Martin Obschonka, Neil Lee, Andrés Rodríguez-Pose, Johannes C. Eichstaedt and Tobias Ebert

Big data methods, social media, and the psychology of entrepreneurial regions: capturing cross-county personality traits and their impact on entrepreneurship in the US

Article (Accepted version) (Refereed)

Original citation:
Obschonka, Martin and Lee, Neil and Rodríguez-Pose, Andrés and Eichstaedt, Johannes C. and Ebert, Tobias (2018) Big data methods, social media, and the psychology of entrepreneurial regions: capturing cross-county personality traits and their impact on entrepreneurship in the US. Small Business Economics. ISSN 0921-898X (In Press)

Big data methods, social media, and the psychology of entrepreneurial regions: Capturing cross-county personality traits and their impact on entrepreneurship in the US

Martin Obschonka[1], Neil Lee[2], Andrés Rodríguez-Pose[2], Johannes C. Eichstaedt[3], Tobias Ebert[4]

Accepted for publication in *Small Business Economics*

[1]Australian Centre for Entrepreneurship Research, QUT Business School, Queensland University of Technology, Brisbane, Australia

[2]Department of Geography and Environment, London School of Economics and Political Science, U.K.

[3]Positive Psychology Center, University of Pennsylvania, Philadelphia, PA, USA

[4]Mannheim Centre for European Social Research (MZES), University of Mannheim, Germany

ABSTRACT

There is increasing interest in the potential of artificial intelligence and Big Data (e.g., generated via social media) to help understand economic outcomes. But can artificial intelligence models based on publicly available Big Data identify geographical differences in entrepreneurial personality or culture? We use a machine learning model based on 1.5 billion tweets by 5.25 million users to estimate the Big Five personality traits and an entrepreneurial personality profile for 1,772 U.S. counties. The Twitter-based personality estimates show substantial relationships to county-level entrepreneurship activity, accounting for 20% (entrepreneurial personality profile) and 32% (Big Five traits) of the variance in local entrepreneurship, even when controlling for other factors that affect entrepreneurship. Whereas more research is clearly needed, our findings have initial implications for research and practice concerned with entrepreneurial regions and eco-systems, and regional economic outcomes interacting with local culture. The results suggest, for example, that social media datasets and artificial intelligence methods have the potential to deliver comparable information on the personality and culture of regions than studies based on millions of questionnaire-based personality tests.

Keywords: Big Data, artificial intelligence, entrepreneurship, counties, U.S., social media, psychological traits, personality, Big Five, Twitter

INTRODUCTION

The economic impact of regional and local cultural characteristics has received steadily increasing attention over the past two decades (Duranton, Rodríguez-Pose, & Sandall, 2009; Greif & Tabellini, 2010; Guiso, Sapienza, & Zingales, 2004). This attention has, according to Huggins & Thompson (2017), spurred an interest in new measures and aspects of culture taken from psychological research, such as regional personality differences (Hofstede & McCrae, 2004; McCrae, 2001; Rentfrow, Gosling, Jokela, Stillwell, Kosinski, & Potter, 2013) and in how personality differences in the cultural make-up of territories affect regional economic trajectories (Lee, 2017; Obschonka et al., 2017).

One example where geographical cultural patterns can play a particularly important role for economic outcomes is the field of entrepreneurship (Fritsch & Wyrwich, 2014). The analysis of local or regional psychological differences has become an important subject of entrepreneurship research (Davidsson, 1995; Davidsson, & Wiklund, 1997; McClelland, 1961; Saxenian, 1994). Consistent with theorizing on the central role of culture for regional entrepreneurship (e.g., Hayton, George, & Zahra, 2002; Sternberg, 2009), this literature has shown that a cultural perspective helps explain spatial variations in entrepreneurship. The debate about what entrepreneurial culture actually is and how we measure it is, however, ongoing (Hayton & Cacciotti, 2013). Earlier studies focused on values, beliefs and need-for-achievement – often with inconsistent findings (Hayton & Cacciotti, 2013). More recent research has delved on the spatial effects of the Big Five personality traits, often by building entrepreneurial personality profiles from a constellation of these traits. The entrepreneurial personality profile most associated with local entrepreneurship includes high values in Extraversion, Conscientiousness and Openness to Experience, and lower values in Agreeableness and Neuroticism (Audretsch et al., 2017; Fritsch et al., 2018; Garretsen et al., 2018; Obschonka et al., 2013; 2015; 2016; Stuetzer et al., 2016).

The Big Five model is the most established and best-validated trait model in psychology (Digman, 1990; John & Srivastava, 1999) and has received considerable attention in entrepreneurship research in recent years (Brandstätter, 2011; Shane et al., 2010; Zhao, Seibert, & Lumpkin, 2010). As argued in Hofstede and McCrae (2004, p. 79), "culture-level traits can be legitimately operationalized as the mean of individual trait levels". Indeed, the Big Five traits have gradually become a more common indicator of the psychological facet of the local culture (McCrae, 2004). However, the actual dynamics between the geographical variation in personality traits and culture might be very complex with reciprocal dynamics. For example, regional and national cultural values are also likely to influence personality at various levels, bringing a collective element to individual personality traits (Hofstede & McCrae, 2004; McAdams & Pals, 2006). But if one assumes a certain overlap between personality and culture, conceptually and empirically, as some sort of psychological climate in the region as the umbrella construct, the study of this psychological climate, measured for example via regional personality features, appears to deliver interesting new insights in economic processes (Garretsen et al., 2018; Huggins & Thompson, 2017).

However, research on personality often faces a significant limitation, as data is only sporadically available at a local level. Psychological surveys are expensive and subject to particular selection biases, making it hard to investigate culture at a local level.

In this paper we address this problem by asking the following research question: Can artificial intelligence models, solely based on publically available Big Data (e.g., language patterns left on social media), reliably identify regional differences in entrepreneurial personality/culture and, in turn, in entrepreneurial activity? And can this be done when applying an established theoretical framework, namely this regional personality approach? We are the first to use and test public domain social media data ("digital footprints", Kosinski, Stillwell, & Graepel, 2013) as a source for the assessment of differences in local personality that may reflect regional differences in entrepreneurial activity. Hence, our aim is to explore the potential and validity of such new Big Data methods (e.g., based on artificial intelligence) for the field of regional and cultural economics and entrepreneurship research, given the increasing interest in the potential and predictive validity of Big Data from social media in these research fields (e.g., to stimulate theory development and to infer new research and practical implications). Specifically, we analyze, using a method approved by the University of Pennsylvania IRB (protocol #813866), county-level personality differences in the U.S., assessed by means of computerized language analyses of millions of short Twitter posts ("tweets"). The aim is to assess whether this new approach to evaluating personality from a geographical perspective delivers robust estimates that are markers of local and regional entrepreneurship. The explanatory power of local psychological characteristics derived from social media is compared with that of a number of economic factors traditionally deemed to be behind differences in entrepreneurship (see Eichstaedt et al., 2015 for a similar approach but on the association between regional Twitter-derived psychological patterns and regional health outcomes). We also compare the effect of the Twitter-based personality estimates (entrepreneurial personality profile) with effects of regional personality differences measured with self-reports (e.g., Obschonka et al., 2013, 2015). Our study is an attempt to embrace the "age of Big Data" in the context of entrepreneurship and economic research (Einav & Levin, 2014; Arribas-Bel, Kourtit & Nijkamp, 2015; Glaeser, Kominers, Luca, & Naik, 2018). By testing whether digital footprints, as a Big Data source, encode valid psychological information on regional personality differences, we assess the links between collective psychology and the local economy – in our case, entrepreneurial rates which, in turn, determine the economic dynamism of cities and regions in the U.S. (Glaeser, Kerr, & Kerr, 2015).

The study makes three central contributions. First, we evaluate the potential and usefulness of public-domain social media as a Big Data source in entrepreneurship research. We also contribute to the emerging literature linking Big Data and social media to entrepreneurship (e.g., Tata, Martínez, García, Oesch, & Brusoni, 2017). Second, we push forward the entrepreneurship research studying the role of personality (Brandstätter, 2011) and regional psychological characteristics (Audretsch et al., 2017; Davidsson, & Wiklund, 1997; Huggins & Thompson, 2017) by using Big Data methods (Zomaya & Sakr, 2017). Third, we extend the research on Big Data's usefulness in the study of personality and behavioural and psychosocial outcomes (e.g., Eichstaedt et al., 2015; Kosinski & Behrend, 2017; Kosinski, Wang, Lakkaraju, & Leskovec, 2016; Wilson, Gosling, & Graham, 2012; Youyou, Kosinski, & Stillwell, 2015) by demonstrating that social media encodes relevant psychological information that can predict *economic* activity, in addition to other outcomes such as health (Eichstaedt et al., 2015), happiness (Curini, Iacus, & Canova, 2015), or political orientation (Sylwester & Purver, 2015). We have to stress though that the analysis, like similar studies in cognate fields (e.g., Eichstaedt, 2015), delivers correlations on the link between digital footprints and regional outcomes that cannot prove causality. The results of the analysis should thus be interpreted as a test of whether social media language, translated into local personality characteristics, can be a meaningful and robust statistical marker of economic activity, in our case entrepreneurship rates (see also Eichstaedt el., 2015). Nevertheless, as our study deals with personality traits (using a new measurement method at the local level) and there is growing evidence that geographical changes in personality shape economic (and other) outcomes of regions (Garretsen et al., 2018; Lee, 2017; McClelland, 1961; Obschonka et al., 2016; 2018; Stuetzer et al., 2018), it may also guide future research examining the actual mechanisms and causal links between digital footprints and economic outcomes. This new research could, of course, also examine how economic factors, in turn, affect regional personality differences (Obschonka et al., 2017).

The paper is structured as follows. In section two, we outline the existing theory on the relationship between local personality traits and entrepreneurship and develop a series of testable hypotheses. Section three outlines the methods we use to identify personality traits via Twitter and our methodology to use this data to test our hypotheses. Section four presents the results of our models, before section five concludes.

HYPOTHESES

Informed by theories that highlight the role of personality factors as drivers of entrepreneurial behavior (Knight, 1921; McClelland, 1961; Schumpeter, 1934, see also Hisrich, Langan-Fox, & Grant, 2007), prior research examining personality traits in entrepreneurship has relied fundamentally on standard techniques to measure individuals' personality – most commonly self-reports collected via typical personality questionnaires (Brandstätter, 2011; Obschonka, 2017).

Such questionnaire-based research at the individual level typically finds that the Big Five traits of Extraversion, Conscientiousness, and Openness to Experience are positively correlated with entrepreneurial behavior, while Neuroticism has a negative connection with it (Brandstätter, 2011; Shane et al., 2010; Zhao, Seibert, & Lumpkin, 2010). The role of the remaining trait, Agreeableness, is less clear (Zhao, Seibert, & Lumpkin, 2010). Some analyses posit that entrepreneurial behavior is negatively linked with Agreeableness (Brandstätter, 2011), as entrepreneurship often requires non-conformism and (mild) rule-breaking (Schumpeter, 1934; Zhang & Arvey, 2009). Research has also stressed that an intra-individual constellation of the Big Five traits (entrepreneurial personality profile) where low Agreeableness and Neuroticism interact with high Extraversion, Conscientiousness, and Openness to Experience is positively correlated with entrepreneurial behavior (Obschonka & Stuetzer, 2017; Schmitt-Rodermund, 2004, 2007). Individual-level research also proposed that not only entrepreneurial behavior, but also underlying entrepreneurial human and social capital, identity, passion, and other specific traits, such as self-efficacy, risk-taking, and internal locus of control are connected with such a personality profile (characterized by higher values in Extraversion, Conscientiousness, and Openness to Experience, and lower values in Agreeableness and Neuroticism). Evaluating data from a longitudinal cohort study, Schmitt-Rodermund (2007) found that even in adolescence such profile is capable of predicting entrepreneurial behavior over the subsequent life course. Hence, this type of research has highlighted that personality does indeed affect entrepreneurship (and not just the other way around). This is in line with studies on the effect of personality traits on various life outcomes (Roberts et al., 2007), including work-related outcomes (Fruyt & Mervielde, 1999), such as occupational choices (Holland, 1997).

