Article:

Resce, G. and Maynard, D. (2018) What matters most to people around the world? Retrieving Better Life Index priorities on Twitter. Technological Forecasting and Social Change, 137. pp. 61-75. ISSN: 0040-1625

https://doi.org/10.1016/j.techfore.2018.06.044

What matters most to people around the world? Exploring Better Life Index priorities on Twitter

Giuliano Resce, Department of Economics, Roma Tre University, giuliano.resce@uniroma3.it

Diana Maynard, Department of Computer Science, University of Sheffield, d.maynard@sheffield.ac.uk

Abstract

Better Life Index (BLI), the measure of well-being proposed by the OECD, contains many metrics, which enable it to include a detailed overview of the social, economic, and environmental performances of different countries. However, this also increases the difficulty in evaluating the big picture. In order to overcome this, many composite BLI procedures have been proposed, but none of them takes into account societal priorities in the aggregation. One of the reasons for this is that at the moment there is no representative survey about the relative priorities of the BLI topics for each country. Using these priorities could help to design Composite Indices that better reflect the needs of the people. The largest collection of information about society is found in social media such as Twitter. This paper proposes a composite BLI based on the weighted average of the national performances in each dimension of the BLI, using the relative importance that the topics have on Twitter as weights. The idea is that the aggregate of millions of tweets may provide a representation of the priorities (the relative appreciations) among the eleven topics of the BLI, both at a general level and at a country-specific level. By combining topic performances and related Twitter trends, we produce new evidences about the relations between people's priorities and policy makers' activity in the BLI framework.

Keywords: Big Data; Better Life Index; Composite Indicators; Twitter

JEL Classification: A13; C43; I31

1. Introduction

These days, there is a general consensus about the limits of the Gross Domestic Product (GDP) in predicting societal well-being. After the seminal work of Easterlin (1974), which clearly shows that GDP and happiness are not always positively correlated, this point has been largely discussed (among others UNDP, 1996; Fleurbaey, 2009; Stiglitz *et al.* 2010; Frey, Stutzer, 2010; Bleys, 2012; Fioramonti, 2013; Costanza *et al.* 2014; De Beukelaer, 2014; Coyle, 2014; Karabell, 2014; Costanza *et al.* 2016; Partizii *et al.* 2016). In particular, as reported in Costanza *et al.* (2016), UNDP (1996) identifies five main negative consequences of growth in GDP: 'jobless growth', 'voiceless growth', 'ruthless growth', 'rootless growth', and 'futureless growth'.

To resolve this, an increasing number of alternative measures of well-being have recently been proposed by the main international institutions, as well as by the national statistics offices (Costanza *et al.* 2014; 2016). Among them, one of the most influential is the Better Life Index (BLI), launched by the OECD in 2011, and measured in 37 countries in 2016. The BLI is based on the idea of Stiglitz *et al.* (2010), that well-being is multi-dimensional and has eight key aspects of life to take into account simultaneously: (i) Material living standards; (ii) Health; (iii) Education; (iv) Personal activities including work; (v) Political voice and governance; (vi) Social connections and relationships; (vii) Environment; (viii) Insecurity of an economic as well as a physical nature.

Based on the framework of Stiglitz *et al.* (2010), the BLI is composed of eleven topics: Housing, Income, Jobs, Community, Education, Environment, Civic engagement, Health, Life Satisfaction, Safety, and Work-Life Balance. Performances in all these topics are measured in 37 countries by means of 24 different metrics.[1] On a dedicated website[2], the OECD presents national performance (i.e. at the country level) on the whole set of topics, rather than a single Composite Index. This is a deliberate choice by the OECD to share information without any statement about the overall well-being. The information is presented in such a way that users can bring their own relative importance for each topic, and estimate their personal composite BLI, combining performance on topics with relative personal preferences (OECD, 2015). The interactive website enables people to create their own index by weighting the performance on topics according to their own viewpoint. Participants have been encouraged to create and share their own composite Better Life Index since its launch in 2011. At the time of this paper, the OECD has received and collected more than 100,000 opinions from 180 different countries.

The opinions related to the different topics (the relative appreciations) are one of the most important factors in multidimensional well-being for at least two reasons. First, the preferences of people interested in the measurement are themselves part of the phenomenon (Helliwell, 2003; Helliwell,

[1] The metrics are: dwellings without basic facilities, housing expenditure, rooms per person, household net adjusted disposable income, household net financial wealth, employment rate, job security, long-term unemployment rate, personal earnings, quality of support network, educational attainment, student skills, years in education, air pollution, water quality, consultation on rule-making, voter turnout, life expectancy, self-reported health, life satisfaction, assault rate, homicide rate, employees working very long hours, and time devoted to leisure and personal care.

[2] www.oecdbetterlifeindex.org

Barrington-Leigh, 2010), since the BLI is a metric to assess "the level of well-being of individuals with different preferences" (Stiglitz *et al.* 2010, p. 143). Second, people's preferences are eventually translated into policies by means of some mechanism of preference aggregation, so that they drive policy makers towards providing specific representations of multidimensional well-being. This issue is far more relevant when a Composite Index (CI, Nardo *et al.* 2008) is designed. The focal point in the literature related to CIs is that in order to aggregate many dimensions into one index, a choice must be made about the relative importance of each dimension, as different weights may give rise to relevant differences in the final synthetic evaluation, and thus in the ranking of countries (Sharpe, 2004; Saisana *et al.* 2005; Cherchye *et al.* 2008; Permanyer, 2011; OECD, 2014; Patrizii *et al.* 2016; Costanza *et al.* 2016; Greco *et al.* 2017).

Recently, several CIs based on the BLI have been proposed, e.g. Mizobuchi (2014), Marković *et al.* (2016), Patrizii *et al.* (2016), and Lorenz *et al.* (2016). However, none of the previous CIs has taken into account the relative appreciations in the aggregation. One of the reasons for this shortcoming is that although their relevance is widely recognized, at present there is no representative survey about the relative preferences of the various BLI dimensions. The opinions collected by the OECD's Better Life Index are not representative, since there is an intrinsic self-selection in people visiting this dedicated website (mainly economic experts).

Nowadays the largest collection of information about society available is in social media (Maynard *et al.* 2017). Twitter has become one of the most popular social media sites in the political arena. In terms of traffic, in June 2018 twitter.com is the fourth most popular social media platform in the world, after facebook.com, youtube.com, and instagram.com (SimilarWeb, 2018). In April 2018 Twitter has 330 millions of active users, ranking 12-th among the most famous social network sites worldwide (Statista, 2018). Although each tweet (an individual user post) is limited to only 140 characters, the aggregate of millions of tweets may provide a representation of public mood and sentiment about priorities (the relative appreciations)[3]. This idea has led to the development of real-time sentiment-tracking indicators such as "Pulse of Nation"[4], and "Mood of the Nation"[5] (Lansdall-Welfare *et al.* 2012; Lampos, 2012).

