practices that reflect an innovative atmosphere – and explore how this may influence regional firm creation. Using information from 180 million geolocated Tweets comprising almost two billion words across NUTS3 regions in the UK for the year 2014, our results offer novel evidence, robust to different model specifications, that regions characterised by a relatively higher intensity of discussion and vibrancy around topics related to innovation may provide a more effective set of informal resources for sharing and recombination of ideas, defining regional capabilities to support and facilitate entrepreneurial processes. The findings contribute to the literature on the intangible dimensions in the geography of innovation and offer new insights on the potential of natural language processing for economic geography research.

Keywords

big data, buzz, entrepreneurship, knowledge spillovers, Twitter

Corresponding author:

Carlo Corradini, University of Birmingham, Birmingham, UK.

Email: c.corradini@bham.ac.uk

Introduction

Building on the seminal analysis of territorial agglomerations by Marshall (1890), the literature has described regions as loci for interactive learning underpinning innovation and new knowledge creation (Bathelt and Glückler, 2003; Camagni, 1991; Malmberg and Maskell, 2002; Storper, 1997). This perspective has also served as a framework of reference for the analysis of entrepreneurial processes at the regional level, where knowledge created but left unexploited by incumbent actors and then recombined with other ideas by spatially embedded agents becomes a critical driver of new firm creation (Acs et al., 2009; Audretsch et al., 2008). These dynamics present a strong spatial dimension as the exchange of tacit and uncodified knowledge which underpins entrepreneurial opportunity recognition occurs more effectively through localised interactions (Jaffe et al., 1993; Polanyi, 1967; Sonn and Storper, 2008). It is through these dense social networks that ‘entrepreneurs tap into knowledge streams they would otherwise not be able to access’ (Spigel, 2017: 4).

Economic geographers have underlined the importance of the complex web of information and knowledge exchange ecology created via co-location through the concept of ‘buzz’ (Bathelt et al., 2004; Storper and Venables, 2004). Yet, partly reflecting the intangible nature of this concept, buzz has often been simplified as a synonym for density of interaction and its analysis has remained subsumed within the broader effects of agglomeration economies and proximity in economic activity (Qian et al., 2013). This has led to researchers ignoring an important dimension of the concept – the *content* of buzz that is reflective of ‘nondeliberate knowledge-exchange and information-exchange propensities’ (Asheim et al., 2007: 659). Connecting to the role of place-specific informal institutions and social and normative contexts in shaping relational structures (Gertler, 2003; Saxenian, 1996; Storper, 1997, 2018), the content of buzz can be seen as indicative of whether, even for the same levels of knowledge and economic density, some places may be defined by a more innovative atmosphere better supporting the propensity to discuss new ideas and engage in innovative practices.

In this paper, we argue that listening for the presence of innovation within localised buzz represents an important element for probing into the intangible character of places to capture the seminal notion of ‘industrial atmosphere’ discussed by Marshall (1890). Building on a conceptual framework based on relational economic geography and the knowledge spillover theory of entrepreneurship (Acs et al., 2009; Bathelt and Glückler, 2003; Storper, 1997), we posit observing the *content* of buzz may reveal the extent of the innovativeness of knowledge-exchange ecologies that define the regional entrepreneurial context. In particular, the content of buzz may be indicative of both an informative and normative dimension (Bathelt et al., 2004; Storper and Venables, 2004). In this sense, the presence of buzz about innovation may reflect a localised relational infrastructure of shared interests and interpretative schemes, surrounding potential entrepreneurs with valuable knowledge and a continuous re-discussion of information about new opportunities. At the same time, it may reflect localised descriptive norms encouraging the discussion and implementation of novel ideas, thereby supporting entrepreneurial action.

To ‘listen’ to the buzz, we leverage the opportunities offered by big data recently highlighted in the scholarly community in the context of regional studies and entrepreneurship research (Indaco, 2020; Schwab and Zhang, 2019; Van Meeteren and Poorthuis, 2018). In the last decade, the increasing availability of geo-tagged data from social network sites has led to a substantial body of research exploring the structure and spatiality of online communities (Crampton et al., 2013; Stephens and Poorthuis, 2015; Tranos, 2020), including digitally mediated interactions of entrepreneurial networks (Wang et al., 2017). Here, we endeavour to contribute to this literature following a complementary perspective focusing on the content of geo-located social media and language patterns in economic geography (Casadei and Lee, 2020; Obschonka et al., 2020). In particular, we explore a set of measures based on natural language processing to investigate whether regions

with buzz characterised by discursive practices reflecting an emphasis on innovation experience higher rates of entrepreneurship, using data from 180 million geolocated Tweets encompassing almost two billion words across NUTS3 regions in the UK for the year 2014.

Controlling for regional factor endowments such as GDP and different measures of human capital as well as density, we present novel evidence suggesting regions characterised by a relatively higher intensity of discussion and vibrancy around topics related to innovation may provide a more effective environment for entrepreneurial processes. While the explorative nature of our analysis does not address causal effects, our findings allow us to advance the concept of buzz by emphasizing the importance of its content and the potential in geo-located big data sources for comparing the ‘innovativeness’ of buzz across places.

The rest of the paper is organised as follows. In the second section, we delineate the theoretical framework of the paper. In the following sections, data and methods are discussed before presenting results and a discussion of the findings. The final section concludes with a review of the key theoretical as well as methodological implications, together with limitations of the paper.

Theoretical framework

Entrepreneurship, spatial proximity and buzz

Scholars have long emphasised the role of localised factor conditions and geographical proximity as fundamental drivers of new firm creation (Armington and Acs, 2002; Fotopoulos, 2013; Fritsch and Falck, 2007). The relationship between local knowledge, proximity and new firm creation is formalised in the knowledge spill-over theory of entrepreneurship, which suggests entrepreneurs start new firms to take advantage of opportunities offered by knowledge created but left unexploited by incumbents (Acs et al., 2009, 2012; Audretsch and Lehmann, 2005; Audretsch et al., 2008; Qian and Jung, 2017). In line with this framework, entrepreneurial opportunities are broadly defined as a function of place-specific knowledge endowments and higher density of economic activity (Qian and Acs, 2013; Qian et al., 2013). Accordingly, empirical studies have pointed to a positive relationship between investment in knowledge creation within a region and the resulting knowledge spillovers on entrepreneurship rates (Audretsch and Lehmann, 2005; Audretsch et al., 2008). Thus, ‘contexts rich in knowledge will generate more entrepreneurial opportunities’ (Audretsch and Keilbach, 2008: 1699). Similarly, the presence of higher levels of human capital allows for both a source of knowledge as well as providing higher levels of entrepreneurial absorptive capacity (Qian and Acs, 2013; Qian et al., 2013). The role of density and agglomeration for entrepreneurial processes is defined by the locally bound nature of knowledge spillovers, whose diffusion is spatially constrained because the transfer of knowledge is not costless across geographic distance (Audretsch and Lehmann, 2005). This is especially the case for ‘tacit’ knowledge, which remains highly context specific due to its uncertain and fuzzy nature (Gertler, 2003). As knowledge diffusion tends to be ‘sticky’, the density of economic activity becomes the conduit for the transmission of new ideas and entrepreneurial opportunities, encouraging the clustering of entrepreneurial and innovative activities (Glaeser et al., 2010).

The importance of geographical proximity and localised interaction for knowledge recombination is notably captured in the concept of ‘buzz’. Bathelt et al. (2004: 38) have defined ‘buzz’ as ‘the information and communication ecology created by face-to-face contacts, co-presence and co-location of people and firms within the same industry and place or region’. Scholars have traditionally underlined spatial proximity as being critical for buzz (Bathelt et al., 2004; Gertler, 2003; Storper and Venables, 2004). Yet, its conceptualisation is not confined to a synonym for density of interaction (Qian et al., 2013). Buzz comprises informal institutions rooted in a place-specific social and cultural milieu (Malmberg and Maskell, 2002). These define an informal relational

infrastructure offering an effective conduit for the diffusion of information through the shared interests, attitudes and interpretative schemes of embedded participants (Bathelt et al., 2004). In this sense, buzz reflects the emergence of a common language that underpins collective learning within 'innovative milieux' (Lawson and Lorenz, 1999). These dynamics distinguish the informative character of buzz as a critical driver of tacit knowledge exchange, also revealing the socialisation effects within the concept (Storper and Venables, 2004). This framework also reflects similar perspectives in the entrepreneurship literature suggesting knowledge about opportunities and new technologies rests upon social attributes that are context specific (Spigel, 2017).