However, the observed individual-level link between Big Traits and entrepreneurial behavior is often not very large. This is an expected pattern for three reasons: a) given the phenomenon of equifinality in career development, i.e., where different initial conditions can lead to the same career outcome (Colarelli, Dean, & Konstans, 1987); b) given the changing nature of today's careers, where the fit between one's personality and a stable career trajectory, i.e., working in the same job that fits one's personality throughout the whole career (Holland, 1997), is less relevant than in the past; and c) given the fact that adaptive capacities and a general entrepreneurial and flexible approach to careers has become more important in general (Fouad, 2007; Sullivan, 1999; Savickas & Porfeli, 2012).

While the individual-level perspective on the link between personality and entrepreneurship has attracted considerable attention in recent years (see, for example, the "meta-analytic" summary of various meta-analyses in the field by Brandstätter, 2011), the analyses covering the geographical dimension of this phenomenon are much more limited. This is notwithstanding the fact that theories on local entrepreneurial eco-systems and regional differences in entrepreneurial activity increasingly highlight the role of local behavioral and psychological foundations as a crucial determinant of economic outcomes (Fitjar and Rodríguez-Pose, 2011; Huggins & Thompson, 2017; Saxenian, 1994; Sternberg, 2009). The mismatch between a booming individual-level research on personality and entrepreneurship and the disregard for its local and regional dimension is possibly a consequence of the persistence phenomenon in regional entrepreneurship research. Substantial and persistent regional differences in entrepreneurial activity over longer periods of time – the result of path dependencies derived from relatively stable cultural structures (Fritsch & Wyrwich, 2014) – can explain the relative lack of interest of the role of psychology for entrepreneurship at a local and regional level. Past research does indeed underline that regional personality differences show considerable stability (e.g., Elleman, Condon, Russin, & Revelle, 2018) and can be linked to historical processes that took place decades or even centuries ago (Duranton, Rodríguez-Pose, & Sandall, 2009; Nunn & Wantchekon, 2011; Obschonka et al., 2017; Talhelm et al., 2014).

What has the existing research linking regional personality differences to entrepreneurship found so far? Existing research has normally measured regional personality by means of self-report questionnaires. These studies have mainly uncovered that the entrepreneurial personality profile (low Agreeableness and Neuroticism, and high Extraversion, Conscientiousness, and Openness to Experience) of a region in a range of countries, such as the U.S., the U.K., or Germany, relates to regional entrepreneurial activity (e.g., Audretsch et al., 2017; Obschonka et al., 2013; see also Carbonara et al., 2018). This is in line with the results of individual-level research. Moreover, as is also the case of research on individuals, there is some evidence that personality (assessed at the Big Five level) is the cause and entrepreneurship the effect. For example, using a natural experiment (the global economic recession of 2007-2008), Obschonka et al. (2016) demonstrated that geographical differences in the entrepreneurial personality profile predicted differences in entrepreneurial activity trajectories during this recession. Places scoring higher in the entrepreneurial profile before the crisis showed a lower or no economic decline (in terms of decreasing start-up rates) during the recession. Hence, specific psychological patterns were behind the economic resilience of territories in the early stages of the crisis. Employing an instrumental variable technique, other research has found that regional personality differences (in this profile) that are tied to an exogenous instrument (e.g., historical coal mining) predict spatial economic growth (Stuetzer, et al., 2018; see also Garretsen et al., 2018), which is stimulated by entrepreneurship (Beugelsdijk, 2010; Glaeser, Kerr, & Kerr, 2015). Other research examining the origins of current differences in entrepreneurial culture assessed by means of geographical differences in the entrepreneurial personality profile indicate that a historical concentration of large-scale industries dominating entire regions over many decades has shaped the local psychology, which then affects the region's entrepreneurial activity today (Stuetzer et al., 2016). The mechanisms behind this impact are agentic and social processes stimulating entrepreneurial thinking and acting as well as a certain entrepreneurial culture in local populations (Huggins & Thompson, 2017).

In short, recent studies have suggested that geographical differences in personality may help explain variations in entrepreneurship and new firm creation. These studies typically find stronger relationships between personality and entrepreneurship at the territorial level than at the individual-level. Yet such region or local-level studies are still rare, as the lack of availability of large personality datasets has limited the ability to carry out such analyses. Datasets need to be large enough (Big Data approach) to study relatively fine-grained spatial levels, such as counties or cities as the units of entrepreneurial eco-systems that may differ in entrepreneurial culture and activity (Audretsch & Keilbach, 2007; Lee, Florida, & Acs, 2004; Spigel, 2017). As already indicated, most of these studies linking personality and entrepreneurship at the territorial level rely on questionnaire-based self-reports. However, the methodological limitations of such self-reports are well known (Baumeister, Vohs, Funder, 2007; Furr, 2009). At the same time, the value of behaviour residue and language patterns left on social media as additional source in objective personality assessment is gaining considerable attention (Kosinski & Behrend, 2017; Wilson, Gosling, & Graham, 2012). Interestingly, recent research has substantiated the validity of social media-based personality assessment, as digital footprints from social media typically reflect the actual and not self-idealised or “faked” personality structure of individuals (Back et al., 2010; Kosinski, Stillwell, & Graepel, 2013; Schwartz et al., 2013a, 2013b).

In this study, we utilize social media as a Big Data source to extract information about the psychology of large numbers of individuals and mapping the dominant psychological patterns that may be linked to entrepreneurship across different areas of the U.S.. Do digital footprints from social media convey information to estimate (in terms of a statistical marker of) the economic vitality of cities and regions? Can entrepreneurial activity be derived from the language used in the Twitter tweets of individuals living in specific locations? Could we then, for example, use publically available social media data to estimate central indicators of the local entrepreneurial culture of entrepreneurial eco-systems (Stam, 2017)? In order to answer these questions, we develop a number of hypotheses drawn from earlier entrepreneurship research on the Big Five, which relied on questionnaire-based self-reports to measure personality. Although we “only” study regional personality indirectly, by solely focusing on psychological patterns in social media language, we expect – informed by similar Big Data research on regional psychological patterns measured with social media and predicable effects on regional outcomes (Curini, Iacus, & Canova, 2015; Sylwester & Purver, 2015) – to find a similar pattern as that from entrepreneurship research based on self-report questionnaires. So, in other words, we assume that artificial intelligence methods are indeed effective in translating local language patterns, used in social media, into reliable markers of regional entrepreneurial activity (e.g., the entrepreneurial personality profile) and thus into markers of entrepreneurial culture (if one accepts the notion that regional personality differences reflect important aspects of regional *cultural* differences, Hofstede & McCrae, 2004; Huggins & Thompson, 2017; McCrae, 2001). We thus expect:

H1: Local language patterns translated via artificial intelligence methods into local differences in an entrepreneurial personality profiles can uncover county entrepreneurship rates (positive correlation between the profile and entrepreneurship rates).

H2: Local language patterns translated via artificial intelligence methods into local differences in Extraversion can uncover county entrepreneurship rates (positive correlation between Extraversion and entrepreneurship rates).

H3: Local language patterns translated via artificial intelligence methods into local differences in Conscientiousness can uncover county entrepreneurship rates (positive correlation between Conscientiousness and entrepreneurship rates).

H4: Local language patterns translated via artificial intelligence methods into local differences in Openness can uncover county entrepreneurship rates (positive correlation between Openness and entrepreneurship rates).

H5: Local language patterns translated via artificial intelligence methods into local differences in Agreeableness can uncover county entrepreneurship rates (negative correlation between Agreeableness and entrepreneurship rates).

H6: Local language patterns translated via artificial intelligence methods into local differences in Neuroticism can uncover county entrepreneurship rates (negative correlation between Neuroticism and entrepreneurship rates).

METHOD

First, to depict the spatial distributional patterns of our outcome variable (i.e., start-up rates) and major independent variable (i.e., entrepreneurial personality profile), we apply a so-called *Hotspot-Analysis* (Kondo, 2016). That is, we calculate the *Getis-Ord G**statistic (Getis & Ord, 1992; Ord & Getis, 1995) as a measure of local clustering for each region. To calculate the Getis-Ord G*, first the connection between the underlying spatial entities needs to be quantified. Thereby, the most widely used approach is to apply a spatial weight matrix indicating whether two regions share a border or not (Jokela et al., 2015; Rentfrow et al, 2015).

Given that our sample does not comprise all counties of the U.S., some counties would end up with no (or a reduced number of) neighbors. Therefore, we instead apply a definition of neighboring in which each cell of the matrix indicates whether the centroid of two regions are more than 100km apart from each other.[1] Finally, we *row standardize* the resulting matrix. That is, we divide each binary weight by the number of neighbors for that county, hence resulting in equal proportional weights for all counties. Based on this spatial definition, the Getis-Ord G* measure can be defined as:

$$G_i^* = \frac{\sum_{j=1}^{n} w_{i,j} x_j - \bar{X} \sum_{j=1}^{n} w_{i,j}}{S \sqrt{\frac{\left[n \sum_{j=1}^{n} w_{i,j}^2 - \left(\sum_{j=1}^{n} w_{i,j} \right)^2 \right]}{n-1}}} \quad (1)$$

[1] The threshold of 100km is informed by previous research showing that commuting (as a proxy for an individual's daily available interaction radius) becomes cumbersome if it exceeds 80-100km (e.g., Helminen & Ristimäki, 2007). We also tested alternative distance thresholds, leading to very similar distributional patterns.

where $w_{i,j}$ is the weight between the regions i and j, x_i is the actual value in the region and S_O is the sum of all weights. In other words, the Getis-Ord G* compares the values of a county and its neighbors against the sum of all regions. The more a local sum deviates positively (or negatively) from its expected value, the more clustering of high (or low) values happens in the referring area. The output is a z-score that directly indicates the statistical significance of the deviation. To interpret the findings, we map the regional z-scores and thereby reveal areas in which positive or negative clustering occurs.

Next, to test the hypotheses outlined above, we estimate a series of regression models which are variations of the following:

$$Entrepreneurship_i = \alpha + \beta_1\, Psychology_i + \beta_2\, Controls_i + \varphi + \varepsilon \quad (2)$$

For county 'i', where *Entrepreneurship* is a variable for the ratio of new firm starts; 'Psychology' depicts, in turn, one of the Big Five personality traits (Openness, Neuroticism, Extraversion, Agreeableness, Conscientiousness) or the entrepreneurial personality profile; 'Controls' represents a vector of other potential local determinants of entrepreneurship; 'φ' is one of 52 state fixed effects, and '$\boldsymbol{\varepsilon}$' is the error term. The unit of analysis is the U.S. county (or equivalent, e.g., parishes in Louisiana). The counties included in the analysis have an average population of around 100,000 people (but a range that spans order of magnitude, from 88 people to over 10 million). Counties tend to be the smallest geographical unit for which nation-wide economic indicators are available. Summary statistics, sources and variable definitions are included in Table 1.

Insert Table 1 around here

Outcome Variable

The dependent variable is *entrepreneurship rate*, measured as the log number of new firm births per 1,000 people. This is probably the most common measure of entrepreneurship and covers all businesses which have at least one employee (therefore excluding self-employment). Data come from the Statistics of U.S. Businesses, with each establishment registered when it hires its first employee. This definition is commonly used in entrepreneurship research, but does not give us information on the quality of new firm starts (see Nightingale and Coad, 2014). To match the time period over which the tweets were collected, we average data for the period 2009-2015.

Personality Traits Estimated via Twitter

Our data for personality traits comes from the World Well-Being Project at the University of Pennsylvania (see Park et al., 2015, and map.wwbp.org). The personality estimates are derived in a sequence of steps: a) a machine learning model is used to estimate personality from language use across a sample of 71,556 Facebook users for which language and survey-based self-reported Big Five scores were available; b) geo-tagging a 10% Twitter language sample using data from 5.25 million users; c) adjusting the Twitter language frequencies appropriately for an application of the Facebook prediction model; and d) applying the Facebook person-level prediction model to counties on Twitter.[2] Figure 1 summarizes this procedure.

[2] The University of Pennsylvania IRB approved this study (protocol #813866).