The aim of this paper is to estimate the societal relative appreciations of the eleven topics of the BLI, using the importance (the volume of trends) that these topics have on large-scale collections of daily Twitter posts. In other words, we use the public interest in the different topics as a proxy of societal relative appreciations. Then we propose a composite BLI based on the weighted average of country-level performances in each topic (estimated by the OECD) in relation to the relative importance that the topic has on Twitter. This procedure allows us to estimate at country level a composite BLI reflecting both the societal relative appreciation and the relative performance for each topic. As a pilot experiment, we have collected and processed a set of tweets related to BLI topics from 30 May 2017 to 30 June 2017.

---

[3] **Recently Twitter started testing the 280-character tweets, doubling the 140 characters limit.**
[4] www.ccs.neu.edu/home/amislove/twittermood
[5] http://geopatterns.enm.bris.ac.uk/mood/about.php

Among the different measures of well-being based on sets of indicators, as the Human Development Index (firstly developed by the UNDP in 1990) and the Happy Planet Index (launched by Marks et al. 2006), the BLI is particularly suitable for this kind of analysis because the indicators are not hierarchically ordered (Kerényi 2011; Lind 2014). BLI avoids the legitimacy by allowing the end user to select the relevant variables. In other words, BLI does not force one to accept a given interpretation of well-being.

It is worth noting that the analysis proposed here has the disadvantages to consider a "convenience samples" that may not be entirely representative, since it depends on who has a Twitter accounts and decides to tweet. The main problem is that what is reflected in Twitter is not necessarily what majority of people looking for it, since Twitter users consist only of a small fraction of the population. Our results can be therefore biased toward opinions of a particular group in society. However, Twitter posts are available at high frequency and granularity compared with other traditional source of data. Moreover, many traditional surveys are not immune to bias concerns (Einav, Levin, 2014). For instance, as mentioned, the opinions collected by the OECD's Better Life Index website may be mainly given by economists and policy experts.

The analysis shows that both the general and the country-level (using geo-localization) Twitter trends make a difference in the composite BLI. This affects both synthetic evaluation and the ranking of countries. The combination of information about Twitter trends and topic performances produces new evidence about the relations between people's priorities and policy makers' activity. Although the pilot study presented in this paper is only small, it nevertheless provides a framework on which further studies could be based.

The paper is organized as follows: Section 2 presents a literature review on the use of big data in social sciences; Section 3 introduces the framework and the dataset; Section 4 presents the analysis; and in Section 5 we end with discussion and conclusions.

2. Big Data in social sciences

The term Big Data is used to define the unstructured mass of data generated every day from sources such as text, voice, and video. **As McAfee et al. (2012) mention, we are "walking data generators", since our Mobile phones, online shopping, social networks, and electronic communications produce torrents of data every day.** According to Laney (2001), 'Big' refers to three main characteristics **of these data**: high speed of generation and use, high variety in terms of range and sources, and high volume in terms of amount of data. Currently there is an increasing interest in these innovative sources among social scientists, since Big Data contains information for complex phenomena that may be difficult to observe using traditional surveys (di Bella *et al.* 2016**; Einav, Levin, 2014**). **According to Einav, Levin (2014), what characterizes Big Data for economic and policy research is that they are available in real time, at a larger scale, on novel types of variables, and with less structure than traditional surveys.**

One of the most influential papers using Big Data for social phenomena is Ginsberg *et al.* (2009), in which influenza epidemics are detected using counts of relevant Google searches. For the first time, this paper made clear that new opportunities are available in real time for social scientists, and opened a debate about the usefulness of Big Data analytics (Boyd and Crawford, 2012). On the one hand, the increasing interest in Big Data comes from its low cost and its possibility to be processed in real-time, closing the time gap between observation and analysis which is a typical problem of studies made on traditional surveys (Giannone *et al.* 2008). On the other hand, some important issues remain unsolved in the use of Big Data in social science, involving in particular the definition of quality measures, and problems of privacy and transparency (di Bella *et al.* 2016).

Big Data allows social scientists to make use of timely and geo-localized information. **A prominent application in economics is the Billion Prices Project (BPP) at MIT, that has been experimenting with daily online price indexes in an increasing number of countries since 2008 (see Cavallo, 2013; Cavallo, Rigobon, 2016). After BPP, both National Statistical Offices and academic research have started to consider the use of online data in official Consumer Price Indices, in the study of price competition, market segmentation, price stickiness, international relative prices, and real exchange rate dynamics (Cavallo, 2017). Choi, Varian (2012) show that some measures of economic time series such as unemployment claims and consumer confidence can be retrieved by means of Google search engine data. Einav et al. (2013; 2014) use eBay data to estimate the effect of sales taxes on Internet commerce, the degree of price dispersion, residual demand curves, and how consumers respond to potentially nontransparent charges such as shipping fees. Other relevant applications of Big Data in economics involve** the estimates of local poverty and socio-economic status (Elvidge *et al.* 2009; Smith-Clarke *et al.* 2014; Blumenstock *et al.* 2015; Mao *et al.* 2015), food security (Dutta *et al.* 2014), and social unrest (Manrique *et al.* 2013).

The most frequent sources of Big Data among social science papers are social networks (di Bella *et al.* 2016). Social media has emerged as a promising source of societal information such as political participation, brand perception, and stock trading (Bond *et al.* 2012; Bollen *et al.* 2011); has been used for monitoring public health and health perceptions (UNICEF 2013; Garcia-Herranz *et al.* 2014; Stoové and Pedrana, 2014), forecasting immigration and mobility flows (Zagheni *et al.* 2014; Lenormand *et al.* 2014), and measuring criminal violence (Monroy-Hernández *et al.* 2013).

Among the social science papers using social networks, Twitter is the most frequent source of data (di Bella *et al.* 2016). Chang and Chu (2013) use Twitter to detect tourism preferences. Rill *et al.* (2014) detect political trends in Germany on Twitter, collecting about 4 million tweets before and during the 2013 parliamentary election. Yazdani and Manovich (2015) use Twitter images to predict socio-economic characteristics. Seabold *et al.* (2015) use Twitter data to analyze public perception of the 2011 reform to the propane gas subsidy in El Salvador. Usherwood and Wright (2017) monitored the three main groups that made extensive use of social media during the UK European Union membership referendum (Brexit referendum). Maynard *et al.* (2017) used Twitter data for the long-term monitoring of UK Members of Parliament and parliamentary candidates throughout the 2015 UK election campaign, and for short-term intensive monitoring of tweets with particular hashtags during the televised leaders' debates during the same period. The content of social media is dynamic, following trends driven by these events (sport, celebration, crises, articles), and topics such as global warming, terrorism, and immigration (Maynard *et al.* 2017).