Building on these insights, we propose that the presence of buzz about innovation may be reflective of an atmosphere characterised by vibrancy around innovation related topics conducive to the exchange of novel ideas and be indicative of local descriptive norms of innovative behaviour, thereby fostering regional entrepreneurship.

Buzz and the innovative atmosphere

In economic geography literature, scholars have long underlined the importance of a broad set of intangible elements for knowledge diffusion within clusters suggesting that, of places with similar levels of knowledge creation and density of interaction, some may be defined by a stronger drive to innovate (Saxenian, 1996; Storper et al., 2015; Urbano et al., 2019). Moving beyond traditional perspectives based on education levels and R&D activity as determinants of such an innovative environment, growing research has started looking at the role of creative occupations in shaping such an innovative atmosphere for new firm formation. Several authors have underlined that the level of human capital is not necessarily a conduit for knowledge creation and entrepreneurship, and its effect is dependent upon whether it is used in a creative manner (Audretsch and Belitski, 2013; Lee et al., 2004; Sleuwaegen and Boiardi, 2014). Firms in creative industries have been associated with the introduction of original innovation across cities (Lee and Rodríguez-Pose, 2014). Building on the arguments by Florida (2002) on the creative class, research shows that higher proportions of people in creative industries and 'bohemians' are linked to regional firm entry (Lee et al., 2004) and city level firm entry (Audretsch and Belitski, 2013).

However, this view has been criticised along two main elements. First, some authors argue the creative capital of regions is simply a proxy for human capital (Glaeser, 2004). Others have pointed out that measuring creativity using creative occupations may lead to excluding many other jobs that are not classified as creative but actually do present significant aspects of innovative content (Boschma and Fritsch, 2009). Indeed, Florida himself recognises this point when he argues that creativity can be required from manufacturing and assembly workers as they engage in problem-solving to improve productivity and deal with breakdowns (Florida, 2012). Levels of creativity and innovation in a region are thus important but may not be accurately captured by measures of human capital and occupational structure (Sleuwaegen and Boiardi, 2014).

In particular, some authors have criticised these approaches for being centred solely on the presence of people with creative and innovative capabilities (Storper and Scott, 2009; Storper et al., 2015). Instead, this stream of research argues that it is place-specific practices and collective routines as well as interests of local communities that define the 'industrial atmosphere' and character of innovative places. Storper (2018) underlines the *content* of networks is a key component of the relational infrastructure of places; thus, observing shared ideas and practices can reveal a broader regional context beyond technology or industry specific factors. These elements define the nature of localised buzz. While buzz is available to the 'insiders' of an industrial cluster, the type of information and signals that are transmitted via buzz may not be easily accessible or interpretable for those that are not part of the cluster. This is because economic actors who are co-located are exposed to a certain 'noise' that can be translated into patterns of signals by insiders who share the same

interpretative schemes and which might not be shared with outsiders (Grabher, 2002). In this sense, the propensity to exchange ideas goes beyond the level of human capital available and it is not simply defined by creative people moving in (Spigel, 2017). Instead, some places may offer a nurturing innovative environment characterised by a higher propensity to discuss new ideas that result from the intensity of discussion and vibrancy around topics related to innovation. In such a situation buzz can play an informative role – surrounding individuals with a context of diffused information where talk on innovation and new ideas is common.

These arguments are reinforced by the role buzz may play in socialisation of people into innovative and entrepreneurial behaviour. As Bathelt et al. (2004) point out, the information and communication flows in buzz also transmit descriptive norms to those who ‘mingle’ in the same economic and social context. Descriptive norms ‘are standards that develop out of our own observation of others’ behaviour’ (Cialdini and Trost, 1998: 152) that inform us of how other people typically act and can guide individuals’ decision-making processes. In this sense, looking at the dominant topics of conversation in a community allows us to explore a mechanism by which cultural prescriptions about expected behaviour are diffused to individual members of a community. An innovative atmosphere with descriptive norms on discussing and creating new ideas as common activities leads potential entrepreneurs to perceive their local environment as hospitable to entrepreneurial efforts. Linking to established research on the role of informal institutions in underpinning entrepreneurial processes, descriptive norms provide a particular logic of action (Lefebvre et al., 2015; Linan et al., 2011; Saxenian, 1996; Storper et al., 2015; Terjesen et al., 2016). Descriptive norms that are shared within a community create cognitive legitimacy for particular actions that are seen as typical, accepted and desirable (Stephan and Uhlaner, 2010). As such, they represent a type of ‘social proof heuristic’. These insights reflect the normative dimension of buzz, which is shaped by ‘shared cultural traditions and habits [, with] actors continuously contributing and benefiting from the diffusion of information, gossip and news by just “being there”’ Bathelt et al. (2004: 38). Similarly, Storper and Venables (2004) identify psychological motivation stimulating imitation and competition as one of the main features of buzz.

In line with these arguments, we posit the actual content of buzz is an important aspect in understanding this intangible dimension of places and the diffusion of entrepreneurial opportunities, as it may reflect the regional *zeitgeist* (Storper, 2018) of collective interests and practices. In some places people may be interested in innovation and have vibrant discussions on the topic whereas in others innovation is rarely spoken about. Therefore, we propose to analyse knowledge exchange more directly by examining to what extent people are ‘buzzing’ about innovation. Such buzz reflects an environment where ‘know how’, information, and opportunities for innovation are being diffused, offering a nurturing environment to nascent entrepreneurs. At the same time, we propose the presence of buzz about innovation implies that innovation is the norm, that it is considered normal to run with new ideas, to experiment and also to fail. This provides social legitimation and normative influence on innovation and experimentation, thereby fostering regional entrepreneurship activities.

Data and methods

Twitter data

In this study, we follow a novel approach to proxy the content of buzz within a region using data from Twitter. Previous research shows that people use Twitter in many different ways. This includes posting about daily routines, conversations between users as well as sharing information or reporting news (Java et al., 2007). A single user may have multiple motivations, discussing both professional and personal life topics at the same time (Java et al., 2007; Tata et al., 2017; Zhao and Rosson, 2009). In particular, our dataset is based on all geolocated tweets in the UK for 2014,

corresponding to almost two billion words across 180 million tweets.¹ The use of geolocated tweets allows to capture the exact location in terms of longitude and latitude GPS coordinates at the time of the tweet, which have been remapped to match NUTS3 region boundaries.

One aspect to consider when using big data from social media is the connection between the content of Tweets and their location.² A stream of recent research of online social networks has shown how these are connected to place and play an important role as local forms of communication (Gross et al., 2017; Stephens and Poorthuis, 2015). Tranos and Nijkamp (2013) explore the relationship between physical distance and virtual geography, highlighting the spatiality of the Internet. Similarly, Grabher (2002) argues that online interaction does not replace offline interaction but can reflect it or even ‘multiply’ it. More recent empirical research from Stephens and Poorthuis (2015) corroborates this view as in analysing Twitter data they found that local online networks are most effective at transmitting information. Accordingly, Twitter data have already been used to provide robust evidence and accurate measures on the spatial variation of a variety of topics, ranging from rates of heart-disease (Eichstaedt et al., 2015), fine-scale electric consumption (Deng et al., 2018), patents and GDP (Corradini, 2021; Indaco, 2020) to dominant demographic dimensions (Bokányi et al., 2016). In our dataset, as an example for the connection between the content of tweets and their location, the correlation between names of cities like Birmingham, Manchester or York and their NUTS3 location yields correlation scores higher than 0.9.

Another important element related to Twitter data is their representativeness. Recent studies have suggested the Twitter gender profile is close to the UK population as captured by the UK Census (Sloan et al., 2013). However, the average age of Twitter users is lower than the general UK population, though there is a significant tail of users aged 40 or above (Sloan, 2017). Data from social media are also skewed towards wealthy and highly educated users (Crampton et al., 2013; Wang et al., 2017). Accordingly, people from the creative industries tend to be over-represented across Twitter users (Sloan and Morgan, 2015). While these elements may create limitations in terms of representativeness of the UK population as a whole,³ the profile of Twitter users aligns well with the dominance of highly educated 25–45 year olds in UK early-stage entrepreneurship (Hart et al., 2014) and it is likely to overlap significantly with most individual engaged in buzz about innovation. Accordingly, Twitter data have been used to explore entrepreneurial networks across (Wang et al., 2017) and capture personality traits for the analysis of regional entrepreneurship culture in the USA (Obschonka et al., 2020).