Building a personality prediction model on Facebook data. A sample of N = 71,556 Facebook users (from the MyPersonality dataset, Kosinski & Stillwell, 2011) took a standard survey to estimate personality traits as defined by the NEO-PI-R five factor model (Costa & McCrae, 1992), using 20-100 items from the International Personality Item Pool (IPIP; Goldberg et al., 2006). These users had consented to share the language of their Facebook posts ("wall posts") for research purposes. Using the methods described in Park et al. (2015) and an open-source code base (Differential Language Analysis ToolKit, see dlatk.wwbp.org, Schwartz et al, 2017), we extracted the relative frequency of words and phrases for every user, in addition to their use of 2,000 topics extracted in previous work (Schwartz et al., 2013a, 2013b). In addition to the relative frequencies, we also derived a binary encoding, i.e., a 0/1 variable for every word and phrase, encoding if a user had mentioned it at least once. This yielded a set of 51,060 language variables for every user, encoding their language use. We fed these variables as predictors into five (machine learning) ridge regression models, one model for each survey-reported personality dimension as the dependent variable. Fundamentally, these are multiple linear regression models that include the language variables as the independent and the personality dimension as the dependent variable.

For example, for person i,

$$Agreebleness_i = \beta_1 * feature_{i1} + \cdots + \beta_k * feature_{ik}$$

Where k is the number of features (k = 51,060), and $feature_{ik}$ gives *user i*'s relative use of $feature_k$. In typical OLS regression, the sum of the square of the residuals make up the loss function to be minimized in the fitting of the coefficients; ridge regression adds the sum of the square magnitude of the coefficients as a "penalty term" to this loss function, encouraging the model to fit conservative, smaller coefficients. Ridge regression models are suited for cases in which there are many more predictors than there are observations and/or the predictors are highly collinear, as in this case of using language variables. The relative balance of OLS loss and the penalty term is determined through a single coefficient (a hyperparameter). This hyperparameter is fit automatically in a search process in a cross-validation framework, yielding a model that strikes the best balance between fitting the data and generalizing well to new data. Accuracies of these models are always reported on new ("testing") data, not the training data used to fit the model. We saved the coefficients ($\beta's$) of this prediction model for later application to Twitter.

Evaluating the personality prediction model. In previous work (Kern et al., 2014; Park et al., 2015), the quality of these Facebook prediction models was evaluated along a number of psychometric dimensions. The language-based estimates showed adequate convergence with the scores obtained through self-report surveys (average r = .39), and a pattern of correlations similar to self-reported personality with a large list of external variables, including self-reported sick days and physician visits, political orientation, satisfaction with life and number of friends. Six-month test-retest reliabilities of the language-based assessments of personality averaged r = .70. Finally, examining the language features (words, phrases, topics) most strongly associated with the Big 5 personality dimension (and thus the most important features in the prediction model) reveal a pattern of associations that is both face valid and coherent with the psychological literature. For example, among the most correlation features for extraversion are *party, can't wait,* and *love you;* for agreeableness *blessed, excited* and *wonderful;* for conscientiousness *ready for, to work* and *workout*; for neuroticism *sick of, f**** and *depression*; and for openness to experience *writing, art* and *universe*. This previous work thus suggests that language-based assessments through these prediction models demonstrate convergent, external, test-retest and face validity.

Twitter data set: random sampling. To assemble a county-level data set of Twitter language, we started with a 10% random sample of all U.S. tweets collected between August 2009 and January 2015. This feed of random Tweets is provided by Twitter (for a fee) and commonly referred to as the "gardenhose" feed. Twitter also makes a 1% random feed available for free ("spritzer"), in addition to providing 100% of the Tweets to paying institutional clients (the "firehose"). For the 10% feed, the exact nature of the random sampling is unknown. While the sampling of the 1% feed has been critiqued as not truly random (Morstatter et al., 2013), we know of no such concerns about the 10% feed, and papers in computer science frequently use the 10% feed as the reference feed to compare the 1% feed against (e.g., Valkanas et al., 2014). In previous work, Eichstaedt et al. (2015) have used a smaller sample of the 10% feed (only spanning 9 months in 2009 to 2010 and including about 150 million tweets) to predict heart disease mortality. They showed that this sample contained enough information about population heart disease mortality at the county level to outpredict a model based on the ten leading, officially-collected risk factors, in an out-of-sample, cross-validation framework. The data set used in this work is about an order of magnitude larger (1.5 billion county-tagged tweets). While we are not able to quantify or investigate the random sampling of the 10% feed, the previous work discussed above suggests that the 10% can serve as a good-enough basis for the estimation of county-level psychological phenomena.

Twitter data set: geo-tagging and feature extraction. Based on the location field used by the users on their Twitter profile page, we determined their most likely location using a set of rules to infer the corresponding U.S. county (see Schwartz et al., 2013a, for full methods of geo-tagging the users and their tweets). We then only retained Twitter users in the data set who had shared at least 30 tweets and extracted the same language features that had previously been extracted in the Facebook sample for every user (words and phrases encoded both as binaries and relative frequencies, in addition to the relative use of the same 2,000 LDA topics). Having extracted these language frequencies for every user, we aggregated (averaged) these language features to the county-level in such a way that every included Twitter user contributed equally to the overall county-level profile of language use, limiting the potential of any one Twitter account to distort the language results (e.g., through bots or other Twitter accounts not representing natural persons). We retained county-level language samples for counties that had at least 100 Twitter users in them (with at least 30 tweets per user, as stated above). This process yielded a final county-level Twitter language sample that included 1.5 billion tweets from 5.25 million Twitter users for 1,772 counties that represented just under 95% of the total U.S. mainland population (see Giorgi et al., 2018 for full methods on creating the Twitter data set). Given the restrictions on the number of users per county, many small, often rural, counties were dropped from the sample.

Application of Facebook prediction model to county-level Twitter data. In the final step, we first estimate the *region-level Big Five traits* for the counties and then the entrepreneurial personality profile for these counties. Namely, using the DLATK codebase (see dlatk.wwbp.org), we apply the personality prediction model trained on the user-level Facebook language features to the same language features derived at the U.S. county level (Twitter data). That is, as in all multiple linear regression models, we apply the coefficients learned by the Facebook prediction model for the relative frequencies of words, phrases and topics to the relative frequencies of the same words, phrases and topics used by the Twitter users, to yield personality estimates for these Twitter users. However, as language use on Twitter may deviate from the language use on Facebook, we for correct outliers in the Twitter language frequencies through a process called Target-Side Domain Adaptation (described below) before applying the prediction model. After this correction and the application of the model, we obtained Big Five personality estimates for the 1,772 counties (see also map.wwbp.org by the World Well-Being Project). [3]

Target-side domain adaptation. As language use on Twitter and Facebook may differ in that some words are used more frequently on one platform but not the other could distort the personality predictions (e.g., "RT" [for retweet] is mentioned very frequently on Twitter, but not on Facebook, where it may be used in a different sense such as for "Russia Today"). Accordingly, the estimate of the prediction model would thus be unduly influenced by such differences in language use. TSDA addresses such outliers in the frequency of single words on Twitter compared to the baseline of frequency observed on Facebook (such as "RT") by replacing them with the global average observed on Facebook, before the prediction model is applied to the Twitter data. In this way, the resulting predictions more conservative; previous work has shown this to result in increases in their year to year stability as well as the external validity of the predictions (Rieman et al., 2017).

Entrepreneurial Reference Profile. The obtained county-level Big Five scores are then used to determine the extent to which a county's personality profile matches an entrepreneurial reference profile. Thereby, following previous research (e.g., Audretsch et al., 2017; Fritsch et al., 2018; Garretsen et al., 2018; Obschonka et al., 2013, 2015, 2016; Stuetzer et al., 2016), an

[3] We note that while we collect Twitter data over a period of time, we do so to maximise the sample rather than collect a panel indicator.

entrepreneurial personality profile is defined as higher scores in Extraversion, Conscientiousness, and Openness, and lower scores in Agreeableness and Neuroticism (e.g., measured via the deviation of the empirical profile to a fixed entrepreneurial reference profile, defined as the highest score in Extraversion, Conscientiousness and Openness, and the lowest scores in Agreeableness and Neuroticism). We use the highest (lowest) observed county-level score for each trait as the extreme points of our scales. To evaluate the goodness of fit between a county's personality profile and the entrepreneurial reference profile, we calculate next the absolute deviation between the actual county score and the reference profile for each single trait. In a following step, we add up the five scores for each county. Finally, we reverse the algebraic sign of the sum so that in our final index a value closer to zero indicates a better fit with the entrepreneurial reference profile (= a *more entrepreneurial* personality profile, Obschonka et al. 2013).

Figure 2 shows word cloud visuals of the words and phrases that mostly distinguished the Big Five personality traits in the World Well-Being Project (social media language that may indicate an entrepreneurial personality structure in the region according to the model we use).

Standard Control Variables

To estimate the unique explanatory power of the personality traits derived from the Twitter tweets, we control for a host of standard predictors of economic vitality/entrepreneurship. Where possible, we use data for the full period 2009-2015, matching the data collection period. The first control variable is the natural log of *county population density,* defined as the total population per square kilometer (US Census Bureau, 2018). We expect, in line with the relevant literature, larger counties to be more entrepreneurial as a consequence the positive externalities derived from agglomeration (see Bosma and Sternberg, 2014). Following Rodríguez-Pose and Hardy (2015), we also control for two variables designed to assess local economic conditions – *unemployment* rate and the natural log of *median household income*. Two additional variables take into account the skill composition of the population, both the share of the working age population *without a high-school diploma* and those qualified to *degree level or above* (Nathan and Lee, 2013). Finally, we also include a variable to account for the *industrial diversity* of the county (Rodríguez-Pose and Hardy, 2015). Following past literature, we use the inverse Hirschman-Herfindahl Index (HHI) which is the sum of squared proportions of each industrial sector. The control variables are taken from the US Census Bureau's American Factfinder portal, and sources are given in table 1. Table 2 presents the bivariate correlations between all variables used in this study.

Insert Table 2 around here

RESULTS

Figure 3 presents the geographical variation of the entrepreneurial personality profile, estimated with social media data (as summarized in Figure 1). The regional distribution of the entrepreneurial profile is very similar to regional distributions across U.S. regions when measured with self-report questionnaires (e.g., Obschonka et al., 2013, 2015). Entrepreneurial personality clusters along both coasts, between Southern Florida and Connecticut on the Atlantic coast, and between Southern California and the State of Washington. More precisely, entrepreneurial personality hotspots are uncovered by our social media-based method not just in the Silicon Valley and the San Francisco

Bay area, but also in Los Angeles, Orange County and San Diego in the West, and in and around New York, Philadelphia, Washington, Richmond, Charlotte, Atlanta and Miami in the East (Figure 3). Inland, Denver and Phoenix also show strong clustering of comparatively high entrepreneurial personality levels. Spatial clustering of low values is particularly evident in the Rust Belt and parts of the Mid-West. This is consistent with studies linking a historical concentration of large-scale industries to a lower entrepreneurial culture (e.g., measured via the entrepreneurial personality profile from self-reports in questionnaire-based studies, Stuetzer et al., 2016; see also Obschonka et al., 2013).

The psychological map of entrepreneurial personality based on tweets in the U.S. (Figure 3) closely matches that of actual entrepreneurial activity (Figure 4), based on start-up rates across counties. The Silicon Valley, Southern California, the main cities in the Mid-Atlantic States, Denver and, to a lesser extent Florida are at the top of entrepreneurial performance in the U.S. Nevertheless, some real entrepreneurial hotspots, such as Minneapolis-Saint Paul, Portland (Or.) or St Louis, are not reflected in Twitter language, while Charlotte or Phoenix have a lower level of entrepreneurship than the analysis of the language used on Twitter would reflect.

Insert Figures 3 and 4 around here

The correlation between the Twitter-based entrepreneurial personality profile of a county and its real level of entrepreneurship is plotted in Figure A1 in Appendix 1 (which also illustrates the correlations between the single regional Twitter-based Big Five traits and entrepreneurial activity). The correlation between the Twitter-based entrepreneurial personality profile and entrepreneurial activity is r = .45 in our data (see Table 2), which is very similar to the region-level correlations between the entrepreneurial personality profile and entrepreneurial activity found in studies analyzing personality data from self-reports (questionnaire-based studies). For example, in an analysis comparing 51 U.S. states a prior study found the entrepreneurial personality profile, measured via self-reports, to correlate r = .39 with the Kauffman index of entrepreneurial activity, r =.53 with the establishment entry rate, and r =.31 with the self-employment rate at the U.S.-state level (Obschonka et a., 2013). Another study comparing 366 MSA's in the U.S. and 375 Local Authority Districts in the U.K. found correlations between the entrepreneurial personality profile, measured with self-reports, and the local start-up rate of r = .36 (within the U.S.) and r = .58 (within the U.K.) (Obschonka et al., 2015).