3. Framework and dataset

The majority of analytic tools used in this paper are based on GATE (Cunningham *et al.* 2002), an open source framework for Natural Language Processing (NLP) developed by the University of Sheffield, and the social media analysis toolkit developed within it (Maynard *et al.* 2017). Our analysis has three main stages: in the first we identify topics and sub-topics related to the eleven dimensions of BLI; in the second we collect tweets containing the topics and related sub-topics selected in the first stage; and in the last we analyze the tweets and estimate two new Composite Indexes of well-being using the Twitter trends as weights.

3.1 Topic identification

The aim of this stage is to select the most relevant topics related to the eleven dimensions in the BLI (first column in Table 1). The starting point for this is the set of topics used by the OECD on their website dedicated to BLI. The website has been translated by the OECD from English into 6 other languages: Spanish, French, Russian, German, Portuguese, and Italian. This is a robust source for our analysis, since the 11 topics and their official translations can be used as baseline definitions (Table 1). These translations, however, do not cover all the languages spoken in OECD areas, but only the official languages of 21 of the 37 countries in which BLI is measured.[6] This is one of the major problems of analyzing the priorities of the BLI, because there are no official definitions of the 11 topics that are understandable to everyone in the relevant countries. This results in two main problems: (1) part of the OECD population does not have the possibility to understand, and they cannot vote on the OECD website; (2) studies that want to infer priorities in BLI with alternative tools have no set of pre-defined topics in all the languages. **Nonetheless, in terms of representativeness, the seven languages considered here cover almost 1230 million people out of 1685 million people living in the 37 countries included in the BLI project (73%), and 67.4% of tweets (Hong, 2011).**

[6] The languages in which the OECD website is translated cover Australia, Austria, almost 50% of Belgium, Brazil, Canada, Chile, France, Germany, Ireland, Italy, Luxembourg, Mexico, New Zealand, Portugal, Russia, South Africa, Spain, Sweden, Switzerland, United Kingdom, United States. The remaining countries: Czech Republic, Denmark, Estonia, Finland, Greece, Hungary, Iceland, Israel, Japan, Korea, Latvia, Netherlands, Norway, Poland, Slovak Republic, Slovenia, and Turkey are instead not covered by translation in OECD website **(Ethnologue, 2017)**.

Table 1. Baseline Topics and translations for Better Life Index on the OECD website

English	Spanish	French	Russian	German	Portuguese	Italian
Housing	Vivienda	Logement	Жилищные условия	Wohnverhältnisse	Moradia	Abitazione
Income	Ingresos	Revenu	Доход	Einkommen	Renda	Reddito
Jobs	Empleo	Emploi	Работа	Beschäftigung	Empregos	Occupazione
Community	Comunidad	Liens sociaux	Общество	Gemeinsinn	Comunidade	Relazioni sociali
Education	Educación	Éducation	Образование	Bildung	Educação	Istruzione
Environment	Medio ambiente	Environnement	Экология	Umwelt	Meio ambiente	Ambiente
Civic Engagement	Compromiso cívico	Engagement civique	Гражданские права	Zivilengagement	Engajamento cívico	Impegno civile
Health	Salud	Santé	Здоровье	Gesundheit	Saúde	Salute
Life Satisfaction	Satisfacción	Satisfaction	Удовлетворенность	Lebenszufriedenheit	Satisfação pessoal	Soddisfazione
Safety	Seguridad	Sécurité	Безопасность	Sicherheit	Segurança	Sicurezza
Work-Life Balance	Balance vida-trabajo	Équilibre travail-vie	Работа / Отдых	Work-Life-Balance	Vida/Trabalho	Equilibrio lavoro-vita

Source: http://www.oecdbetterlifeindex.org (Accessed 23 June 2017)

The main problem with the topics in Table 1 is that they may be represented in different ways in the text. For instance, the topic "housing" includes mentions of words such as "landlord", "rent", "homebuilding" etc., while the topic "environment" comprises words like "climate change", "global warming", "sustainability" and so on. To deal with this, after the selection of topics in Table 1, we make an expansion for each of them by following a two-step procedure. In the first step, we use the lists of political sub-topics used in Maynard *et al.* (2017) for monitoring the UK 2015 Election. The topic detection there is performed by means of gazetteer lists manually created and then extended semi-automatically. Terms are matched in text under any morphological variant, including also hyponyms and hypernyms. Since these lists are only available for English, we translate them into all the remaining six languages with Google Translate, and manually check them for possible mistakes.

After this process, we follow the procedure in (Maynard *et al.* 2017) for expanding them using Word2vec (Mikolov *et al.* 2013). Word2vec finds words semantically similar to others by means of similarity measures. The similarity is measured on a large unlabeled corpus, based on the notion that semantically similar words have similar context. The main output of Word2vec is a fixed-length vector for each word (word representation). These vectors are then used to find similarity among words with standard similarity measures. In this study we use the approach proposed by Bojanowski (2016), which is based on the skip-gram model, where each word is represented as a bag of character n-grams. This method outperforms the baseline (Mikolov *et al.* 2013), since it takes into account sub-word information, rare words, and morphologically rich languages (Bojanowski, 2016).

We use the pre-trained word vectors, trained on Wikipedia by Bojanowski *et al.* (2016), which are available for 294 languages. This source allows us to use Word2vec for each of the 11 topics in each of our 7 languages. To prevent over-generation, we limit our expansion to 10 sub-topics for each topic in each language, so that we end up with 770 topics and sub-topics in total (11 topics for 7 languages for 10 sub-topics). These are then used as keywords for our Twitter collector.

3.2 Data collection

We collected the tweets using the GATE Twitter Collector.[7] This is an easy-to-use service which enables tweets to be collected based on a number of different criteria such as keywords, hashtags, geolocations, authors etc. Topics in tweets are often denoted by hashtags, e.g. #OlympicGamesLondon2012, and so we manually converted all our topics into hashtags in order to use them as input for the Twitter collector. This transforms for example "Life Satisfaction" to "#LifeSatisfaction". The tweet collector finds and stores tweets containing at least one of these 770 hashtags (the complete lists of hashtags are in Appendix A.1).