Differences in the socio-economic profile of Twitter users are reflected in well-known digital divides and uneven spatial distribution of content creation, which is markedly higher in places with higher education and economic activity (Graham and Zook, 2011; Indaco, 2020; Stephens and Poorthuis, 2015; Tranos, 2020). This requires adjusting text analysis for regional differences in the level of Twitter activity, in combination with ancillary data (Crampton et al., 2013). Accordingly, our Twitter measures are normalised for the total number of words per NUTS3 region, and are analysed controlling for economic development, population density as well as education levels across locations. We discuss variables and controls for our analysis in the next section.

Variables

In line with previous research (Armington and Acs, 2002; Corradini and De Propriis, 2015; Lee et al., 2004), we measure entrepreneurship (ENTRY) by looking at the relative rate of new firms over the regional population. Hence, our dependent variable is calculated as the number of firms registered as founded in 2014 on the Business Structure Database (ONS, 2019), over NUTS3 population in 10 thousand people.

The key explanatory variable in the model is a novel measure of buzz about innovation (INNOBUZZ). This is defined as the relative count for the word ‘innovation’ over the total

number of words (in million) within geolocated tweets in each region across 2014.⁴ We note this measure is not confined to tweets from entrepreneurs or specifically dedicated to business activities. Conversely, it is defined over the population of geolocated tweets from all users to capture what is ‘in the air’. As discussed in the previous section, we underline our measure is normalised for the total number of words per NUTS3 region to take into account the uneven distribution in the level of content creation on social media (Graham and Zook, 2011; Indaco, 2020; Malik et al., 2015). Thus, INNOBUZZ is a relative measure reflecting the extent to which the innovation is a topic of interest across all conversations within a specific area. Focusing on the aggregate content of place-specific communication ecologies, our approach is intended to emphasise the spatial embeddedness that is inherent to local buzz. This is in contrast to the analysis of knowledge flows occurring exclusively within the structure of online exchanges, as reflected in the concept of virtual buzz (Bathelt and Turi, 2011; Tranos, 2020). In this sense, our measure offers a complementary perspective to previous research using Twitter data to explore entrepreneurial networks across digitally mediated interactions (Wang et al., 2017).

The straightforward operationalisation for INNOBUZZ is based on the word ‘innovation’ being used in a broad set of contexts, spanning across business and other areas of creative and innovative activities, as well as in more common situations. However, the use of a specific keyword exploits only limited information to capture the extent to which innovation is a topic of discussion in a specific place. Thus, we also explore an alternative approach considering all words within our corpus for each region. To this end, we follow a linguistics-based approach exploiting word co-occurrence analysis to identify other words that may be synonyms to ‘innovation’ or convey similar meanings. We start calculating the relative frequency of all words in each region weighted by their similarity to the word ‘innovation’.⁵ Following an established approach in natural language processing applications, these weights are obtained by applying the *Word2Vec* algorithm introduced by Mikolov et al. (2013a, 2013b). *Word2Vec* trains a shallow neural network to obtain ‘neural word embeddings’ reflecting n -dimensional numerical vector representations for each word. This allows to transform a large corpus of words as input into a vector space where words that are used in similar contexts are assigned to vectors close to each other in the vector space (for a methodological discussion, see Mikolov et al., 2013a). This approach captures word-to-word relations based on their different contexts providing a more accurate measure of semantic similarity between words.⁶ Thus, using pre-trained 300-dimensional vectors for 3 million words trained on a corpus of roughly 100 billion words from Google News (Mikolov et al., 2013b), we obtain similarity scores based on the cosine distance across word vectors numerically representing the relationship of each pre-trained word embedding to all others in the corpus.

A sample of words and their cosine similarity score is reported in Table 1. Based on this approach, we define INNOBUZZ_5 as the average regional cosine similarity of words whose cosine similarity score is higher than the 0.5 threshold in order to reduce potential noise from words with broader meanings. We also report and discuss results based on different threshold levels for robustness. In line with text analysis methods, the values are adjusted by the inverse

Table 1. Sample of words and associated cosine similarity to innovation.

Innovate	0.7123	Visionary	0.4757	IT	0.2916
Innovative	0.6424	Invention	0.4737	Inspiring	0.2614
Creativity	0.6018	Industry	0.4302	Thinking	0.2219
Technology	0.5921	Design	0.3867	Growing	0.1899
Excellence	0.5496	Business	0.3757	Internet	0.1645
Ingenuity	0.5047	Ideas	0.354	Brilliant	0.1553

word frequency formula⁷ in order to confer a smaller weight to words that are more common in the corpus before calculating an average score for each region.

We add a set of explanatory variables to account for established determinants of entrepreneurship reflecting human capital, knowledge creation and density of interaction in the region. These measures are based on socio-economic regional data from the UK Office for National Statistics (ONS). We include a measure of human capital (EDUCATION) following the established measure defined as the share of the working population (people aged 25–64) with tertiary education, which is equivalent to at least degree-level qualifications. To test our measure of INNOBUZZ against previous work trying to capture aspects that are ‘in the air’, we follow previous work on the role of the creative class and include an index of creativity labelled CREATIVE INDEX (Lee et al., 2004; Sleuwaegen and Boiardi, 2014). This is defined as the number of people working in creative industries identified following the ONS methodology (ONS, 2016) over the total working population (Lee and Rodríguez-Pose, 2014). We also add R&D intensity (at the NUTS2 level⁸) in logs to further capture scientific investment within the region. To proxy intensity of interaction, we use the common approach in the literature including a measure of population density in logs (POP_DENS), defined as the number of people per square kilometre at the NUTS3 level (Corradini and De Propris, 2015). We also test an alternative measure – labelled (FIRM_DENS) – reflecting the density of interaction in pooled labour markets, based on the number of firms over total population (Armington and Acs, 2002; Fang, 2020).

To control for other potential confounding factors, we add GDP per capita in purchasing power standard ($\log(\text{GDP}_{\text{ppp}})$) to control for regional factor endowments reflecting the availability of capital as well as the development of the region. Similarly, we control for GDP growth (GDP_GROWTH), measured as the annual rate of change in GDP, to account for demand effects as a result of a growing economy (Armington and Acs, 2002; Rodríguez-Pose and Hardy, 2015). Finally, we proxy diversity in regional economic activity including an index of sectoral diversification (DIVERSIF) defined as an inverse Herfindhal index, calculated using employment data across SIC industries at 4-digit level. As discussed in the following section, our measure of INNOBUZZ reflects imbalances across the UK, presenting moderate levels of correlation⁹ with control variables on regional development, education and density in labour markets. However, variance inflation factor (VIF) values suggest that multicollinearity is not a significant concern in this study. Descriptive statistics and VIF values for the variables in the sample are reported in Table 2.

Table 2. Descriptive statistics.

	N	Mean	SD	VIF
ENTRY	173	106.83	66.39	
INNOBUZZ	173	836.97	878.00	3.18
INNOBUZZ_5	173	11.67	10.97	1.51
LOG(GDP_{ppp})	173	10.17	0.39	3.95
GDP_GROWTH	173	10.00	2.98	1.21
LOG(R&D)	173	6.03	0.79	1.36
EDUCATION	168	35.20	9.30	2.13
CREATIVE_INDEX	173	3.29	2.56	2.59
LOG(POP_DENS)	173	5.68	1.63	1.34
FIRM_DENS	173	100.83	53.05	3.43
HERF	173	0.97	0.04	1.10

Results and discussion

Exploring the spatial distribution of INNOBUZZ, which is shown in Figure 1 (left panel), we find evidence of strong heterogeneity across the various NUTS3 regions in the UK. Values for INNOBUZZ are particularly high for the largest cities, starting from central London as well as Greater London, but also Manchester, the region around Birmingham, Nottingham as well as Belfast. The regions of Oxfordshire and Cambridgeshire are similarly included in the top decile as well as the Thames Estuary region, which has seen an increasing investment and development in creative industries in recent years. The regional distribution is skewed at the top, with values in central London being significantly higher than the average value. This reflects well-known differences across UK regional economies and the special case of the capital city. Thus, we find that the distribution of INNOBUZZ across UK regions is not random, but effectively reflects the agglomerated nature of economic activities in the UK. This variable presents higher values in areas defined by high population density and high GDP per capita, both elements characterising larger cities where buzz is expected to be more prominent (Bathelt et al., 2004; Storper and Venables, 2004). These stylised facts are quite similar when looking at INNOBUZZ_5, reported in the right panel of Figure 1. The spatial distribution of both measures also resembles that of new firms across UK regions. Exploring the correlation between ENTRY and INNOBUZZ as well as INNOBUZZ_5, we observe a strong correlation between the first pair ($r=0.78$); similarly, the association between ENTRY and INNOBUZZ_5 is also high ($r=0.52$), though we note increasing noise.