Figure 5 shows the regions in the US where Twitter-based entrepreneurial personality and actual entrepreneurship mainly cluster (regions where the spatial clustering is statistically significant; z-score > 1.96 or < -1.96). These maps underscore our general results that speak for a substantial overlap between the psychological and economic maps.

Insert Figure 5 around here

To test our hypotheses, we estimate the effects of the local personality estimates on entrepreneurial activity by U.S. county. Table 3 presents the results of the OLS regressions. Column 1 considers simply the entrepreneurial personality profile, without any controls and no state fixed effects. We see a positive, significant connection that explains 20% of the variance. This supports H1. Column

2 considers state fixed effects. The entrepreneurial personality profile (together with the state fixed effects) now accounts for 49% of the variance.

Column 3 tests the single Big Five traits (instead of the profile), without any controls and no state fixed effects. Four of the personality traits are statistically and significantly associated with entrepreneurship at county level: Openness, Agreeableness, and Conscientiousness positively, and Neuroticism negatively. While the results for Openness, Neuroticism and Conscientiousness are in line with the stated hypotheses (H3, H4, H6), that is not the case for Agreeableness (H5). This model accounts for 32% of the variance. When considering state fixed effects (column 4), the model explains 51% of the variance. Openness and Conscientiousness still show a positive connection with entrepreneurship rates, but the coefficients for Agreeableness and Neuroticism become non-significant. Extraversion, in turn, becomes positive and significant in this model. This model thus supports H2, H3 and H4.

Columns 5 and 6 introduce the county-level control variables (without and with state fixed effects). We then test the entrepreneurial personality profile against these control variables (column 7), with state fixed effects. The profile is still positively associated with actual entrepreneurship, which again supports H1. Even when controlling for a range of economic standard factors that according to the literature should affect entrepreneurship, the entrepreneurial personality profile, as mapped in Figure 3, shows a significant and positive correlation with local entrepreneurship. A one standard deviation increase in the Twitter-based entrepreneurial personality variable is associated with a .18 standard deviation increase in new firm births.

Column 8 to 12 test the robustness of each of the Big Five trait – introduced consecutively – coefficients, when control variables are introduced in the analysis. The aim is to assess the connection of each individual Big Five trait, independently from each other, with entrepreneurship. Openness, Extraversion and Conscientiousness show positive coefficients, while that for Neuroticism is negative and significant (supporting H2, H3, H4, and H6). Only Agreeableness is insignificant, which concurs with the weaker link between Agreeableness and entrepreneurship identified in survey- and questionnaire-based research (relying on self-reported traits) at the individual level (Zhao, Seibert, & Lumpkin, 2010).

Finally, when all single Big Five traits are included together with controls and state fixed effects in one model (column 13), Openness, Extraversion, Neuroticism and Conscientiousness are all positively and statistically significantly associated with entrepreneurship; Agreeableness negatively so. Hence, in this regression, H2, H3, H4 and H5 are supported, but not H6 (which assumed a negative sign for Neuroticism).

Taken these regression results together, we see a clear and robust, positive relationship between the entrepreneurial personality profile of a county and its level of entrepreneurship, as expected in H1, even when controlling for the standard economic factors behind entrepreneurship. The picture for the single Big Five traits is less consistent, which concurs with prior research comparing the effects of the regional variation in the entrepreneurial personality profile with the single Big Five traits (e.g., Garretsen et al., 2018; Obschonka et al., 2013, 2015, 2016). The most consistent picture in the present analysis is delivered by Conscientiousness and Openness. In line with H2, H3 and H4, Conscientiousness, Openness, and, to a slightly lower degree, Extraversion are connected with local entrepreneurship, even when a host of social and economic variables are controlled for and state fixed effects are included. However, the coefficients for Agreeableness and Neuroticism are less consistent across the various regression models.

The controls mostly follow expectations. There are greater levels of entrepreneurship in areas of the U.S. with a better endowment of human capital – proxied by the percentage of the population with a university degree – and entrepreneurship is lower in countries with high levels of unemployment. Only the coefficient for the percentage of the population without a high school diploma challenges previous views, although in fairness part of the literature on entrepreneurship has highlighted the role of the low-skilled as entrepreneurs, especially among the immigrant community (Kloosterman, 2010; Kloosterman & Rath, 2001; Lofstrom, 2013).

DISCUSSION

What makes some places more entrepreneurial than others? Until now research suggested that entrepreneurship was the result of a combination of individual and place-level characteristics. Individual factors, such as education, experience, age, and previous employment status determined the likelihood of a person becoming an entrepreneur (Audretsch, 2003). Psychological characteristics also ranked highly. The need for personal development, zest for learning, and personal traits, such as risk-taking, independence, charisma and leadership, have featured highly in this line of research (Carter, Gartner, Shaver & Gatewood, 2003). The socio-economic environment is considered to shape the level of entrepreneurship in specific places. Local macroeconomic conditions, industry structure, the financial environment as well as local institutions, the education system and local culture facilitate or deter the propensity to become entrepreneurs in specific territories (Cuervo, 2005).

The role played by collective psychological patterns – something that is "in the air" in a region, imprinted in the behavior of people and most likely also in the local language style – attracted less attention in existing research determining whether regions are more or less entrepreneurial. While it has been highlighted that regional personality differences play an important role for local entrepreneurship and the persistence of regional differences in entrepreneurship rates over time (e.g., Stuetzer et al., 2016, 2018), the appetite for new research analyzing regional personality differences (Huggins & Thompson, 2017) was thwarted by problems in obtaining information about the prevailing psychological patterns across a wide range of cities and regions (e.g., counties). This meant that most research on the topic to date relied on self-reports only, implying important limitations when it comes to the measurement of the actual personality of individuals (and regions) (Baumeister, Vohs, Funder, 2007; Furr, 2009).

Our research has addressed this gap by using Big Data methods and digital footprints from social media – 1.5 billion tweets by 5.25 million users – in order to estimate regional personality differences that, in turn, may reflect aspects of the local economic culture. The psychological patterns, measured by the Big Five personality traits derived from the tweets, were then connected to differences in entrepreneurship rates across U.S. counties. While our study cannot address causality and direction of effects, the results indicate that counties that rank higher in the entrepreneurial personality profile, and in the personality traits Conscientiousness, Openness and Extraversion, as reflected in the local language patterns used in social media, are also more entrepreneurial in terms of behavior. The most consistent and robust results were delivered by the entrepreneurial personality profile and by Conscientiousness and Openness. In industrial psychology, Conscientiousness is typically the most important Big Five trait in analyses predicting job motivation and performance of individuals (Barrick & Mount, 1991). This trait reflects the typical virtues that are valued in today's work such as self-control, self-management and a strong motivation to achieve outcomes and be productive. These virtues may particularly matter for entrepreneurship, which often relies on the motivation and skills of the entrepreneurs (Brandstätter, 2011; McClelland, 1961). Survey-based regional research found that regions with a history of economic hardship (and lower entrepreneurship rates) score lower in this trait (Obschonka et al., 2017; Stuetzer et al., 2016).

The finding that regional Openness expressed in social media language stimulates entrepreneurship is consistent with research and approaches giving creativity and a proactive and open approach to change and innovation a unique role (Lee, Florida, & Acs, 2004). A regional or local environment open to new ideas and change seems conducive to entrepreneurship because, by its very definition, entrepreneurship is about the discovery, development and application of new solutions. Moreover, regional Openness can be also linked to a "taste for variety"-tendency in the local population (Åstebro & Thompson, 2011; Lazear, 2005).

While our data did not deliver clear conclusions regarding Agreeableness and Neuroticism, it is noteworthy to stress that the effect size of local psychological traits solely derived from publicly available social media is substantial: County-level personality estimates accounted for one fifth (20% - the entrepreneurial personality profile) and one third (32% - the Big Five traits when studied as single traits in one model) of the variance in county-level entrepreneurial activity, which are substantial effects of practical relevance (Ferguson, 2009). Even when including socio-economic controls that traditionally have been considered the main determinants of regional entrepreneurship and accounting for unobserved differences across U.S. states (e.g., differences in policy), meaningful impressions of regional personality differences derived from social media language remain. Put differently, the language people use in their conversations and posts on publically available social media channels, such as Twitter, reveal relevant information about the entrepreneurial vitality and capacity of any given location.

In regions where the language indicates a more entrepreneurial character (entrepreneurial constellation of all Big Five traits as studied in the profile), or (when focusing on Conscientiousness and Openness and single traits) more self-control, personal motivation to excel, better executive skills and more creativity and openness to new ideas, change and variety, there is more manifest entrepreneurial vitality (e.g., start-up rates), which is typically linked to subsequent economic growth and development (Beugelsdijk, 2010; Glaeser, Kerr, & Kerr, 2015; Stuetzer et al., 2018). Our results may also indicate that research based on language patterns revealing personality patterns – which show substantial stability at the individual and regional level (Elleman, Condon, Russin, & Revelle, 2018; Obschonka et al., 2017; Talhelm et al., 2014) – can be a useful and important tool to analyses the *future* economic trajectories of territories, helping to assess, for example, how they will react to macro-economic crises (Obschonka et al., 2016) and respond to changes in policies (Audretsch, 2003). Our research could inform the growing literature on entrepreneurial eco-systems, which is interested in the role (and measurement) of the local psychology and culture as markers, and shapers, of said eco-systems (Spigel, 2017; Stam 2017).

This future research could also test, for example, if emerging entrepreneurial eco-systems with a rapidly growing entrepreneurial vitality would show a corresponding increase in regional entrepreneurial personality over time. After all, regional personality – just like individual-level personality – should not be perfectly stable (see, for example, Specht, Egloff, & Schmukle, 2011). Systematic migration patterns can also play a major role in determining changes in regional personality (e.g., the influx of entrepreneurially-minded people as shaper of the regional personality structure) (Jokela, 2009; Obschonka et al., 2017; Rentfrow et al., 2008).

Finally, our results underscore the potential of a regional personality approach in research on regional entrepreneurial activity (Huggins & Thompson, 2017; McClelland, 1961, Obschonka et al., 2016; Stuetzer, et al., 2018). Whereas prior research relied on self-reports, which can have important limitations, our study uses a very different method – with in part very similar results (e.g., similar correlation for the entrepreneurial personality profile). So whereas the prior research had to rely on hundreds of thousands, or even millions, of people that filled out personality tests for research purposes, our study indicates that one can achieve similar results when analyzing publically available social media data by using artificial intelligence methods. This has important implications for research and practice, given that such social media datasets can deliver personality estimates for a relatively *fine-grained* spatial level (e.g., counties), as demonstrated in the present study.

Our study has several limitations. First, although we have linked regional personality differences to entrepreneurial outcomes, we cannot establish causality. As stressed earlier, our main goal was to assess whether the language-based Big Data approach analyzing digital footprints would deliver the expected links to county-level entrepreneurship. When taking the existing empirical literature pointing to causal effects of personality on entrepreneurship into account (e.g., McClelland, 1961, Obschonka et al., 2016; Stuetzer, et al., 2018), we have some indications that the effect could indeed run in the expected direction in our data, with culture as the cause. But future studies should also explore how local entrepreneurship may form and shape language patterns (that are indicative of local personality and culture) over time. For example, a region that attracts a lot of entrepreneurial talent and start-ups may also see an increase in an "entrepreneurial language" and entrepreneurial topics in social media due to the increased prevalence and social acceptance of entrepreneurial activities in the region. From this perspective, it is probably safest to say that our present results highlight that local language patterns in social media reveal a (previously unknown) statistical marker of "hard" economic activity, in our case of regional entrepreneurship.

Second, our study did not address mechanisms. *How* do regional personality traits, derived from the local social media language, affect economic outcomes such as entrepreneurship in the region? Future studies need to examine motivational aspects, such as entrepreneurial attitudes, norms and self-efficacy beliefs, and how they are shaped by the local, collective Big Five traits characterizing a territory as a whole. More research is needed in order to understand how individual psychological traits manifest in a place and transform its entrepreneurial and economic profile and how, in turn, the predominant psychological characteristics of a given area affect the economic behaviour of individuals. It would also be interesting to examine whether the language used in a region today can reliably predict its economic trajectory (e.g., boom or decline) in the next years and even decades. This would have important implications for economic policy (Audretsch et al., 2007) and research on economic growth (Glaeser, Kerr, & Kerr, 2015).