We collected in this way all tweets containing at least one of these selected hashtags from 30 May 2017 to 30 June 2017. **The root-level data structures for Tweet activities, are made of 15 elements: 1. id: a unique IRI for the tweet; 2. actor: an object representing the twitter user who tweeted; 3. verb: the type of action being taken by the user (e.g., Tweets, Retweets, and Deleted Tweets); 4. generator: an object representing the utility used to post the Tweet; 5. provider: a JSON object representing the**

[7] https://cloud.gate.ac.uk/info/about/twitter.html

provider of the activity; 6. inReplyTo: a JSON object referring to the Tweet being replied to, if applicable; 7. location: a JSON object representing the Twitter "Place" where the tweet was created; 8. geo: point location where the Tweet was created; 9. twitter_entities: the entities object from Twitter's data format which contains lists of urls, mentions and hashtags; 10. twitter_extended_entities: an object from Twitter's native data format containing "media"; 11. link: a Permalink for the tweet; 12. body: the tweet text; 13: objectType: "activity"; 14. object: an object representing tweet being posted or shared; 15. postedTime: the time the action occurred, e.g. the time the Tweet was posted.

GATE Twitter Collector makes use of the statuses/filter API, which allows the user to specify certain constraints and then delivers up to a maximum of around 50 tweets per second that match those constraints (sleeted hashtags in our case). Since we have been collecting tweets from different GATE collectors (one of which having a certain number of hashtags), the tweets containing keyword stored by different collectors have been collected twice or more times. Before performing the text mining and further processing the data, we excluded tweet collected more than one time by means of the id element (the unique IRI for the tweet). After this cleaning procedure, the analysis presented in section 4 is made on the body element of each tweet.

3.3 Composite Index of Well-Being

Following the OECD Handbook on Composite Index (Nardo et al, 2008), we have the set A of 37 countries to be evaluated on the set G of topics (the 11 dimensions of the BLI):

$$(1) \quad A = \{a_1, \dots, a_m\}$$

$$(2) \quad G = \{g_1, \dots, g_n\}$$

Where $\boldsymbol{m = 37}$ and $\boldsymbol{n = 11}$. The performance in each of the 11 topics is estimated by the OECD by means of 24 variables. Each topic is composed of one or more of the 24 variables (see Table A.4.1 in the Appendices), of which 16 have a positive effect on well-being (e.g. rooms per person) and 8 have a negative effect on well-being (e.g. long-term unemployment rate). In order to group 24 variables into 11 topics, the OECD first normalizes the value each variable takes, so that they all are in the $(0{:}\,1)$ range (min max method):

$$(3) \quad index = \left(\frac{observed\ value - minimum\ value}{maximum\ value - minimum\ value}\right)$$

Secondly, variables that have a negative effect on well-being undergo a unit translation $(1 - index)$ in order to make the complement to one comparable with the variables that have a positive effect on well-being. Thirdly, the indices so obtained are aggregated into 11 topics by simple average:

$$(4) \quad g_i = \left(\frac{\sum_{j=1}^{N} index_j}{N}\right);\ (i = 1, \dots, n)$$

Where $\boldsymbol{N}$ represents the number variables in the $\boldsymbol{i}$-th topic (see Table A.4.1 in the Appendices). The individual function that aggregates topics can be assumed as the weighted sum of topic performance multiplied by the relative weight (relative appreciation). Given the individual relative appreciations expressed by a vector of weights $w = \{(w_1, \dots, w_n); \sum_{i=1}^{n} w_i = 1\}$, for each country $a_k \in A$, we can estimate the following individual composite BLI:

$$(5) \quad CI(a_k, w) = \sum_{i=1}^{n} w_i g_i(a_k);\ k = 1, \dots, m$$

where w_i reflects the importance that the citizen gives to the topic i, and $g_i(a_k)$ is the performance of the country a_k in the topic i. The problem is that the order of importance changes among people across countries, which implies that we do not have the vector of w.

The simplest way would be to assume that each citizen gives the same importance to each topic. In this case, a composite BLI can be obtained by the simple arithmetic mean; this assumption would amount to having $w_1 = w_2 = w_3 = \cdots = w_i$ in equation (5). This is the baseline Composite Index used by the OECD, and one of the most popular ways to build Composite Indices (see among others Floridi *et al.* 2011). However, this method implicitly assumes equal preferences among the 11 dimensions related to well-being. If this is not the case (as we assume), then the weights should also be different.

To estimate the composite BLI on the basis of trends detected on Twitter, we (a) determine the frequency f_i of hashtags associated with the ith topic in the **body element of each tweet** collected; and (b) compute a weighted average using the relative frequency:

$$(6) \quad \frac{f_i}{\sum_{i=1}^{n} f_i}$$

The relative frequency (6) is used as weight (w_i) associated to the ith topic:

$$(7) \quad CI(a_k, f) = \sum_{i=1}^{n} \left(f_i \Big/ \sum_{i=1}^{n} f_i \right) g_i(a_k)$$

where $(f_i / \sum_{i=1}^{n} f_i)$ reflects the relative importance that the topic i has on Twitter, and $g_i(a_k)$ is the performance of the country a_k in the topic i.

There are two ways in which we can use Twitter relative appreciations in our CI model. The first is to assume that the importance of each topic is the same for all countries, and to use the general trend. However, a better way might be to assume that each country has a potentially different relative appreciation, e.g. that people in France might consider jobs their top priority, but people in Germany might consider health. In economic theory, this point has been partially addressed with the seminal work of Tiebout (1956) regarding the public services. Adapting the Tiebout's framework to multidimensional well-being, we argue that citizens/voters have their optimal subjective mix of well-being, and that policy makers act in providing a specific proportion among the single dimensions of well-being (mix of well-being). For instance, in the same country there could be a relevant share of people interested in a specific aspect of well-being, such as health care, and at the same time there could be policy makers who are devoting more resources to education than to health. In this context, since the objective function of the policy maker is to be (re-)elected, policy makers should act according to local preferences (for a broad review of political economic models see Persson and Tabellini, 2002; Dranzen, 2004; Alesina and Giuliano, 2009).