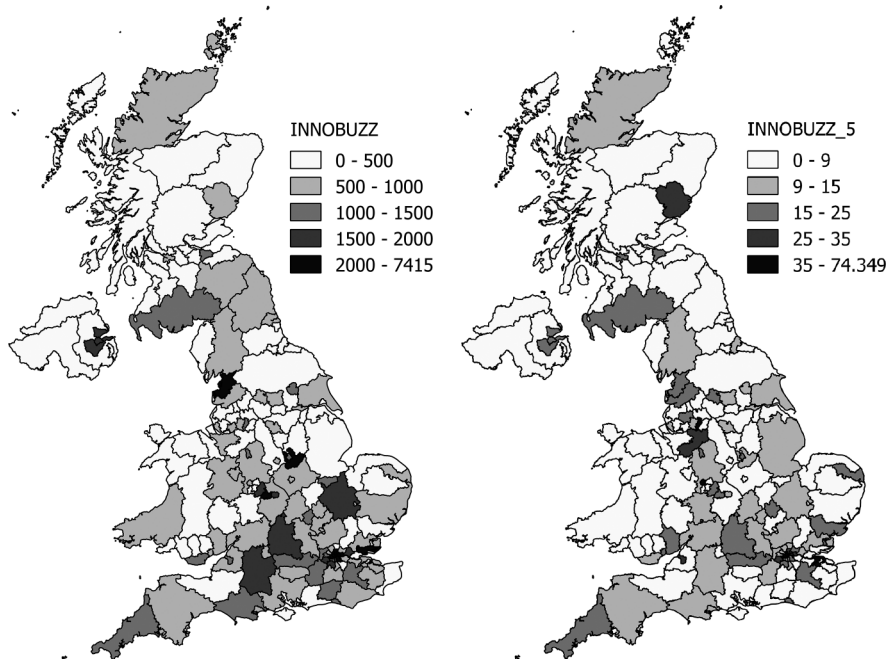


Figure 1. Spatial distribution of INNOBUZZ and INNOBUZZ_5 across NUTS3 regions.

Multivariate analysis

To disentangle the relationship between buzz about innovation and new firm creation from other confounding factors, we move to the results from OLS regression analysis. Estimates with cluster robust standard errors to control for heteroscedasticity are reported in Table 3.¹⁰ In column 1, we include controls for traditional determinants of entrepreneurship reflecting factor conditions, proxied by GDP, knowledge endowments in the form of human capital, and population density to capture proximity effects. Looking at the coefficient for INNOBUZZ, we observe a significant and positive effect on FIRM ENTRY, suggesting an innovative atmosphere may be an important determinant of entrepreneurship, *ceteris paribus*. In column 2, we also introduce R&D expenditure and the share of people working in creative industries (CREATIVE INDEX) to further control for knowledge creation across regions. Interestingly, we find a negative and significant effect of R&D on entrepreneurship in column 4. A possible interpretation of this finding is that the other variables, in particular GDP but also INNOBUZZ, may already be picking up any variance due to knowledge spillovers in the region so that the residual effect of R&D may effectively reflect the presence of large companies in the area, which has been associated with a negative effect on firm entry in previous literature (Armington and Acs, 2002). Indeed, removing GDP from the model yields a statistically significant and positive effect for our measure of R&D. In contrast, CREATIVE INDEX is not significant in the model. This may reflect the discussion in the literature

Table 3. OLS regression estimates.

	OLS					
	(1) ENTRY	(2) ENTRY	(3) ENTRY	(4) ENTRY	(5) ENTRY	(6) ENTRY
INNOBUZZ	0.019*** (0.005)	0.018*** (0.005)	0.008** (0.003)			
INNOBUZZ_5				0.424** (0.208)	0.377* (0.22)	0.34*** (0.114)
LOG(GDP _{ppp})	79.67*** (21.119)	88.908*** (22.274)	25.964*** (9.673)	104.165*** (23.066)	112.308*** (25.772)	30.312*** (9.048)
GDP_GROWTH	1.228 (0.94)	1.251 (0.957)	1.522* (0.847)	1.368 (0.968)	1.428 (0.967)	1.636* (0.856)
LOG(R&D)		-6.987** (3.513)	-6.752** (2.995)		-8.979** (3.604)	-7.365** (2.978)
EDUCATION	1.34*** (0.445)	1.362*** (0.502)	0.373 (0.421)	1.413*** (0.452)	1.311** (0.51)	0.337 (0.419)
CREATIVE_INDEX		-0.057 (1.587)	4.336*** (0.864)		0.966 (1.924)	4.708*** (0.856)
LOG(POP_DENS)	5.311*** (1.443)	5.075*** (1.446)		6.11*** (1.536)	5.346*** (1.501)	
FIRM_DENS			0.754*** (0.09)			0.782*** (0.08)
DIVERSIF		3.494 (75.051)	-59.688 (49.631)		-16.077 (76.515)	-70.225 (49.048)
F test	20.96***	17.97***	98.05***	18.87***	15.57***	102.78***
R ²	0.79	0.80	0.89	0.77	0.78	0.89
N	168	168	168	168	168	168

Cluster robust errors in parentheses, * $P < 0.1$ ** $P < 0.05$ *** $P < 0.01$.

on the complementary aspects of traditional measures of human capital based on education and creative occupations (Boschma and Fritsch, 2009). However, when we use firm density instead of population density in column 3, the results for CREATIVE_INDEX turn positive and statistically significant in line with previous research, whilst EDUCATION is no longer significant. This likely reflects the higher correlation between these variables and population as opposed to firm density. GDP GROWTH is also significant and positive, in line with previous evidence on the role of demand growth across regional economic systems (Armington and Acs, 2002). Finally, our measure of sectoral dispersion, DIVERSIF, is not statistically significant, reflecting the sensitivity of regional diversity measures as found in previous studies on entrepreneurship (Bishop, 2012; Colombelli and Quatraro, 2018). Even after controlling for these elements, results for INNOBUZZ remain consistent.

We report the results based on the alternative measure INNOBUZZ_5 in columns 4 to 6 of Table 3. Following this approach does not alter our findings on the relationship between buzz about innovation and new firm creation. Coefficients for INNOBUZZ_5 are found to be positive and statistically significant, though we note this holds only at the .10 level for the model reported in column 5. We argue this point to a trade-off in the construction of this index. On one side, this measure is less sensitive to the occurrence of specific keywords within the Twitter corpus. At the same time, this implies it captures a broader conceptualisation of buzz about innovation, including aspects that are more connected to other variables for localised knowledge. We find further support for this interpretation exploring INNOBUZZ across different threshold levels of word similarity. We report in Table 4 estimates for INNOBUZZ, alongside coefficients based on a 0.6, 0.5, 0.3 threshold as well as including all words in the Twitter corpus. We observe that as the weighted average underlying the index encompasses a greater set of words, the precision of the estimates reduces.¹¹ Whilst the results remain statistically significant, we note the magnitude of the coefficients and the R^2 are also lower.

Overall, these results confirm previous evidence on the knowledge spillover theory of entrepreneurship, pointing to the level of economic development, human capital in its various forms, and the

Table 4. Estimates across different thresholds for INNOBUZZ.

	OLS				
	(1) ENTRY	(2) ENTRY	(3) ENTRY	(4) ENTRY	(5) ENTRY
INNOBUZZ	0.018*** (0.005)				
INNOBUZZ_6		2.722** (1.195)			
INNOBUZZ_5			0.377* (0.22)		
INNOBUZZ_3				0.011* (0.006)	
INNOBUZZ_ALL					0.002* (0.001)
<i>F</i> test	20.96***	16.64***	15.57***	16.79***	16.78***
R^2	0.80	0.79	0.78	0.78	0.78
<i>N</i>	168	168	168	168	168

Cluster robust errors in parentheses, * $P < 0.1$ ** $P < 0.05$ *** $P < 0.01$. Controls included are: $\log(\text{GDP}_{\text{PPP}})$, GDP growth, $\log(\text{R\&D})$, EDUCATION, CREATIVE_INDEX, $\log(\text{POP_DENS})$.

intensity of localised interactions as significant determinants of entrepreneurial processes (Acs et al., 2012; Qian et al., 2013). Yet, even whilst controlling for other traditional determinants of firm entry including creative industries (Boschma and Fritsch, 2009), our measures of buzz about innovation remain consistently significant. We argue that this points to INNOBUZZ reflecting differences in norms, attitudes and propensity to discuss topics related to innovation across places with a similar economic structure. An alternative interpretation could be that this variable is reflecting other aspects of human capital and knowledge creation that are not effectively captured by traditional variables by such as education level, economic development or employment in creative industries. Thus, in this interpretation, INNOBUZZ allows us to look into the functional structure of regions by bringing to the fore the extent of unobserved innovative content in the regional economy.