Third, our study did not address the complex interplay between regional personality, on the one hand, and entrepreneurial conditions and policy factors, on the other. Existing research found indications that regions with the highest entrepreneurship rates show both a more entrepreneurial personality and entrepreneurial eco-system (Obschonka et al., 2013, 2015; see also Carbonara et al., 2018). Our study could thus inform future research targeting this interplay when analyzing large datasets from social media sources.

To conclude, the present findings, while substantial and novel, represent only an initial step in understanding how artificial intelligence methods utilizing publicly available Big Data can "measure" the local collective psychology that is encoded in local social media language and that may shape, or interact with, economic outcomes of regions. Our psychological analysis of digital footprints underscores the usefulness of a regional personality perspective for research interested in the link between such digital footprints, analyzed by means of artificial intelligence methods, and economic factors. But the main contribution is probably the demonstrated potential, and the predictive validity of, new Big Data methods and social media data in entrepreneurship research. Indeed, we have shown that social media data, when analyzed with Big Data methods, can encode "hard-wired" psychological information (traits) that is characteristic for a region and, as such, is a marker of *economic* activity, in addition to other local outcomes (Curini, Iacus, & Canova, 2015; Eichstaedt et al., 2015; Sylwester & Purver, 2015). In any case, we hope to have planted a seed for a branch of research focused on the psychology of places, Big Data methods, and economic factors.

REFERENCES

Åstebro, T., & Thompson, P. 2011. Entrepreneurs, Jacks of all trades or Hobos? ***Research Policy,*** *40*(5): 637–649.

Audretsch, D. B. 2003. ***Entrepreneurship: A survey of the literature***. Luxembourg: Office for Official Publications of the European Communities.

Audretsch, D. B., Grilo, I., & Thurik, A. R. (Eds.). 2007. ***Handbook of research on entrepreneurship policy***. Aldershot: Edward Elgar Publishing.

Audretsch, D. B., & Keilbach, M. 2007. The localisation of entrepreneurship capital: Evidence from Germany. ***Papers in Regional Science,*** *86*(3): 351-365.

Audretsch, D. B., Obschonka, M., Gosling, S. D., & Potter, J. 2017. A new perspective on entrepreneurial regions: linking cultural identity with latent and manifest entrepreneurship. ***Small Business Economics***, *48*(3): 681-697.

Arribas-Bel, D., Kourtit, K. & Nijkamp, P. 2016. The sociocultural sources of urban buzz. ***Environment and Planning C: Government and Policy***, *34*(1): 188-204.

Back, M. D., Stopfer, J. M., Vazire, S., Gaddis, S., Schmukle, S. C., Egloff, B., & Gosling, S. D. 2010. Facebook profiles reflect actual personality, not self-idealization. ***Psychological Science,*** *21*(3): 372-374.

Barrick, M. R., & Mount, M. K. 1991. The big five personality dimensions and job performance: A meta-analysis. ***Personnel Psychology,*** *44*: 1–26.

Baumeister, R. F., Vohs, K. D., Funder, D. C. 2007. Psychology as the science of self-reports and finger movements: Or, whatever happened to actual behavior? ***Perspectives on Psychological Science,*** *2*: 96–403.

Beugelsdijk, S. 2010. Entrepreneurial culture, regional innovativeness and economic growth. In ***Entrepreneurship and culture*** (pp. 129-154). Berlin, Heidelberg: Springer.

Bosma, N. and Sternberg, R., 2014. Entrepreneurship as an urban event? Empirical evidence from European cities. ***Regional Studies***, *48*(6), pp.1016-1033.

Brandstätter, H. 2011. Personality aspects of entrepreneurship: A look at five meta-analyses. ***Personality and Individual Differences***, *51*(3): 222-230.

Carbonara, E., Santarelli, E., Obschonka, M., Tran, H. T., Potter, J., & Gosling, S. D. 2018. Agency culture, constitutional provisions and entrepreneurship: a cross-country analysis. ***Industrial and Corporate Change***, forthcoming.

Carter, N. M., Gartner, W. B., Shaver, K. G., & Gatewood, E. J. 2003. The career reasons of nascent entrepreneurs. ***Journal of Business Venturing***, *18*(1): 13-39.

Colarelli, S. M., Dean, R. A., & Konstans, C. 1987. Comparative effects of personal and situational influences on job outcomes of new professionals. ***Journal of Applied Psychology***, *72*(4): 558.

Costa, P. T., Jr., & McCrae, R. R. 1992. **Revised NEO Personality Inventory (Neo-PI-R) and NEO Five-Factor Inventory (NEO-FFI): Professional manual**. Odessa, FL: Psychological Assessment Resources.

Cuervo, A. 2005. Individual and environmental determinants of entrepreneurship. *The* ***International Entrepreneurship and Management Journal***, *1*(3): 293-311.

Curini, L., Iacus, S., & Canova, L. 2015. Measuring idiosyncratic happiness through the analysis of twitter: An application to the Italian case. ***Social Indicators Research***, *121*(2): 525-542.

Davidsson, P. 1995. Culture, structure and regional levels of entrepreneurship. ***Entrepreneurship & Regional Development***, *7*(1): 41-62.

Davidsson, P., & Wiklund, J. 1997. Values, beliefs and regional variations in new firm formation rates. ***Journal of Economic Psychology***, *18*(2-3): 179-199.

Digman, J. 1990. Personality structure: Emergence of the five-factor model. ***Annual Review of Psychology****, 41*: 417–440.

Duranton, G., Rodríguez-Pose, A., & Sandall, R. 2009. Family types and the persistence of regional disparities in Europe. ***Economic Geography***, *85*(1): 23-47.

Elleman, L. G., Condon, D. M., Russin, S. E., & Revelle, W. 2018. The personality of US states: Stability from 1999 to 2015. ***Journal of Research in Personality****, 72*: 64-72.

Eichstaedt, J. C., Schwartz, H. A., Kern, M. L., Park, G., Labarthe, D. R., Merchant, R. M., ... & Weeg, C. 2015. Psychological language on Twitter predicts county-level heart disease mortality. ***Psychological Science***, *26*(2): 159-169.

Einav, L., & Levin, J. 2014. Economics in the age of big data. ***Science****, 346*(6210): 1243089.

Ferguson, C. J. 2009. An effect size primer: a guide for clinicians and researchers. ***Professional Psychology: Research and Practice****, 40*(5): 532.

Fitjar, R.D. & Rodríguez-Pose, A. 2011. When local interaction does not suffice: Sources of firm innovation in urban Norway. ***Environment and Planning A****, 43*(6): 1248-1267.

Fouad, N. A. 2007. Work and vocational psychology: theory, research, and applications. ***Annual Review of Psychology****, 58*: 543.

Fritsch, M., & Wyrwich, M. 2014. The long persistence of regional levels of entrepreneurship: Germany, 1925–2005. ***Regional Studies****, 48*(6): 955-973.

Fritsch, M., Obschonka, M., Wyrwich, M., Gosling, S. D., Rentfrow, P. J., & Potter, J. 2018. Regionale Unterschiede der Verteilung von Personen mit unternehmerischem Persönlichkeitsprofil in Deutschland–ein Überblick [Regional differences in entrepreneurial personality in Germany]. ***Raumforschung und Raumordnung| Spatial Research and Planning***, *76*(1): 65–81.

Fruyt, F., & Mervielde, I. 1999. RIASEC types and Big Five traits as predictors of employment status and nature of employment. ***Personnel Psychology***, *52*(3): 701-727.

Furr, R. M. 2009. Personality psychology as a truly behavioural science. ***European Journal of Personality**, 23:* 369–401.

Garretsen, H., Stokera, J. I., Soudisa, D., Martin, R., & Rentfrow, J. P. 2018. The Relevance of Personality Traits for Urban Economic Growth: Making Space for Psychological Factors. ***Journal of Economic Geography***. Forthcoming.

Getis, A., &, Ord, J.K. 1992. The Analysis of Spatial Association by Use of Distance Statistics. G**eographical Analysis**, 24(3), 189-206.

Giorgi, S., Preotiuc-Pietro, D., Buffone, A., Rieman, D., Ungar, L.H., & Schwartz, H.A. 2018. The Remarkable Benefit of User-Level Aggregation for Lexical-based Population-Level Predictions. In Proceedings of the 2018 Conference on **Empirical Methods in Natural Language Processing**.

Glaeser, E. L., Kerr, S. P., & Kerr, W. R. 2015. Entrepreneurship and urban growth: An empirical assessment with historical mines. ***Review of Economics and Statistics**, 97*(2): 498-520.

Glaeser, E. L., Kominers, S. D., Luca, M., & Naik, N. 2018. Big data and big cities: The promises and limitations of improved measures of urban life. ***Economic Inquiry***, 56(1): 114-137

Goldberg, L. R., Johnson, J. A., Eber, H. W., Hogan, R., Ashton, M. C., Cloninger, C. R., & Gough, H. G. 2006. The international personality item pool and the future of public domain personality measures. ***Journal of Research in Personality**,* 40: 84-96.

Greif, A., & Tabellini, G. 2010. Cultural and institutional bifurcation: China and Europe compared. ***American Economic Review***, *100*(2): 135-140.

Guiso, L., P. Sapienza, & L. Zingales 2004. Does Culture Affect Economic Outcomes? ***Journal of Economic Perspectives**, 20*: 23-48.

Hayton, J. C., George, G., & Zahra, S. A. 2002. National culture and entrepreneurship: A review of behavioral research. ***Entrepreneurship Theory and Practice**, 26*(4): 33-52.

Hayton, J. C., & Cacciotti, G. 2013. Is there an entrepreneurial culture? A review of empirical research*.* ***Entrepreneurship & Regional Development**, 25*(9-10): 708-731.

Helminen, V. & Ristimäki, M. 2007. Relationships between commuting distance, frequency and telework in Finland. **Journal of Transport Geography**, *15(5),* 331-342.

Hisrich, R., Langan-Fox, J., & Grant, S. 2007. Entrepreneurship research and practice: a call to action for psychology. ***American Psychologist**, 62*(6): 575.

Hofstede, G., & McCrae, R. R. 2004. Personality and culture revisited: Linking traits and dimensions of culture. ***Cross-Cultural Research***, *38*(1): 52-88.

Holland, J. L. 1997. ***Making vocational choices. A theory of vocational personalities and work environments.*** Odessa, FL: PAR.

Huggins, R., & Thompson, P. 2017. The behavioural foundations of urban and regional development: culture, psychology and agency. ***Journal of Economic Geography***, forthcoming.

John, O. P., & Srivastava, S. 1999. The Big Five trait taxonomy: History, measurement, and theoretical perspectives. In L. A. Pervin, & O. P. John (Eds**.), Handbook of personality: Theory and research** (2nd ed., pp. 102-138). New York: Guilford.

Jokela, M. 2009. Personality predicts migration within and between US states. ***Journal of Research in Personality**, 43*(1): 79-83.

Jokela, M., Bleidorn, W., Lamb, M.E., Gosling, S.D., & Potter, J. 2015. Geographically varying associations between personality and life satisfaction in the London metropolitan area. **Proceedings of the National Academy of Sciences**, *112*(3), 725-730.

Kern, M. L., Eichstaedt, J. C., Schwartz, H. A., Dziurzynski, L., Ungar, L. H., Stillwell, D. J., Kosinski, M., Ramones, S. M., & Seligman, M. E. (2014). The Online Social Self: An Open Vocabulary Approach to Personality. ***Assessment,*** *21*(2), 158-169.

Kloosterman, R. C. 2010. Matching opportunities with resources: A framework for analysing (migrant) entrepreneurship from a mixed embeddedness perspective. ***Entrepreneurship and Regional Development***, *22*(1): 25-45.

Kloosterman, R., & Rath, J. 2001. Immigrant entrepreneurs in advanced economies: mixed embeddedness further explored. ***Journal of Ethnic and Migration Studies***, *27*(2): 189-201.

Knight, F. H. 1921. ***Risk, uncertainty and profit***. New York: Hart, Schaffner and Marx.

Kondo, K. 2016. Hot and cold spot analysis using Stata. ***The Stata Journal,*** *16*(3), 613–631.

Kosinski, M., & Stillwell, D. J. 2011. **myPersonality Research Wiki**. *myPersonality Project.* Retrieved from http://mypersonality.org/wiki

Kosinski, M., & Behrend, T. 2017. Editorial overview: Big data in the behavioral sciences. ***Current Opinion in Behavioral Sciences,*** *18*: iv–vi.