With this model in mind, the best way to evaluate the overall national BLI should be using local preferences. However, the provision of some dimensions of the BLI go beyond national borders (the sustainability topics are the most evident case), and indeed many of them are regulated by supranational institutions. Moreover, in our dataset only a small proportion (approximately 2.5%) of tweets is geo-localized (which is in line with other findings, see Jurgens *et al.* 2015). We therefore decided to estimate the composite BLI with both global and local trends, as follows:

1. we estimate the general $f_i/\sum_{i=1}^{n} f_i$ for the whole dataset, and we use this to estimate a $CI(a_k, f)$ that reflects relative importance in the whole Twitter community;

2. we estimate the country level $f_i/\sum_{i=1}^{n} f_i$ using the geo-localizations associated with tweets, and use them to estimate a $CI(a_k, f)$ in which each country is ranked on the basis of the relative importance in tweets from people living in that country.

In the next section we present the results of the analysis, showing the composite BLI for both methods.

4. Analysis

Our final dataset contains 7,905,317 tweets, of which 197,612 (2.5%) are geo-localized. We first introduce the analysis without geo-localization, proposing a Composite Index of well-being using the general frequency representing the BLI relative importance in all the tweets (section 4.1). Then, using the subset of geo-located tweets, we propose a Composite index in which each country is evaluated with relative importance based on the people tweeting from that country (section 4.2). We assume that most people tweeting from a country are living there.

4.1 Composite index using general frequency

In order to estimate the Composite index with general frequency, we estimate first the topic frequency in the 7,905,317 tweets using (6). In Figure 1 we show the topic frequency of all the collected tweets. In our tweet collection, the most frequent topic is Jobs (23%), followed by Environment (18%), Health (16%), Housing (9%), Safety (7%), Civic Engagement (7%), Community (6%), Education (6%), Income (4%), Life Satisfaction (3%), and Work Life Balance (0.05%).

These global trends are roughly in line with the two main existing large-scale surveys about priorities and views on multidimensional well-being: "the Values of Europeans", and "MY World".

"The Values of Europeans" (Eurobarometer, 2012) is a 2012 study conducted in the 27 Member States of the European Union, the 6 candidate states, and in the Turkish Cypriot Community, about the values of Europeans. In the section "the idea of happiness", the value which Europeans consider the most important to their happiness is Health (75%), followed by Love (41%), and Work (40%). While Love is not in the BLI, Health and Jobs are also in the first three positions in our dataset.

"MY World" is a global survey for citizens led by the United Nations, that asks people which six of sixteen possible issues they think would make the most difference to their lives.[8] Currently, MY World has more than 9 million votes from all over the world. Globally, the top 3 topics in this survey are: good education, better healthcare, and better job opportunities. Again, 2 of our top 3 topics are in the top 3 ranks in MY World.

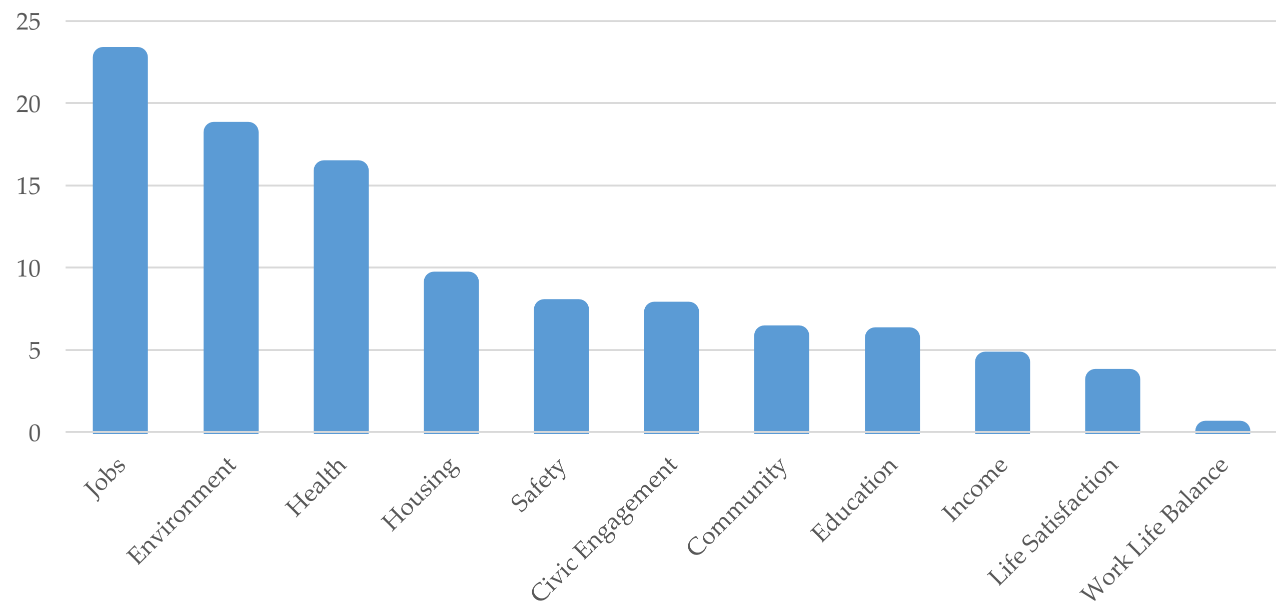
The results from our dataset are thus similar to these two surveys, but quite different from the main results of the OECD dataset. In order to compare our results with this data, we downloaded the OECD dataset (127,136 relative appreciations in 23 June 2016) and estimated the average and median for each topic (Table A.2.1 in the Appendices). Comparing the averages in Table A.2.1, people voting on the OECD website put Health in first place, then Life Satisfaction, Education, Work-Life Balance, Safety, Environment, Jobs, Housing, Income, Community, and Civic Engagement. The main difference between Twitter trends and the OECD data lies in four topics: Jobs, Environment, Life Satisfaction, and Work Life Balance. Jobs and Environment are in the first two ranks on Twitter, but 7th and 6th on the OECD website. On the other hand, Life Satisfaction and Work Life Balance are in the last two ranks in Twitter trends, and in 2nd and 4th rank in the OECD dataset.

One of the explanations for the high frequency of the topic Jobs in our dataset is that a significant number of vacancies and job offers are posted with the hashtag #jobs. While this can be perceived as a problem for our analysis, it shows that the jobs are nevertheless an important dimension in our lives, maybe more than people's perception when they are directly asked about their priorities on the BLI website. One confirmation of this is the relevant role of topics related to jobs and immigration in the successful Donald Trump presidential campaign in 2016 in the USA (e.g. "put American workers first"), and in the successful Brexit campaign in 2016 in UK (Usherwood and Wright 2017; Maynard *et al.* 2017).