Robustness analysis

In this section, we complement our main findings considering the heterogeneous nature of entrepreneurial activities across sectors and possible spatial effects for buzz about innovation along with further robustness checks. We start by exploring our model looking at new firms in manufacturing (MANUF) as opposed to service (SERVICE) industries. Results are reported in columns 1 to 4 in Table 5. We note that the results for manufacturing sectors are not significant. Here, we also point out a negative coefficient for density, suggesting these industries may not be connected to proximity effects of buzz. Thus, we also look at as well as an alternative classification separating innovative to traditional sectors. Accordingly, we consider firm entry in low knowledge intensive (L_KI) and high knowledge intensive (H_KI) industries¹² in columns 5 to 8. For low knowledge intensive sectors results are significant only when looking at INNOBUZZ, but not for INNOBUZZ_5. Even in this case, the magnitude of the coefficient is much smaller than in the case of knowledge intensive sectors. In line with the literature on knowledge-based entrepreneurial activity (Audretsch and Belitski, 2013; Qian et al., 2013), these findings suggest the atmosphere captured by INNOBUZZ is mostly conducive to new firms in knowledge intensive industries.

We also explore potential spatial lags in the effect of buzz. Results are reported in Table 6 for INNOBUZZ as well as INNOBUZZ_5 and INNOBUZZ_ALL. Spatial thresholds are defined at 50 and 100 km from centroids of NUTS3 regions. In line with the literature on knowledge spillovers (Fang, 2020; Jaffe et al., 1993; Sonn and Storper, 2008), we find the presence of spatial spillovers across different measures of INNOBUZZ. However, these are quite localised reflecting insights on the localised nature of buzz in supporting exchange of uncodified knowledge (Storper and Venables, 2004). The size of the effect is much reduced at a 50 km distance, and spatial lags are no longer significant once we move beyond the 50 km threshold.

Finally, we performed two further robustness checks. Previous studies have highlighted the relationship between more wealthy and educated demographics and the uneven spatial distribution of content creation (Crampton et al., 2013; Sloan and Morgan, 2015), as reflected in recent studies showing Twitter activity to be a valid proxy for GDP (Indaco, 2020). Thus, it is important to check INNOBUZZ may not be capturing confounding effects due to the volume of Tweets, which may represent a significant determinant of new business creation on its own. By design, our measure of INNOBUZZ is based on the relative frequency of keywords in the Twitter corpus and should not reflect overall Twitter activity. We also control for GDP level and growth in our analysis. However, we further test for this including an additional variable (TWEET_DENSITY), defined as the number of Tweets per population. Results are reported with and without INNOBUZZ and INNOBUZZ_5 variables for models with population and firm density measures in columns 1 to 6 of Table A2 (in the Appendix). TWEET_DENSITY does not yield significant coefficients, which likely reflects the correlation between Twitter volume

Table 5. Estimates across manufacturing and service, low and high knowledge intensive sectors.

	OLS							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	MANUF	SERVICE	MANUF	SERVICE	L_KI	H_KI	L_KI	H_KI
INNOBUZZ	0.001 (0.001)	0.017*** (0.004)	-0.015 (0.051)	0.391** (0.207)	0.006** (0.002)	0.012*** (0.003)	0.085 (0.091)	0.292** (0.152)
INNOBUZZ_5								
LOG(GDP _{PPP})	9.023** (4.153)	79.885*** (19.505)	10.959*** (3.9)	101.35*** (22.596)	37.911*** (12.217)	50.997*** (11.055)	45.593*** (12.755)	66.715*** (14.257)
GDP_GROWTH	0.239 (0.176)	1.012 (0.893)	0.241 (0.174)	1.187 (0.91)	1.124 (0.71)	0.128 (0.383)	1.171 (0.711)	0.257 (0.392)
LOG(R&D)	-2.484*** (0.89)	-4.502 (3.05)	-2.63*** (0.861)	-6.349** (3.12)	-5.256** (2.086)	-1.731 (1.783)	-5.894*** (2.149)	-3.085* (1.802)
EDUCATION	0.039 (0.089)	1.323*** (0.463)	0.033 (0.094)	1.278*** (0.462)	0.522 (0.341)	0.84*** (0.228)	0.505 (0.344)	0.807** (0.231)
CREATIVE_INDEX	0.293 (0.295)	-0.35 (1.398)	0.381 (0.348)	0.585 (1.671)	0.25 (0.874)	-0.308 (0.93)	0.589 (0.967)	0.376 (1.145)
LOG(POP_DENS)	-1.422*** (0.437)	6.497*** (1.301)	-1.365*** (0.424)	6.71*** (1.379)	1.87** (0.938)	3.205*** (0.726)	1.989** (0.952)	3.357*** (0.779)
DIVERSIF	-12.193 (15.786)	15.686 (69.986)	-13.21 (15.92)	-2.867 (69.72)	-34.021 (49.668)	37.515 (29.133)	-39.932 (49.725)	23.855 (31.101)
F test	4.64***	21.46***	4.38***	19.42***	5.97***	3.42***	5.84***	30.61***
R ²	0.34	0.82	0.34	0.80	0.64	0.85	0.63	0.83
N	168	168	168	168	168	168	168	168

Cluster robust errors in parentheses, * P<0.1 ** P<0.05 *** P<0.01.

Table 6. Estimates with spatial lags across different threshold distances.

OLS						
	(1)	(2)	(3)	(4)	(5)	(6)
	ENTRY	ENTRY	ENTRY	ENTRY	ENTRY	ENTRY
INNOBUZZ	0.018*** (0.004)	0.018*** (0.004)				
INNOBUZZ_50 km	0.001*** (0.001)	0.001*** (0.001)				
INNOBUZZ_100 km		0.000 (0.000)				
INNOBUZZ_5			0.436** (0.218)	0.436** (0.219)		
INNOBUZZ_5_50 km			0.109*** (0.019)	0.108*** (0.025)		
INNOBUZZ_5_100 km				0.001 (0.013)		
INNOBUZZ_ALL					0.002** (0.001)	0.002** (0.001)
INNOBUZZ_ALL_50 km					0.0001*** (0.000)	0.0001*** (0.000)
INNOBUZZ_ALL_100 km						0.000 (0.000)
F test	23***	22.74***	19.59***	20.15***	21.01***	20.92***
R ²	0.83	0.83	0.81	0.81	0.82	0.82
N	168	168	168	168	168	168

Cluster robust errors in parentheses, * $P < 0.1$ ** $P < 0.05$ *** $P < 0.01$. Controls included are: $\log(\text{GDP}_{\text{PPP}})$, GDP growth, $\log(\text{R\&D})$, EDUCATION, CREATIVE_INDEX, $\log(\text{POP_DENS})$.

activity and GDP variables in the model. At the same time, our main results are robust to the inclusion of this control.

Results are also robust to the inclusion of a dummy variable for London, whose significant weight in the UK economy may affect the average values in the sample (Lee and Rodríguez-Pose, 2014). Estimates are reported in columns 7 to 10 of Table A2 in the Appendix. While education and population density variables are no longer significant, reflecting their high correlation with the London region, results for both INNOBUZZ or INNOBUZZ_5 remain consistent in both models with population and firm density measures.

Conclusions

This paper contributes to our understanding of the role that regional context plays in entrepreneurial action offering an exploratory analysis on the opportunities available in big data to listen to the buzz – the vibrancy and knowledge exchange propensity of localised informal communication flows. Renewing theoretical insights indicating buzz encapsulates more than just density of interaction (Asheim et al., 2007; Storper and Venables, 2004), we propose looking at the *content* of buzz to capture the presence of a localised innovative atmosphere. In particular, we argue the presence of buzz about innovation may comprise an informative dimension conducive to exchange of novel ideas and entrepreneurial opportunities as well as a normative dimension of innovative attitudes and behaviour encouraging entrepreneurial processes.