Kosinski, M., Stillwell, D., & Graepel, T. 2013. Private traits and attributes are predictable from digital records of human behavior. ***Proceedings of the National Academy of Sciences,*** *110*(15): 5802-5805.

Kosinski, M., Wang, Y., Lakkaraju, H., & Leskovec, J. 2016. Mining big data to extract patterns and predict real-life outcomes. ***Psychological Methods,*** *21*(4): 493-506.

Lazear, E. P. 2005. Entrepreneurship**.** ***Journal of Labor Economics,*** *23*(4): 649–680.

Lee, N. 2017. Psychology and the geography of innovation. ***Economic Geography,*** 93 (2): 106-130.

Lee, S. Y., Florida, R., & Acs, Z. 2004. Creativity and entrepreneurship: A regional analysis of new firm formation. ***Regional Studies,*** *38*(8): 879-891.

Lofstrom, M. 2013. Does self-employment increase the economic well-being of low-skilled workers? ***Small Business Economics***, *40*(4): 933-952.

McAdams, D. P., & Pals, J. L. 2006. A new big five: Fundamental principles for an integrative science of personality. ***American Psychologist***, 61(3): 204-217.

McClelland, D. C. 1961. ***The Achieving Society***. Princeton (NJ): Van Nostrand Reinhold.

McCrae, R. R. 2004. Human nature and culture: A trait perspective. ***Journal of Research in Personality,*** *38*(1): 3–14.

McCrae, R. R. 2001. Trait psychology and culture: Exploring intercultural comparisons. ***Journal of Personality***, *69*(6): 819-846.

Morstatter, F., Pfeffer, J., Liu, H., & Carley, K. M. (2013, July). Is the Sample Good Enough? Comparing Data from Twitter's Streaming API with Twitter's Firehose. In **ICWSM**.

Nathan, M. and Lee, N., 2013. Cultural diversity, innovation, and entrepreneurship: Firm-level evidence from London. ***Economic Geography***, *89*(4), pp.367-394.

Nunn, N., & Wantchekon, L. 2011. The slave trade and the origins of mistrust in Africa. ***American Economic Review,*** *101*(7): 3221-52.

Obschonka, M. 2017. The quest for the entrepreneurial culture: psychological Big Data in entrepreneurship research. ***Current Opinion in Behavioral Sciences***, *18*: 69-74.

Obschonka, M., Schmitt-Rodermund, E., Silbereisen, R. K., Gosling, S. D., & Potter, J. 2013. The regional distribution and correlates of an entrepreneurship-prone personality profile in the

United States, Germany, and the United Kingdom: A socioecological perspective. ***Journal of Personality and Social Psychology***, *105*(1): 104-122.

Obschonka, M., & Stuetzer, M. 2017. Integrating psychological approaches to entrepreneurship: The Entrepreneurial Personality System (EPS). ***Small Business Economics**, 49*: 203-231.

Obschonka, M., Stuetzer, M., Audretsch, D. B., Rentfrow, P. J., Potter, J., & Gosling, S. D. 2016. Macropsychological factors predict regional economic resilience during a major economic crisis. ***Social Psychological and Personality Science**, 7*(2): 95-104.

Obschonka, M., Stuetzer, M., Gosling, S. D., Rentfrow, P. J., Lamb, M. E., Potter, J., & Audretsch, D. B. 2015. Entrepreneurial regions: Do macro-psychological cultural characteristics of regions help solve the "knowledge paradox" of economics?. ***PloS One**, 10*(6): e0129332.

Obschonka, M., Stuetzer, M., Rentfrow, P. J., Shaw-Taylor, L., Satchell, M., Silbereisen, R. K., Potter, J., & Gosling, S. D. 2017. In the shadow of coal: How large-scale industries contributed to present-day regional differences in personality and well-being. ***Journal of Personality and Social Psychology**,* forthcoming. DOI: 10.1037/pspp0000175

Ord, J.K., & Getis, A. 1995. Local spatial autocorrelation statistics: Distributional issues and an application. **Geographical Analysis**, *27*, 286–306.

Park, G., Schwartz, H. A., Eichstaedt, J. C., Kern, M. L., Kosinski, M., Stillwell, D. J., ... & Seligman, M. E. 2015. Automatic personality assessment through social media language. ***Journal of Personality and Social Psychology**, 108*(6): 934-952.

Rentfrow, P. J., Gosling, S. D., & Potter, J. 2008. A theory of the emergence, persistence, and expression of geographic variation in psychological characteristics. ***Perspectives on Psychological Science**, 3*(5): 339–369.

Rentfrow, P. J., Gosling, S. D., Jokela, M., Stillwell, D. J., Kosinski, M., & Potter, J. 2013. Divided we stand: Three psychological regions of the United States and their political, economic, social, and health correlates. ***Journal of Personality and Social Psychology***, *105*(6): 996-1012.

Rentfrow, P. J., Jokela, M., & Lamb, M. E. 2015. Regional personality differences in Great Britain. ***PloS One****, 10*(3): e0122245.

Rieman, D., Jaidka, K., Schwartz, H. A., & Ungar, L. 2017. Domain Adaptation from User-level Facebook Models to County-level Twitter Predictions. In ***Proceedings of the Eighth International Joint Conference on Natural Language Processing***, Vol. 1, pp. 764-773.

Rodríguez-Pose, A. and Hardy, D., 2015. Cultural diversity and entrepreneurship in England and Wales. ***Environment and Planning A*****,** *47*(2), pp.392-411.

Roberts, B. W., Kuncel, N. R., Shiner, R., Caspi, A., & Goldberg, L. R. 2007. The power of personality: The comparative validity of personality traits, socioeconomic status, and cognitive ability for predicting important life outcomes. ***Perspectives on Psychological Science***, *2*(4) : 313-345.

Savickas, M. L., & Porfeli, E. J. 2012. Career Adapt-Abilities Scale: Construction, reliability, and measurement equivalence across 13 countries. ***Journal of Vocational Behavior****, 80*(3): 661-673.

Saxenian, A. 1994. ***Regional Advantage. Culture and Competition in Silicon Valley and Route 128***. Cambridge, MA: Harvard University Press.

Schmitt-Rodermund, E. 2004. Pathways to successful entrepreneurship: Parenting, personality, entrepreneurial competence, and interests. ***Journal of Vocational Behavior****, 65*: 498-518.

Schmitt-Rodermund, E. 2007. The long way to entrepreneurship: Personality, parenting, early interests, and competencies as precursors for entrepreneurial activity among the 'Termites'. In R. K. Silbereisen & R. M. Lerner (eds.), ***Approaches to Positive Youth Development*** (pp. 205-224). London: Sage.

Schumpeter, J. A. 1934. ***The theory of economic development***. Cambridge, MA: Harvard University Press.

Schwartz, H. A., Eichstaedt, J. C., Kern, M. L., Dziurzynski, L., Agrawal, M., Park, G.,...Lucas, R. E. 2013a. Characterizing geographic variation in well-being using tweets. In ***Seventh International AAAI Conference on Weblogs and Social Media***, Boston, MA.

Schwartz, H. A., Eichstaedt, J. C., Kern, M. L., Dziurzynski, L., Ramones, S. M., Agrawal, M., ... & Ungar, L. H. 2013b. Personality, gender, and age in the language of social media: The open-vocabulary approach. ***PloS One***, 8(9): e73791.

Schwartz, H. A., Giorgi, S., Sap, M., Crutchley, P., Ungar, L., & Eichstaedt, J. 2017. DLATK: Differential Language Analysis ToolKit. In **Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing: System Demonstrations** (pp. 55-60).

Schwartz, H.A., Giorgi, S., Sap, M., Eichstaedt, J. C., Kern, M. L., Ungar, L. H. (in preparation). Assessment of Regional Personality through Twitter.

Shane, S., Nicolaou, N., Cherkas, L., & Spector, T. D. 2010. Genetics, the Big Five, and the tendency to be self-employed**.** ***Journal of Applied Psychology***, *95*(6): 1154-1162.

Specht, J., Egloff, B., & Schmukle, S. C. 2011. Stability and change of personality across the life course: The impact of age and major life events on mean-level and rank-order stability of the Big Five. ***Journal of Personality and Social Psychology****, 101*(4): 862.

Spigel, B. 2017. The relational organization of entrepreneurial ecosystems. ***Entrepreneurship***

***Theory and Practice**, 41*(1): 49-72.

Stam, E. 2017. Measuring the Entrepreneurial Ecosystem. ***Utrecht School of Economics, Discussion paper series***: 17–11.

Sternberg, R. 2009. Regional dimensions of entrepreneurship. ***Foundations and Trends in Entrepreneurship**, 5*(4): 211-340.

Stuetzer, M., Audretsch, D. B., Obschonka, M., Gosling, S. D., Rentfrow, P. J., & Potter, J. 2018. Entrepreneurship culture, knowledge spillovers and the growth of regions. ***Regional Studies**, 52*(5): 608-618.

Stuetzer, M., Obschonka, M., Audretsch, D. B., Wyrwich, M., Rentfrow, P. J., Coombes, M., ... & Satchell, M. 2016. Industry structure, entrepreneurship, and culture: An empirical analysis using historical coalfields. ***European Economic Review***, *86*: 52-72.

Sullivan, S. E. 1999. The changing nature of careers: A review and research agenda. ***Journal of Management**, 25*(3): 457-484.

Sylwester, K., & Purver, M. 2015. Twitter language use reflects psychological differences between democrats and republicans. ***PloS One***, *10*(9): e0137422.

Tata, A., Martínez, D. L., García, D., Oesch, A., & Brusoni, S. 2017. The psycholinguistics of entrepreneurship. ***Journal of Business Venturing Insights***, *7:* 38-44.

Talhelm, T., Zhang, X., Oishi, S., Shimin, C., Duan, D., Lan, X., & Kitayama, S. 2014. Large-scale psychological differences within China explained by rice versus wheat agriculture. ***Science**, 344*(6184): 603-608.

Valkanas, G., Katakis, I., Gunopulos, D., & Stefanidis, A. (2014, August). Mining twitter data with resource constraints. In Proceedings of the 2014 IEEE/WIC/ACM **International Joint Conferences on Web Intelligence (WI) and Intelligent Agent Technologies (IAT**)-Volume

01 (pp. 157-164). IEEE Computer Society.

Youyou, W., Kosinski, M., & Stillwell, D. 2015. Computer-based personality judgments are more accurate than those made by humans. ***Proceedings of the National Academy of Sciences***, *112*(4): 1036-1040.

Wilson, R. E., Gosling, S. D., & Graham, L. T. 2012. A review of Facebook research in the social sciences. ***Perspectives on Psychological Science***, *7*(3): 203-220.

Zhang, Z., & Arvey, R. D. 2009. Rule breaking in adolescence and entrepreneurial status: An empirical investigation. ***Journal of Business Venturing**, 24*(5): 436-447.

Zhao, H., Seibert, S. E., & Lumpkin, G. T. 2010. The relationship of personality to entrepreneurial intentions and performance: A meta-analytic review. ***Journal of Management***, *36*(2): 381-404.

Zomaya, A. Y., & Sakr, S. 2017. ***Handbook of Big Data Technologies***. Berlin and Heidelberg: Springer.

TABLE 1
Variable Description and Summary Statistics

Variable	Description	Mean	SD	Min	Max
Start-up rate	New business starts per 1,000 population (natural log), from Statistics of U.S. Businesses	0.937	0.330	-0.108	2.467
Entrepreneurial personality profile	Goodness of fit with entrepreneurial reference profile, calculated from Twitter	-7.584	1.179	-11.003	-3.839
Openness	Big Five measure of Openness calculated from Twitter	0.108	0.656	-1.366	2.304
Neuroticism	Big Five measure of Neuroticism calculated from Twitter	0.138	0.456	-1.169	1.747
Extraversion	Big Five measure of Extraversion calculated from Twitter	-0.060	0.381	-1.235	1.097
Agreeableness	Big Five measure of Agreeableness calculated from Twitter	0.020	0.404	-1.187	1.414
Conscientiousness	Big Five measure of Conscientiousness calculated from Twitter	-0.057	0.511	-2.212	1.660
Population density (ln)	Population per square kilometre (natural log), American Factfinder	-3.314	1.374	-7.665	3.182
Unemployment rate (%)	Unemployment rate (%), American Community Survey	0.082	0.029	0.014	0.027
Median HH Income	Median Household Income (natural log) , American Community Survey	11.048	0.227	10.368	12.013
Low skill (%)	% of working age population without high-school diploma, American Community Survey	0.140	0.062	0.023	0.537
Degree + (%)	% of working age population with degree and above, American Community Survey	0.231	0.100	0.064	0.788
Herfindahl-Hirschman	Herfindahl-Hirschman Index of industrial concentration, County Business Patterns	0.047	0.023	0.004	0.245

Note. Variables cover the period 2009-2015.