The high frequency of the environment topic in our dataset is another outlier. This may be partly due to the timing of our collection, which coincides with the period in which Donald Trump pulled the US out of the Paris climate agreement. This event triggered a huge reaction in the Twitter community, and results in a massive increase of tweets containing hashtags such as #climatechange and related sub-topics. In line with Maynard (2017), this confirms that the BLI opinions detected on Twitter are dynamic, and fluctuate according to recent events. Nonetheless, this shows that people care about the environment, maybe again more than is reflected by the BLI website. A larger study over a more extended time period should help smooth out these temporal biases.

[8] The sixteen dimensions are: Better job opportunities, Support for people who can´t work, A good education, Better healthcare, Affordable and nutritious food, Phone and internet access, Better transport and roads, Access to clean water and sanitation, Reliable energy at home, Action taken on climate change, Protecting forests, rivers and oceans, Equality between men and women, Protection against crime and violence, Political freedom, An honest and responsive government, and Freedom from discrimination and persecution.

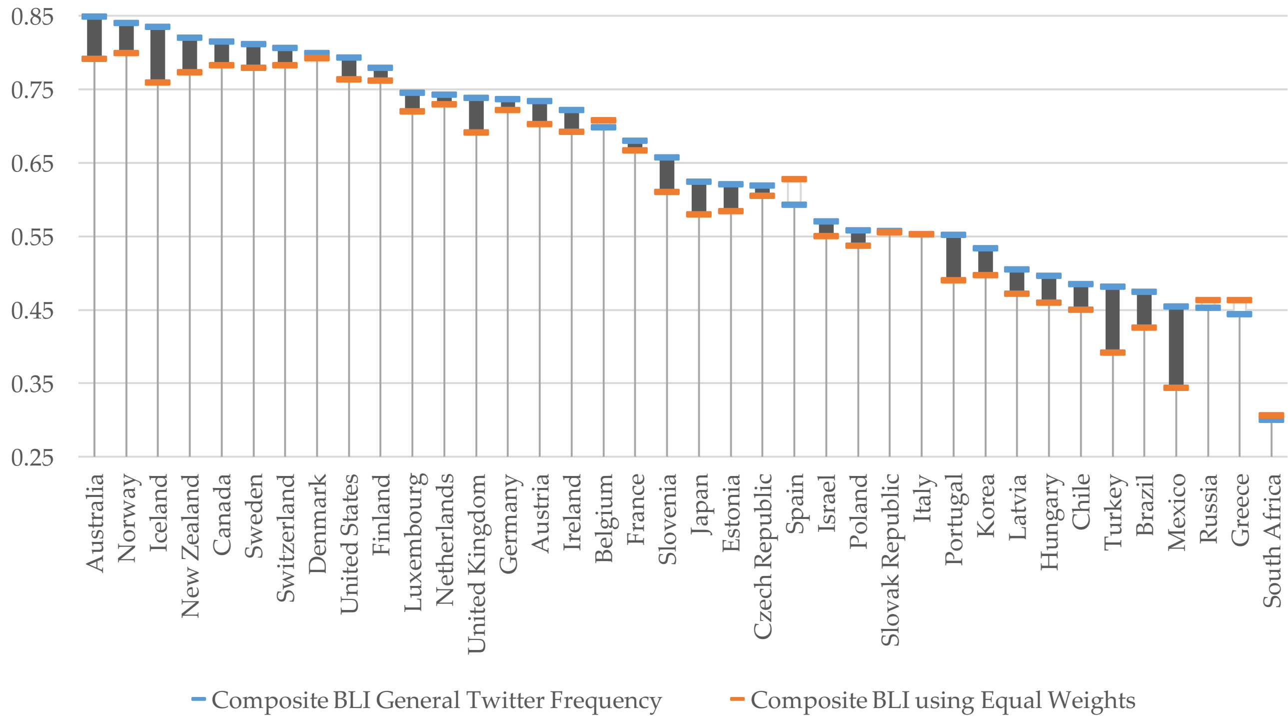
Figure 1 Better Life Index Topics frequency in Twitter (percentage)



Using the topic frequency in Figure 1, we estimate the composite BLI for each of the 37 countries included in the OECD dataset. In Figure 2 we present the composite BLI using Twitter trends as weights, and the one using equal weights (as the baseline presented by the OECD). Surprisingly, there is a pervasive difference both in the indices and in the rank that countries obtain with the different assumptions (rank correlations are shown in Table A.3.1 in the Appendices). Comparing the composite BLI using Twitter trends with the one using equal weights, there emerges a huge improvement (at least 3%) in the composite BLI of 21 countries.[9] Some of these countries, in particular the Scandinavian ones, New Zealand, and Australia, score well in topics that are more frequent on Twitter (Jobs and Environment). The improvement in Mexico, Brazil, and Turkey is instead due to their low performances in topics less frequent on Twitter (Life Satisfaction and Work Life Balance). In general, countries that have improvement in the composite BLI have relative performance among topics that reflect general frequency on Twitter. In other words, they have high performance in topics more frequent on Twitter, and low performance in topics less frequent.

[9] Mexico, Turkey, Iceland, Portugal, Australia, Brazil, Slovenia, United Kingdom, New Zealand, Japan, Norway, Estonia, Hungary, Korea, Chile, Latvia, Sweden, Austria, Canada, Ireland, and United States

Figure 2 Composite BLI using global Twitter trends as Weights and using Equal Weights, black bar means gain in using Twitter frequency, white bar means loss in using Twitter frequency



Four countries in Figure 2 show differences between the indices that are not statistically significant (less than 1%): Denmark, Slovak Republic, Italy, and South Africa. This is partly due to their low variability among performances on different topics, which mitigates the impact of differentiating weights on the composite BLI.

Relevant worsening in the composite BLI when the Twitter trends are taken into account are seen for Russia, Belgium, Greece, and Spain. The relative performance of these countries is unbalanced on topics less frequent on Twitter. For instance, in Greece and Spain, this is due to their low performance in Jobs (which is the most important topic on Twitter).