We provide a novel quantitative measurement of buzz based on local language patterns and natural language processing and test its relationship with entrepreneurship rates using data on 180 million geolocated Tweets across NUTS3 regions in the UK for the year 2014. Controlling for established determinants of firm entry such as economic development, traditional measures of human capital and density, we find that regions which are characterised by higher levels of buzz reflecting an innovative atmosphere – more discussion and vibrancy about topics related to innovation – do have higher levels of entrepreneurship. In line with a knowledge-based view of localisation economies (Fang, 2020; Malmberg and Maskell, 2002), we also find buzz about innovation matters more for encouraging start-ups in high knowledge intensive sectors than in low knowledge intensive sectors and its effect is markedly localised, dissipating beyond the 50 km threshold.

Our research underlines that by listening to the buzz we can deepen our understanding of how co-location of economic actors matters. It is not just their co-presence, but the content of their informal communication that contributes to the emergence of ‘innovative milieus’. Looking at the content of buzz may also reveal a discursive dimension that has consequences for what is ‘in the air’ in localized communities. Similarly, listening to the buzz may allow us to capture normative aspects in the relational infrastructure revealing social norms that define the innovative character of places. By highlighting the role of an innovative atmosphere in the exchange of ideas and knowledge conducive to the diffusion and exploitation of entrepreneurial opportunities, our results could inform the literature on informal institutions within entrepreneurial ecosystems (Audretsch and Belitski, 2013; Qian, 2018; Spigel, 2017). At the same time, this paper advances theories of buzz as intangible infrastructure shaping both knowledge spillovers and social norms by connecting research on creativity and innovative endowments of regions (Boschma and Fritsch, 2009; Florida, 2012; Sleuwaegen and Boiardi, 2014) with economic geography perspectives on the relational dimension of localised learning dynamics (Bathelt et al., 2004; Gertler, 2003; Storper and Venables, 2004).

Our results build on previous insights on the potential in using big data from social media for spatial analysis (Graham and Zook, 2011; Indaco, 2020; Tranos, 2020; Wang et al., 2017). In particular, by providing a novel approach based on text analysis of geo-located tweets to capture the

content of buzz, the paper points to opportunities in merging big data, applied linguistics and natural language processing methods in economic geography as well as entrepreneurship research (Corradini, 2021; Obschonka et al., 2020). We note this interdisciplinary approach may be especially valuable for exploring informal and intangible aspects of localised regional structures. This may be also useful for innovation policy, as such elements have been difficult to capture with traditional and survey data.

At the same time, the results presented must be interpreted considering various limitations of this study. Similarly to recent papers exploring big data from social media (Casadei and Lee, 2020; Obschonka et al., 2020), the lack of longitudinal Twitter data does not allow us to control for potential unobserved heterogeneity across regions or shed light on the dynamics of buzz. Thus, while our results indicate regions characterised by a stronger buzz about innovation - as reflected by language patterns in social media - also show higher entrepreneurship rates, we cannot address causality in their relationship. This is a critical element that requires further investigation. Indeed, our contribution is explorative in nature and several questions remain open. Potential routes for future research could include using big data to capture more nuanced insights on the specific mechanisms within buzz, disentangling its role as knowledge exchange infrastructure from the effects of localised descriptive norms. Also, our analysis would benefit from complementary perspectives exploring the evolution and trends of topics and discussions across different time spans and lower levels of spatial aggregation (Bathelt and Henn, 2014; Crampton et al., 2013). Future studies may also investigate linkages and potential synergies between the content of local buzz defined within spatially embedded communities and virtual buzz across digitally mediated interactions (Tranos, 2020; Wang et al., 2017). Finally, while our approach allows to consider all the information within the word corpus available, it relies on prior selection of keywords for the analysis. Further research may investigate alternative methods from natural language processing such as unsupervised learning algorithms and recent advances in topic modelling. More broadly, we point to significant opportunities from applied linguistics and big data analytics for the development of novel indicators to track tangible and intangible dimensions in the geography of innovation. This could allow to understand their evolution and design more comprehensive policies to stimulate their development.

Acknowledgments

The statistical data used here is from the Office of National Statistics (ONS) and is Crown copyright and reproduced with the permission of the controller of HMSO and Queens Printer for Scotland. The use of the ONS statistical data in this work does not imply the endorsement of the ONS in relation to the interpretation or analysis of the statistical data. The analysis upon which this paper is based uses research datasets which may not exactly reproduce National Statistics aggregates.


Declaration of conflicting interests

The authors 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: This research was funded by the British Academy/Leverhulme Trust under the Small Research Grants scheme (SG171674). Dr Corradini also acknowledges funding by a Leverhulme Trust research fellowship (RF-2020-505\7).

ORCID iD

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

Notes

1. Our data are based on a spatial aggregation of a Twitter corpus collected as part of the Digging into Data Challenge initiative, as discussed by Grieve et al. (2019). Whilst presenting some limitations in volume cap, Twitter data can be accessed through the recently introduced Academic Research product track using the Twitter API.
2. Twitter only allows for original content to be geo-tagged; this means that re-tweets are not geo-tagged and are not used in this analysis (Sloan and Morgan, 2015).
3. Considering differences between those who use Twitter and the sub-section of geolocated tweets, Sloan and Morgan (2015) point to limited differences in gender, age and occupation category.
4. Focusing on tweets across 1 year, we aim to smooth out short-term trends and capture topics of conversation beyond the 'here and now' (Crampton et al., 2013).
5. For robustness, we also calculated our INNOBUZZ measure using all words from the stem 'innov' (INNOBUZZ_S). Similarly, we have also calculated a composite index based on the weights for each individual word with stem 'innov' aggregated at NUTS3. Results are robust to these approaches and are reported in the Appendix, Table A3.
6. Synonyms, for example, have similar word vectors and are closer to each other in vector space because they are used in similar contexts.
7. This is defined as follows: $TF(\text{term frequency}) * IDF(\text{inverse document frequency})$.
8. Results are consistent when this variable is replaced with the number of patents per capita, at NUTS3 level.
9. Correlations among all variables are reported in the Appendix, Table A1.
10. Constant terms for all regressions are not reported for confidentiality reasons due to the use of restricted access ONS data.
11. Differences in the size of coefficients for INNOBUZZ and weighted INNOBUZZ measures reflect the variation in the set of words included and methodology to calculate these variables.
12. These groups are defined following the classification at NACE 2-digit level by EUROSTAT.

References

- Acs ZJ, Audretsch DB, Braunerhjelm P, et al. (2012) Growth and entrepreneurship. *Small Business Economics* 39(2): 289–300.
- Acs ZJ, Braunerhjelm P, Audretsch DB, et al. (2009) The knowledge spillover theory of entrepreneurship. *Small Business Economics* 32(1): 15–30.
- Armington C and Acs ZJ (2002) The determinants of regional variation in new firm formation. *Regional Studies* 36(1): 33–45.
- Asheim B, Coenen L and Vang J (2007) Face-to-face, buzz, and knowledge bases: Sociospatial implications for learning, innovation, and innovation policy. *Environment and Planning C: Government and Policy* 25(5): 655–670.
- Audretsch DB and Belitski M (2013) The missing pillar: The creativity theory of knowledge spillover entrepreneurship. *Small Business Economics* 41(4): 819–836.
- Audretsch DB, Bönte W and Keilbach M (2008) Entrepreneurship capital and its impact on knowledge diffusion and economic performance. *Journal of Business Venturing* 23(6): 687–698.
- Audretsch DB and Keilbach M (2008) Resolving the knowledge paradox: Knowledge-spillover entrepreneurship and economic growth. *Research Policy* 37(10): 1697–1705.
- Audretsch DB and Lehmann EE (2005) Does the knowledge spillover theory of entrepreneurship hold for regions? *Research Policy* 34(8): 1191–1202.
- Bathelt H and Glückler J (2003) Toward a relational economic geography. *Journal of Economic Geography* 3(2): 117–144.