TABLE 2
Correlation Table

	Start-up rate	Entrepr. Persona-lity profile	Openness	Neuro-ticism	Extra-version	Agree-ableness	Conscien-tiousness	Popula-tion density (ln)	Unemploy-ment rate (%)	Median HH Income	Low skill (%)	Degree + (%)	Herfindahl-Hirschman
Start-up rate	1.0000												
Entrepreneurial personality profile	0.4512	1.0000											
Openness	0.4970	0.5932	1.0000										
Neuroticism	-0.3678	-0.8475	-0.2904	1.0000									
Extraversion	-0.2097	0.1062	-0.5692	-0.1825	1.0000								
Agreeableness	0.1968	-0.0788	0.0886	-0.1625	-0.2512	1.0000							
Conscientiousness	0.3871	0.6487	0.3204	-0.6832	-0.1313	0.5379	1.0000						
Population density (ln)	0.2360	0.5151	0.5171	-0.3000	-0.1123	-0.2605	0.1348	1.0000					
Unemployment rate (%)	-0.3806	0.0114	-0.1181	0.0371	0.1812	-0.3492	-0.2004	-0.0387	1.0000				
Median HH Income	0.5049	0.2262	0.4192	-0.1255	-0.2565	0.1034	0.1450	0.4045	-0.5603	1.0000			
Low skill (%)	-0.4204	-0.2484	-0.3912	0.1691	0.2891	-0.2677	-0.3474	-0.2338	0.5158	-0.6340	1.0000		
Degree + (%)	0.6435	0.5796	0.6556	-0.4623	-0.3100	0.1422	0.4272	0.5199	-0.3919	0.6829	-0.6359	1.0000	
Herfindahl-Hirschman	-0.3707	-0.4352	-0.4701	0.2822	0.1575	0.1166	-0.1741	-0.2594	0.0326	-0.2438	0.2879	-0.5118	1.0000

TABLE 3

OLS Regression results: Twitter-based personality traits and entrepreneurship across U.S. counties

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)
Dependent variable:	New Business Starts per 1,000 population (ln)												
Openness			0.377***	0.338***				0.0843**					0.135***
			(0.0281)	(0.0265)				(0.0259)					(0.0293)
Neuroticism			-0.120***	-0.0502					-0.0876***				0.0498
			(0.0284)	(0.0271)					(0.0182)				(0.0256)
Extraversion			0.0464	0.124***						0.0688***			0.127***
			(0.0289)	(0.0263)						(0.0180)			(0.0223)
Agreeableness			0.0771***	-0.0241							0.0292		-0.0545*
			(0.0214)	(0.0218)							(0.0186)		(0.0214)
Conscientiousness			0.0899**	0.167***								0.120***	0.169***
			(0.0315)	(0.0323)								(0.0201)	(0.0284)
Entrepreneurial personality profile	0.382***	0.378***					0.180***						
	(0.0194)	(0.0178)					(0.0221)						
Population (ln)					-0.135***	0.0606*	-0.0137	0.0168	0.0469	0.0713*	0.0729*	0.0694*	0.00754
					(0.0314)	(0.0306)	(0.0316)	(0.0342)	(0.0305)	(0.0296)	(0.0325)	(0.0303)	(0.0362)
Unemployment %					-0.144***	-0.146***	-0.147***	-0.138***	-0.142***	-0.153***	-0.136***	-0.124***	-0.137***
					(0.0240)	(0.0282)	(0.0282)	(0.0283)	(0.0281)	(0.0284)	(0.0290)	(0.0280)	(0.0291)
Mean HH income (ln)					0.0945***	0.0890**	0.139***	0.108***	0.115***	0.0820**	0.0932**	0.123***	0.132***
					(0.0278)	(0.0308)	(0.0312)	(0.0317)	(0.0312)	(0.0311)	(0.0309)	(0.0308)	(0.0315)
Low skill (%)					0.0866***	0.0565	0.0857**	0.0604*	0.0649*	0.0591*	0.0598*	0.0911**	0.105***
					(0.0253)	(0.0295)	(0.0285)	(0.0295)	(0.0294)	(0.0293)	(0.0298)	(0.0296)	(0.0291)
Degree + (%)					0.446***	0.347***	0.266***	0.315***	0.300***	0.359***	0.339***	0.295***	0.287***
					(0.0338)	(0.0326)	(0.0342)	(0.0351)	(0.0346)	(0.0325)	(0.0335)	(0.0343)	(0.0357)
H-H index					-0.0969***	-0.0115	0.00591	-0.00273	-0.0121	-0.00902	-0.0169	-0.0207	0.00448
					(0.0203)	(0.0235)	(0.0231)	(0.0237)	(0.0233)	(0.0233)	(0.0238)	(0.0236)	(0.0235)
Constant	0.0219	0.0219	0.0218	0.0218	-0.0437*	-0.111***	-0.0500*	-0.0803***	-0.0948***	-0.118***	-0.118***	-0.108***	-0.0656**
	(0.0180)	(0.0146)	(0.0166)	(0.0143)	(0.0209)	(0.0202)	(0.0215)	(0.0224)	(0.0205)	(0.0197)	(0.0209)	(0.0200)	(0.0236)
State FE's	No	Yes	No	Yes	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	1,772	1,772	1,772	1,772	1,772	1,772	1,772	1,772	1,772	1,772	1,772	1,772	1,772
R-squared	0.204	0.489	0.321	0.513	0.454	0.601	0.619	0.604	0.607	0.605	0.602	0.613	0.626

Note. All variables standardized. Robust standard errors in parentheses. *** p<0.001, ** p<0.01, * p<0.05.

FIGURE 1
Big Data methods used to estimate regional personality differences based on Twitter language

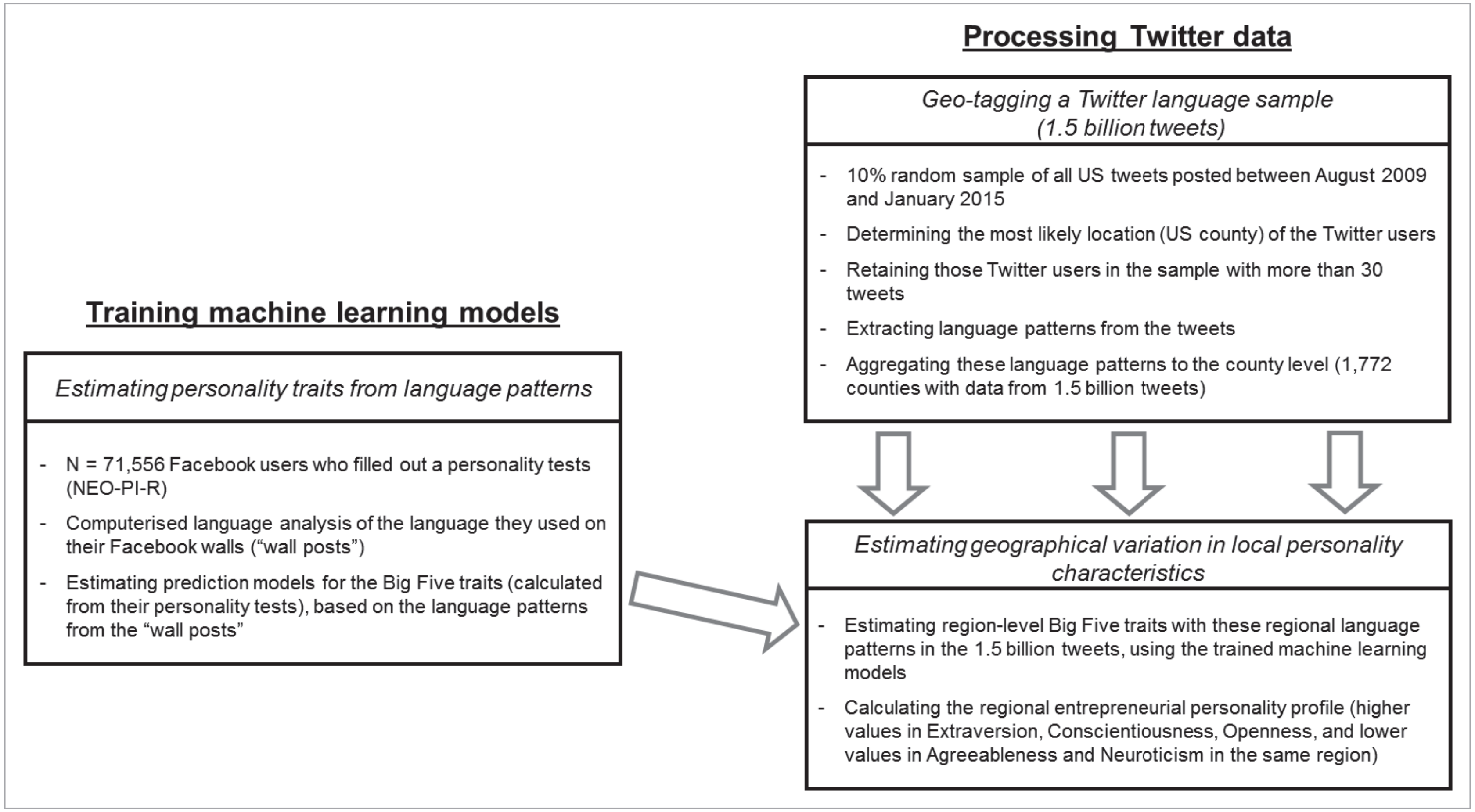


FIGURE 2

Word cloud visuals of the words and phrases most distinguishing the Big Five personality traits, adjusted for age.

Note. Please be aware that the material contained below is of a mature nature (i.e., profanity). The material was taken from http://wwbp.org/personalit y_wc.html (see also Park et al., 2015).

Low Neuroticism

Low neuroticism, also called emotional stability, is characterized by traits such as being calm, relaxed, at ease, not envious, stable, contented, and unemotional. The word cloud reflects positive social relationships (e.g., team, game, success), activities that could build life balance (e.g., blessed, beach, sports), and sport-related words (e.g., lakers, basketball, soccer).

High Extraversion

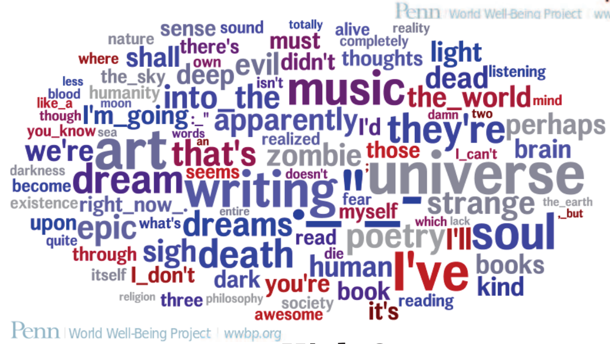
High extroversion is characterized by traits such as being energetic, talkative, bold, active, assertive, and adventurous. The word cloud reflects positive emotions (e.g., ;), excited), and social words and phrases such as party, girls, and can't_wait.

High Conscientiousness

High conscientiousness is characterized by traits such as being organized, responsible, practical, thorough, hardworking, and thrifty. The word cloud includes words reflecting achievement, school, and work (e.g., success, finals, to_work, work_tomorrow, long_day), and activities that support relaxation and balance (e.g., weekend, family, workout, vacation, day_off, lunch_with) and general enjoyment (e.g., much_fun, blessed, enjoying, wonderful).

Low Agreeableness

Low agreeableness is characterized by traits such as being cold, unkind, uncooperative, selfish, distrustful, stingy, and hostile. In the image, swear words are quite prevalent. The words surrounding these words reflect aggressiveness, substance abuse, and other words reflecting a hostile approach to the world (e.g., kill, punch, knife, drunk, i_hate, racist, idiots).