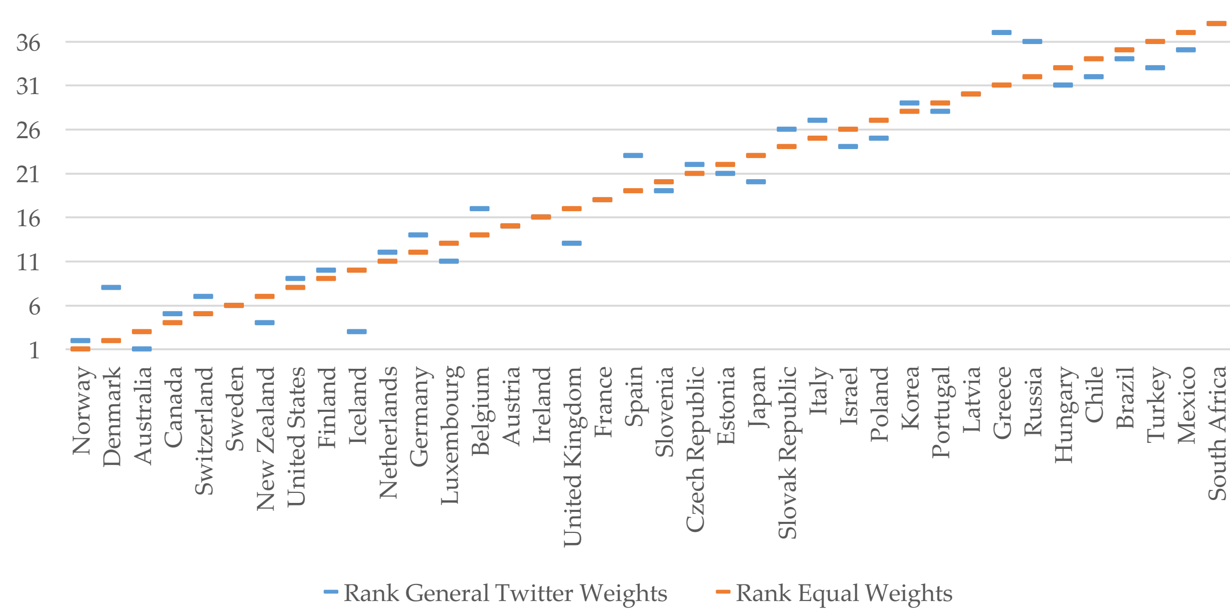
In Figure 3 we show the difference in rank between the composite BLI using equal weights, and the one using Twitter frequency as weights. The differences between the two ranks are significant. Sixteen countries show a gain in the rank[10]. Six countries show no difference[11]. The remaining sixteen have losses in the rank[12].

[10] Iceland, United Kingdom, New Zealand, Japan, Turkey, Australia, Luxembourg, Israel, Poland, Portugal, Hungary, Chile, Mexico, Slovenia, Estonia, and Brazil

[11] Sweden, Austria, Ireland, France, Latvia, and South Africa

[12] Norway, Canada, United States, Finland, Netherlands, Czech Republic, Korea, Switzerland, Germany, Slovak Republic, Belgium, Italy, Spain, Russia, Denmark, and Greece

Figure 3 Rank in Composite BLI using global Twitter trends as Weights and using Equal Weights



The main result in this section is that the general Twitter trends make a difference to the composite BLI. Measuring performance taking into account Twitter trends affects both the level of Composite Indices and the rank obtained by countries. The directions of these differences can provide information about the relationship between relative performance and the relative trends in the Twitter community. They can help us understand the extent to which policy makers act to provide a mix of well-being that is in line with the global societal priorities. Losses (or gains) in the composite BLI when Twitter trends are taken into account reflect a mismatch (or match) between people's priorities and policy makers' activity. Since policy makers are supposed to act in the interests of their citizens, in the next section we propose a composite BLI that takes into account only the local preferences for each country.

Table 2 Better Life Index Topics frequency at county level in Twitter (percentage)

Country	HO	IN	JO	CO	ED	EN	CE	HE	LS	SA	WLB
Australia	7.98	3.19	14.01	2.35	3.72	26.86	11.04	7.58	20.66	2.62	0.00
Austria	5.58	1.40	7.44	0.00	1.86	65.58	8.84	4.19	3.72	1.40	0.00
Belgium	5.84	2.35	11.24	10.71	6.10	30.05	13.50	4.97	7.58	7.49	0.17
Brazil	5.55	1.18	11.22	2.07	5.40	43.75	10.64	4.79	13.75	1.63	0.02
Canada	41.70	1.09	10.99	7.26	3.48	19.10	3.96	4.80	3.38	4.17	0.07
Chile	7.85	2.22	16.51	1.26	4.07	34.79	9.25	7.70	13.69	2.66	0.00
Czech Republic	9.90	0.50	5.94	3.47	3.96	47.52	11.88	4.95	5.45	6.44	0.00
Denmark	5.50	0.61	7.95	2.14	3.36	52.91	10.70	7.03	7.03	2.75	0.00
Estonia	7.96	0.00	7.96	1.77	4.42	59.29	6.19	1.77	3.54	7.08	0.00
Finland	2.43	1.62	46.81	1.62	1.82	32.15	5.36	2.33	3.44	2.43	0.00
France	6.88	1.55	34.27	2.56	2.50	26.39	10.40	4.73	6.38	4.30	0.04
Germany	4.44	0.93	54.37	1.74	1.62	21.16	6.75	2.76	3.60	2.23	0.39

Greece	5.20	0.78	4.29	1.69	2.34	67.62	5.98	1.82	9.75	0.52	0.00
Hungary	7.26	0.99	10.23	1.98	4.95	33.99	19.47	9.24	3.63	8.25	0.00
Iceland	2.89	0.00	1.24	1.65	0.00	89.26	2.07	0.41	2.07	0.41	0.00
Ireland	11.07	1.72	7.71	7.99	5.44	41.65	6.81	8.35	4.90	4.26	0.09
Israel	7.37	4.21	7.89	4.74	8.42	30.53	7.37	2.11	8.42	18.95	0.00
Italy	6.60	1.21	9.12	3.05	3.70	51.36	8.10	4.30	9.07	3.47	0.02
Japan	3.78	0.69	4.98	1.55	4.12	57.90	9.79	3.78	11.00	2.41	0.00
Korea	16.79	0.73	8.03	2.19	5.84	43.80	7.30	5.84	6.57	2.92	0.00
Latvia	8.93	1.79	3.57	0.00	9.82	61.61	4.46	1.79	6.25	1.79	0.00
Luxembourg	3.64	5.45	20.91	8.18	7.27	27.27	10.91	3.64	6.36	6.36	0.00
Mexico	5.71	0.87	12.22	1.96	3.36	42.22	10.77	8.05	12.22	2.62	0.00
Netherlands	8.23	1.71	7.77	3.89	5.03	42.06	12.46	4.69	6.86	7.20	0.11
New Zealand	8.42	1.63	7.34	4.08	3.80	50.00	4.62	7.34	10.33	2.17	0.27
Norway	3.59	0.60	5.09	5.39	1.80	68.56	4.49	2.40	5.09	2.99	0.00
Poland	6.26	1.50	8.03	0.82	2.04	59.18	6.53	8.98	4.76	1.90	0.00
Portugal	8.29	1.11	9.53	2.35	3.84	54.08	8.42	2.72	8.66	0.87	0.12
Russia	4.28	2.02	7.05	1.51	1.76	58.94	12.59	2.27	6.80	2.77	0.00
Slovak Republic	6.06	3.03	5.05	2.02	6.06	43.43	22.22	2.02	5.05	5.05	0.00
Slovenia	5.26	0.66	4.61	0.66	1.97	62.50	5.26	15.79	3.29	0.00	0.00
South Africa	6.31	1.99	14.94	6.64	9.63	31.87	5.23	8.80	11.04	3.57	0.00
Spain	7.02	3.50	12.64	2.43	4.69	39.33	10.09	8.13	9.66	2.51	0.01
Sweden	4.45	0.77	4.61	5.38	3.99	59.75	8.60	6.14	3.69	2.46	0.15
Switzerland	6.31	1.99	14.94	6.64	9.63	31.87	5.23	8.80	11.04	3.57	0.00
Turkey	5.03	0.24	8.53	5.70	2.27	61.75	3.81	4.69	7.31	0.66	0.00
United Kingdom	10.38	3.74	8.42	10.53	7.03	26.70	11.85	10.04	4.26	6.90	0.15
United States	5.21	1.57	14.06	5.07	4.59	27.68	6.15	18.99	4.61	11.99	0.06
No-Geol.	9.15	4.35	23.01	5.90	5.76	17.83	7.31	16.03	3.14	7.48	0.05