- Bathelt H and Henn S (2014) The geographies of knowledge transfers over distance: Toward a typology. *Environment and Planning A* 46(6): 1403–1424.
- Bathelt H, Malmberg A and Maskell P (2004) Clusters and knowledge: Local buzz, global pipelines and the process of knowledge creation. *Progress in Human Geography* 28(1): 31–56.
- Bathelt H and Turi P (2011) Local, global and virtual buzz: The importance of face-to-face contact in economic interaction and possibilities to go beyond. *Geoforum* 42(5): 520–529.
- Bishop P (2012) Knowledge, diversity and entrepreneurship: A spatial analysis of new firm formation in Great Britain. *Entrepreneurship and Regional Development* 24(7–8): 641–660.
- Bokányi E, Kondor D, Dobos L, et al. (2016) Race, religion and the city: Twitter word frequency patterns reveal dominant demographic dimensions in the United States. *Palgrave Communications* 2: 16010.
- Boschma RA and Fritsch M (2009) Creative class and regional growth: Empirical evidence from seven European countries. *Economic Geography* 85(4): 391–423.
- Camagni R (1991) *Innovation Networks: Spatial Perspectives*. London/New York: Bellhaven Press.
- Casadei P and Lee N (2020) Global cities, creative industries and their representation on social media: A micro-data analysis of Twitter data on the fashion industry. *Environment and Planning A* 52(6): 1195–1220. DOI: 10.1177/0308518X20901585.
- Cialdini RB and Trost MR (1998) Social influence: Social norms, conformity, and compliance. In: Gilbert D, Fiske J and Lindzey G (eds) *The Handbook of Social Psychology*. New York: McGraw-Hill, 151–192.
- Colombelli A and Quatraro F (2018) New firm formation and regional knowledge production modes: Italian evidence. *Research Policy* 47(1): 139–157.
- Corradini C (2021) The geography of innovation as reflected by social media. *Environment and Planning A: Economy and Space* 53(2): 227–229. DOI: 10.1177/0308518X20961107.
- Corradini C and De Propriis L (2015) Technological diversification and new innovators in european regions: Evidence from patent data. *Environment and Planning A: Economy and Space* 47(10): 2170–2186.
- Crampton JW, Graham M, Poorthuis A, et al. (2013) Beyond the geotag: Situating ‘big data’ and leveraging the potential of the geoweb. *Cartography and Geographic Information Science* 40(2): 130–139.
- Deng C, Lin W, Ye X, et al. (2018) Social media data as a proxy for hourly fine-scale electric power consumption estimation. *Environment and Planning A: Economy and Space* 50(8): 1553–1557.
- 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.
- Fang L (2020) Agglomeration and innovation: Selection or true effect? *Environment and Planning A: Economy and Space* 52(2): 423–448.
- Florida R (2002) *The Rise of the Creative Class*. New York: Basic Books.
- Florida R (2012) *The Rise of the Creative Class Revisited*. New York: Basic Books.
- Fotopoulos G (2013) On the spatial stickiness of UK new firm formation rates. *Journal of Economic Geography* 14(3): 651–679.
- Fritsch M and Falck O (2007) New business formation by industry over space and time: A multidimensional analysis. *Regional Studies* 41(2): 157–172.
- Gertler MS (2003) Tacit knowledge and the economic geography of context, or the undefinable tacitness of being (there). *Journal of Economic Geography* 3(1): 75–99.
- Glaeser E (2004) Book Review of Florida’s *The Rise of the Creative Class*. Available at: <http://posteconomics.harvard.edu/faculty/glaeser/papers.html>.
- Glaeser EL, Kerr WR and Ponzetto GAM (2010) Clusters of entrepreneurship. *Journal of Urban Economics* 67(1): 150–168.
- Grabher G (2002) Cool projects, boring institutions: Temporary collaboration in social context. *Regional Studies* 36(3): 205–214.
- Graham M and Zook M (2011) Visualizing global cyberscapes: Mapping user-generated placemarks. *Journal of Urban Technology* 18(1): 115–132.
- Grieve J, Montgomery C, Nini A, et al. (2019) Mapping lexical dialect variation in British English using twitter. *Frontiers in Artificial Intelligence* 2: 11.
- Gross AJ, Murthy D, and Varshney LR (2017) Pace of life in cities and the emergence of town tweeters. *Sage Open* 7(4): 2158244017745113.

- Hart M, Levie J, Bonner K, et al. (2014) Global Entrepreneurship Monitor 2014 United Kingdom Report. GEM UK 2014.
- Indaco A (2020) From twitter to GDP: Estimating economic activity from social media. *Regional Science and Urban Economics* 85: 103591.
- Jaffe AB, Trajtenberg M and Henderson R (1993) Geographic localization of knowledge spillovers as evidenced by patent citations. *The Quarterly Journal of Economics* 108(3): 577–598.
- Java A, Song X, Finin T, et al. (2007) Why we twitter: understanding microblogging usage and communities. In: proceedings of the 9th WebKDD and 1st SNA-KDD 2007 workshop on Web mining and social network analysis, pp. 56–65.
- Lawson C and Lorenz E (1999) Collective learning, tacit knowledge and regional innovative capacity. *Regional Studies* 33(4): 305–317.
- Lee SY, Florida RL and Acs ZJ (2004) Creativity and entrepreneurship: A regional analysis of new firm formation. *Regional Studies* 38(8): 879–891.
- Lee N and Rodríguez-Pose A (2014) Creativity, cities, and innovation. *Environment and Planning A: Economy and Space* 46(5): 1139–1159.
- Lefebvre V, Radu Lefebvre M and Simon E (2015) Formal entrepreneurial networks as communities of practice: A longitudinal case study. *Entrepreneurship & Regional Development* 27(7–8): 500–525.
- Linan F, Santos FJ and Fernandez J (2011) The influence of perceptions on potential entrepreneurs. *International Entrepreneurship and Management Journal* 7(3): 373–390.
- Malik M, Lamba H, Nakos C, et al. (2015) Population bias in geotagged tweets. In: proceedings of the international AAAI conference on web and social media, Vol. 9, No. 1.
- Malmberg A and Maskell P (2002) The elusive concept of localization economies: Towards a knowledge-based theory of spatial clustering. *Environment and Planning A: Economy and Space* 34(3): 429–449.
- Marshall A (1890) *Principles of Economics*. London and New York: Macmillan.
- Mikolov T, Chen K, Corrado G, et al. (2013a) Efficient estimation of word representations in vector space. In: proceedings of international conference on learning representations (ICLR).
- Mikolov T, Sutskever I, Chen K, et al. (2013b) Distributed representations of words and phrases and their compositionality. arXiv preprint arXiv:1310.4546.
- Obschonka M, Lee N, Rodríguez-Pose A, et al. (2020) 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* 55(3): 567–588. DOI: 10.1007/s11187-019-00204-2.
- Office for National Statistics (2016) Creative industries economic estimates methodology. Available at: <https://www.gov.uk/government/publications/creative-industries-economic-estimates-methodology>.
- Office for National Statistics (2019) Business structure database, 1997–2018: Secure access. [data collection]. 10th Edition. UK Data Service. SN:6697, <http://doi.org/10.5255/UKDA-SN-6697-10>.
- Polanyi M (1967) *The Tacit Dimension*. New York: Anchor Books.
- Qian H (2018) Knowledge-Based regional economic development: A synthetic review of knowledge spillovers, entrepreneurship and entrepreneurial ecosystems. *Economic Development Quarterly* 32(2): 163–176.
- Qian H and Acs ZJ (2013) An absorptive capacity theory of knowledge spillover entrepreneurship. *Small Business Economics* 40(2): 185–197.
- Qian H, Acs ZJ and Stough RR (2013) Regional systems of entrepreneurship: The nexus of human capital, knowledge and new firm formation. *Journal of Economic Geography* 13(4): 559–558.
- Qian H and Jung H (2017) Solving the knowledge filter puzzle: Absorptive capacity, entrepreneurship and regional development. *Small Business Economics* 48(1): 99–114.
- Rodríguez-Pose A and Hardy D (2015) Cultural diversity and entrepreneurship in England and Wales. *Environment and Planning A: Economy and Space* 47(2): 392–411.
- Saxenian A (1996) *Regional Advantage*. Cambridge, MA: Harvard University Press.
- Schwab A and Zhang Z (2019) A new methodological frontier in entrepreneurship research: Big data studies. *Entrepreneurship Theory and Practice* 43(5): 843–854.
- Sleuwaegen L and Boiardi P (2014) Creativity and regional innovation: Evidence from EU regions. *Research Policy* 43(9): 1508–1522.
- Sloan L (2017) Who tweets in the United Kingdom? Profiling the twitter population using the British social attitudes survey 2015. *Social Media + Society* 3(1): 2056305117698981.