High Openness

High openness is characterized by traits such as being intelligent, analytical, reflective, curious, imaginative, creative, and sophisticated. The word cloud reflects an artistic domain (e.g., soul, dreams, universe, music). High openness also included more pronoun-type words (e.g., we're, you're, I've, I'll).

FIGURE 3
Regional differences in the entrepreneurial personality profile, calculated by means of Big Data methods analyzing 1.5 billion Twitter tweets

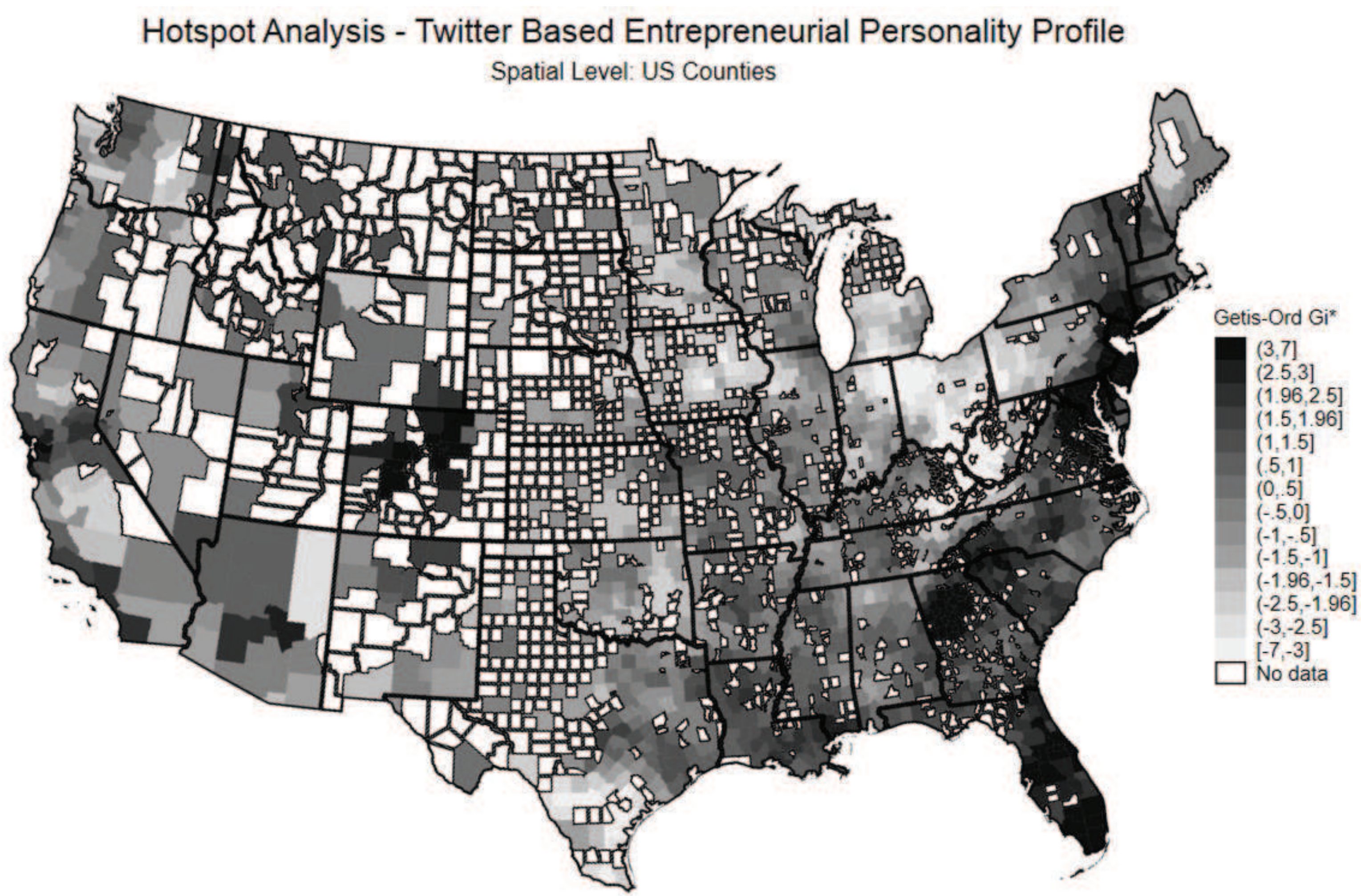


FIGURE 4
Regional differences in entrepreneurial activity (Start-up rates across U.S. counties)

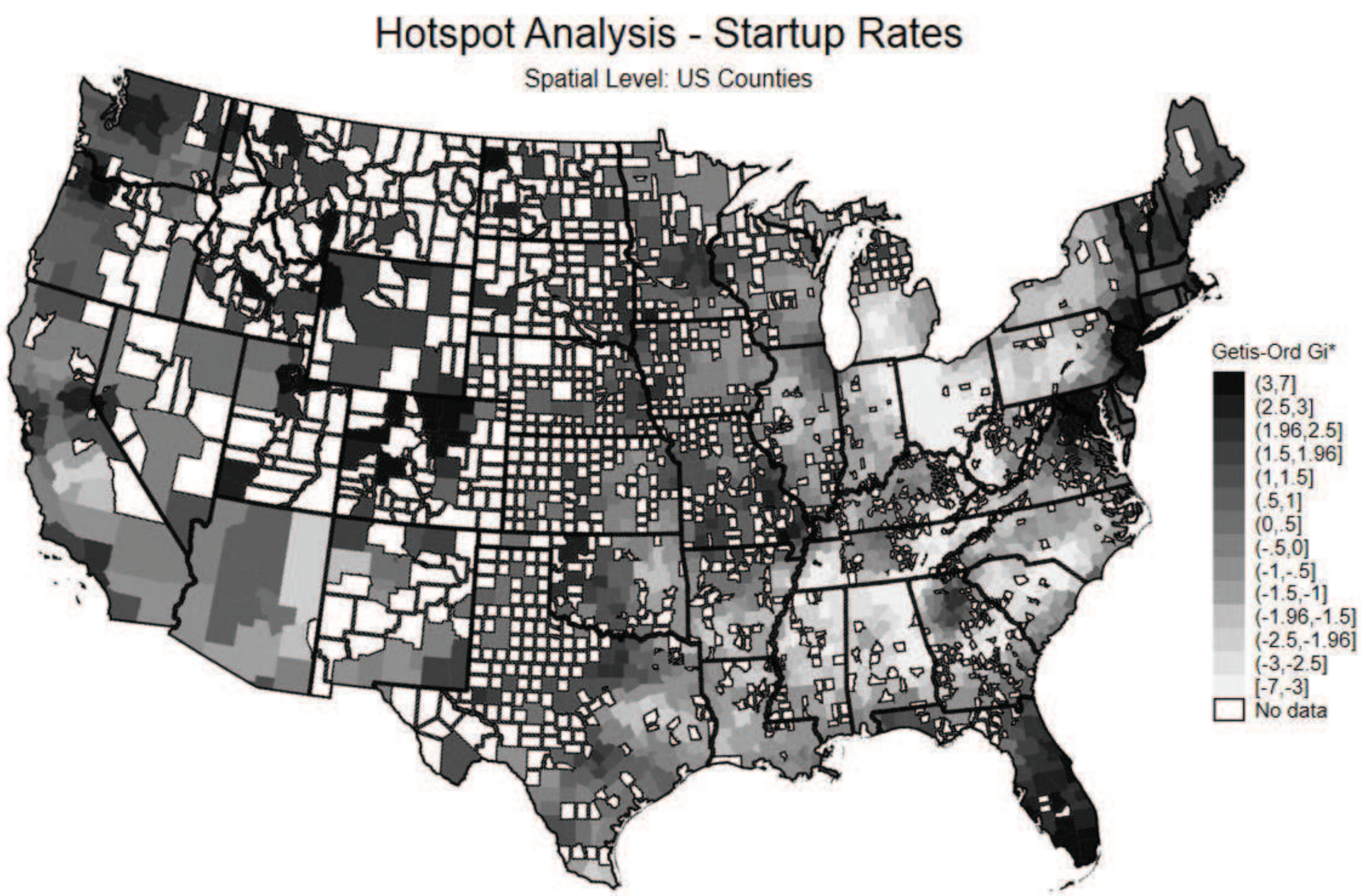
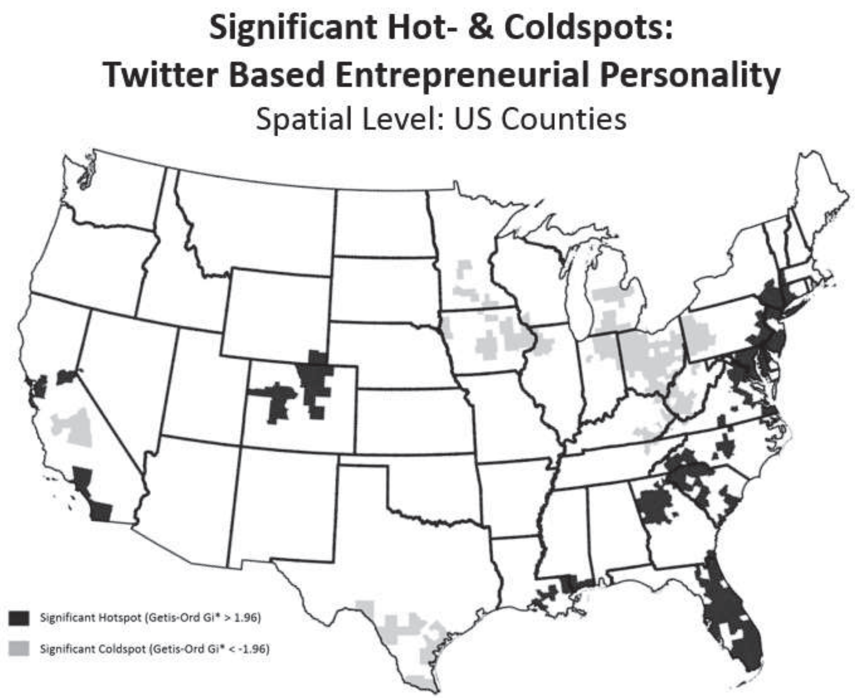
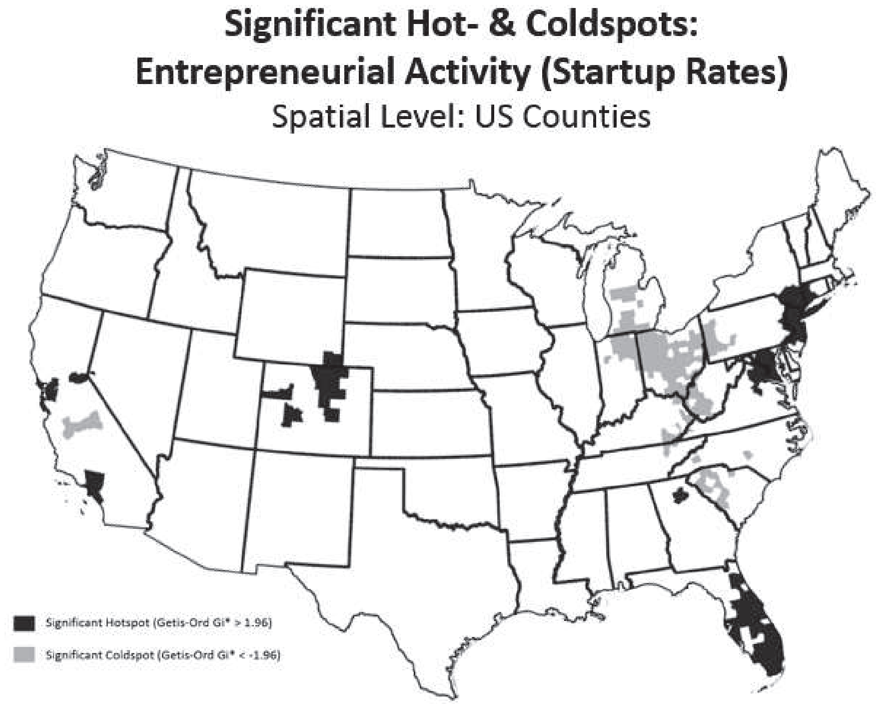


FIGURE 5
Significant hotspots and coldspots of the Twitter-based regional entrepreneurial personality profile and actual entrepreneurial activity





Appendix 1.

FIGURE A1
Bivariate relationships between Twitter calculated Big Five personality characteristics and county-level entrepreneurial activity