- Sloan L and Morgan J (2015) Who tweets with their location? Understanding the relationship between demographic characteristics and the use of geoservices and geotagging on twitter. *PLoS ONE* 10(11): e0142209.
- Sloan L, Morgan J, Housley W, et al. (2013) Knowing the Tweeters: Deriving sociologically relevant demographics from Twitter. *Sociological Research Online* 18(3): 1–11.
- Sonn JW and Storper M (2008) The increasing importance of geographical proximity in knowledge production: An analysis of US patent citations, 1975–1997. *Environment and Planning A: Economy and Space* 40(5): 1020–1039.
- Spigel B (2017) The relational organization of entrepreneurial ecosystems. *Entrepreneurship Theory and Practice* 41(1): 49–72.
- Stephan U and Uhlaner LM (2010) Performance-based vs socially supportive culture: A cross-national study of descriptive norms and entrepreneurship. *Journal of International Business Studies* 41(8): 1347–1364.
- Stephens M and Poorthuis A (2015) Follow thy neighbor: Connecting the social and the spatial networks on Twitter. *Computers, Environment and Urban Systems* 53: 87–95.
- Storper M (1997) *The Regional World: Territorial Development in A Global Economy*. New York/London: Guilford Press.
- Storper M (2018) Regional innovation transitions. In: Glückler J, Suddaby R and Lenz R (eds) *Knowledge and Institutions*. Cham: Springer, 197–225.
- Storper M, Kemeny T, Makarem N, et al. (2015) *The Rise and Fall of Urban Economies: Lessons from San Francisco and Los Angeles*. Stanford, CA: Stanford University Press.
- Storper M and Scott AJ (2009) Rethinking human capital, creativity and urban growth. *Journal of Economic Geography* 9(2): 147–167.
- Storper M and Venables AJ (2004) Buzz: Face-to-face contact and the urban economy. *Journal of Economic Geography* 4(4): 351–370.
- Tata A, Martinez DL, Garcia D, et al. (2017) The psycholinguistics of entrepreneurship. *Journal of Business Venturing Insights* 7: 38–44.
- Terjesen S, Hessels J and Li D (2016) Comparative international entrepreneurship: A review and research agenda. *Journal of Management* 42(1): 299–344.
- Tranos E (2020) Social network sites and knowledge transfer: An urban perspective. *Journal of Planning Literature* 35(4): 408–422.
- Tranos E and Nijkamp P (2013) The death of distance revisited: Cyber-place, physical and relational proximities. *Journal of Regional Science* 53(5): 855–873.
- Urbano D, Aparicio S and Audretsch D (2019) Twenty-five years of research on institutions, entrepreneurship, and economic growth: What has been learned? *Small Business Economics* 53(1): 21–49.
- Van Meeteren M and Poorthuis S (2018) Christaller and “big data”: Recalibrating central place theory via the geoweb. *Urban Geography* 39(1): 122–248.
- Wang F, Mack EA and Maciejewski R (2017) Analyzing entrepreneurial social networks with big data. *Annals of the American Association of Geographers* 107(1): 130–150.
- Zhao D and Rosson MB. (2009) How and why people Twitter: The role that micro-blogging plays in informal communication at work. In proceedings of the ACM 2009 international conference on Supporting group work, pp. 243–252.

Appendix

Table A1. Correlation matrix.

	1	2	3	4	5	6	7	8	9	10
ENTRY	1									
INNOBUZZ	0.78	1.00								
INNOBUZZ_5	0.52	0.61	1.00							
LOG(GDP _{ppp})	0.85	0.78	0.53	1.00						
GDP_GROWTH	0.40	0.33	0.17	0.41	1.00					
LOG(R&D)	0.31	0.28	0.20	0.48	0.20	1.00				
EDUCATION	0.65	0.51	0.34	0.63	0.25	0.31	1.00			
CREATIVE_INDEX	0.63	0.61	0.44	0.62	0.23	0.30	0.66	1.00		
LOG(POP_DENS)	0.44	0.39	0.36	0.35	0.19	0.08	0.22	0.44	1.00	
FIRM_DENS	0.90	0.72	0.45	0.82	0.33	0.35	0.58	0.48	0.11	1.00
DIVERSIF	0.12	0.09	0.11	0.11	0.09	0.00	0.12	0.26	0.07	0.11

Table A2. OLS results with Twitter density controls and London_Area fixed effects.

	OLS									
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
	ENTRY	ENTRY	ENTRY	ENTRY	ENTRY	ENTRY	ENTRY	ENTRY	ENTRY	ENTRY
TWEET_DENSITY	0.025 (15.155)	11.385 (7.134)	-1.618 (15.075)	10.24 (7.372)	-0.577 (15.158)	10.591 (7.132)				
INNOBUZZ			0.018*** (0.005)	0.007** (0.003)			0.014*** (0.004)	0.004* (0.003)		
INNOBUZZ_5					0.377* (0.218)	0.317*** (0.115)			0.379** (0.187)	0.252** (0.102)
LOG(GDP _{PPP})	117.143*** (24.121)	27.902*** (9.231)	89.565*** (18.605)	20.787** (10.216)	112.555*** (22.61)	24.683** (9.501)	92.113*** (21.067)	31.643*** (8.341)	109.571*** (23.239)	33.516*** (8.109)
GDP_GROWTH	1.342 (0.972)	1.658* (0.857)	1.237 (0.952)	1.603* (0.857)	1.423 (0.968)	1.712** (0.866)	0.405 (0.924)	0.668 (0.752)	0.498 (0.944)	0.734 (0.756)
LOG(R&D)	-9.203** (3.632)	-7.167** (3.024)	-7.032* (3.569)	-6.392** (3.032)	-8.996** (3.639)	-6.952** (3.015)	3.787 (3.222)	2.959 (2.468)	3.098 (3.338)	2.761 (2.456)
EDUCATION	1.288** (0.51)	0.218 (0.425)	1.374*** (0.501)	0.298 (0.431)	1.315** (0.513)	0.262 (0.428)	0.212 (0.507)	-0.375 (0.32)	0.09 (0.518)	-0.397 (0.315)
CREATIVE_INDEX	1.213 (1.798)	5.376*** (0.825)	-0.115 (1.407)	4.668*** (0.829)	0.946 (1.786)	5.028*** (0.807)	-1.779 (1.343)	1.543** (0.731)	-1.165 (1.609)	1.665** (0.729)
LOG(POP_DENS)	5.746*** (1.506)	5.116*** (1.453)			5.36*** (1.522)		0.186 (1.518)		-0.065 (1.612)	
FIRM_DENS		0.798*** (0.07)		0.764*** (0.081)		0.791*** (0.071)		0.696*** (0.089)		0.711*** (0.079)
LONDON_FE							71.425*** (10.895)	59.421*** (6.029)	76.856*** (11.885)	60.072*** (6.12)
HERF	-14.148 (77.344)	-68.817 (49.677)	3.441 (75.25)	-59.461 (48.494)	-16.115 (76.734)	-69.291 (48.427)	6.166 (56.066)	-44.114 (36.824)	-9.383 (58.686)	-50.031 (37.328)
F test	20.56***	137.49***	22.37***	133.89***	18.48***	138.38***	22.97***	91.98***	20.54***	101.01***
R2	0.77	0.89	0.80	0.89	0.78	0.89	0.84	0.93	0.83	0.93
N	168	168	168	168	168	168	168	168	168	168

Cluster robust errors in parentheses, * P<0.1 ** P<0.05 *** P<0.01.

Table A3. OLS results with based on stemmed words within tweets.

	OLS			
	(1) ENTRY	(2) ENTRY	(3) ENTRY	(4) ENTRY
INNOBUZZ_S	0.009** (0.004)			
INNOBUZZ_S_5		2.301*** (0.956)		
INNOBUZZ_S_3			0.027* (0.015)	
INNOBUZZ_S_ALL				0.002** (0.001)
LOG(GDP _{ppp})	105.351*** (25.293)	101.101*** (25.293)	110.063*** (25.461)	108.897*** (24.965)
GDP_GROWTH	1.245 (0.925)	1.237 (0.917)	1.232 (0.961)	1.134 (0.992)
LOG(R&D)	-7.904** (3.534)	-7.315** (3.563)	-8.472** (3.599)	-7.537** (3.598)
EDUCATION	1.237** (0.499)	1.071** (0.486)	1.177** (0.501)	1.072** (0.511)
CREATIVE_INDEX	0.675 (1.837)	0.385 (1.732)	0.703 (1.916)	0.293 (1.971)
LOG(POP_DENS)	5.337*** (1.483)	4.744*** (1.495)	5.461*** (1.463)	5.420*** (1.459)
HERF	-13.674 (74.163)	-15.586 (72.853)	-22.368 (74.922)	-18.087 (74.073)
F test	16.90***	18.29***	16.93***	16.66***
R ²	0.78	0.79	0.78	0.78
N	168	168	168	168

Cluster robust errors in parentheses, * $P < 0.1$ ** $P < 0.05$ *** $P < 0.01$.