Notes: HO=Housing, Income, JO=Jobs, CO=Community, ED=Education, EN=Environment, CE=Civic engagement, HE=Health, LS=Life Satisfaction, SA=Safety, WLB=Work-Life Balance; Red represents High Value and blue represents Low Value

4.2 Composite index using country-specific frequency

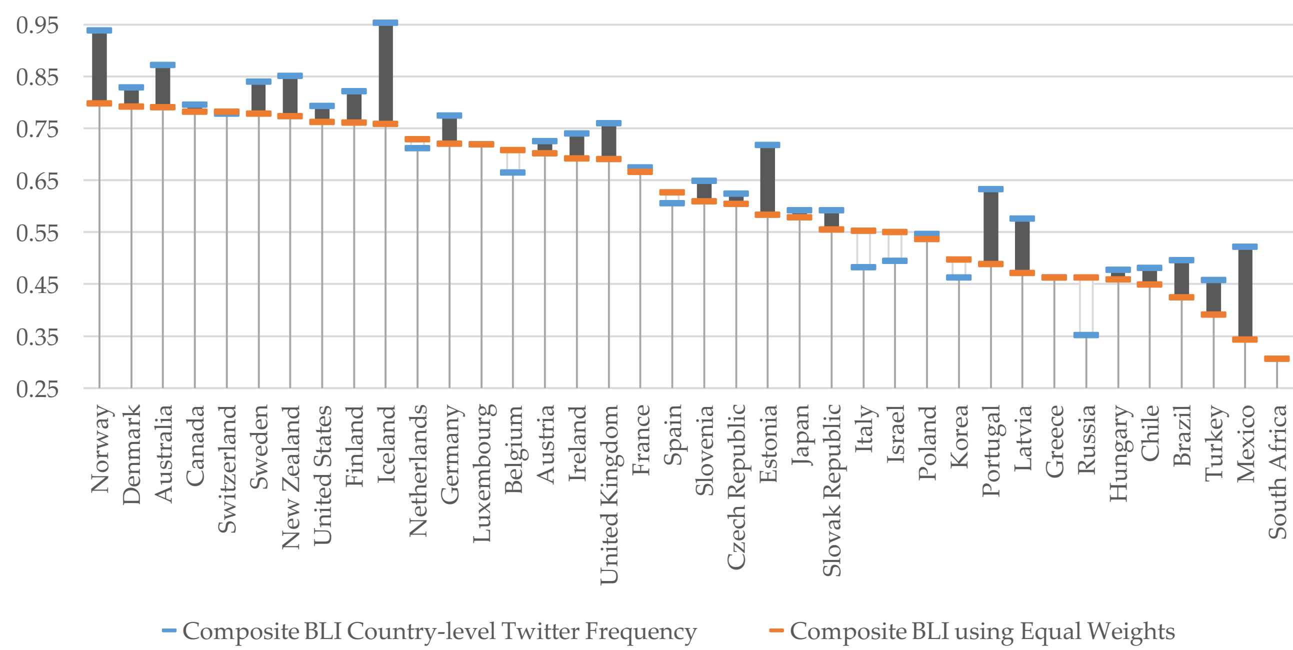
In this section, we estimate the composite BLI using differentiated weights for different countries. The weights used here are the relative frequency of tweets geo-localized in the country. In this case each country is evaluated on the basis of internal trends. Note that this analysis involves just a small part of our dataset: 197612 tweets (2.5%) containing geo-localizations[13].

In Table 2, we show the country-level relative frequency of BLI topics on Twitter. There are significant differences among countries. In the geo-localized tweets, the most frequent topic is "Environment".

[13] The selected geo-localized tweets are: Australia 3,370, Austria 202, Belgium 1,094, Brazil 5,700, Canada 9,657, Chile 1,209, Czech Republic 184, Denmark 307, Estonia 104, Finland 954, France 9,773, Germany 6,066, Greece 725, Hungary 279, Iceland 234, Ireland 1,045, Israel 183, Italy 11,321, Japan 560, Korea 129, Latvia 104, Luxembourg 103, Mexico 3,972, Netherlands 829, New Zealand 349, Norway 304, Poland 673, Portugal 751, Russia 369, Slovak Republic 90, Slovenia 143, South Africa 1,148, Spain 7,969, Sweden 566, Switzerland 887, Turkey 2,578, United Kingdom 18,444, and United States 65,661. 7,707,705 tweets are no geo-localized.

High frequency on this topic (above 60%) is found in Iceland, Norway, Greece, Austria, Slovenia, Turkey, and Latvia. Relatively low frequency (less than 30%) is found in United States, Luxembourg, Australia, United Kingdom, France, Germany, Canada, and in the non-geo-localized tweets. The topic "Jobs" is the second most important topic on average among the geo-localized tweets, but with much variation among countries. Its relative frequency is more than 40% in Germany and Finland, and less than 5% in Japan, Sweden, Slovenia, Greece, Latvia, and Iceland. Work Life Balance is confirmed as the least frequent in all countries (possibly because people only talk about this topic implicitly).

Figure 4 Composite BLI using national Twitter trends as weights and using equal weights, black bar means gain in using Twitter frequency, white bar means loss in using Twitter frequency

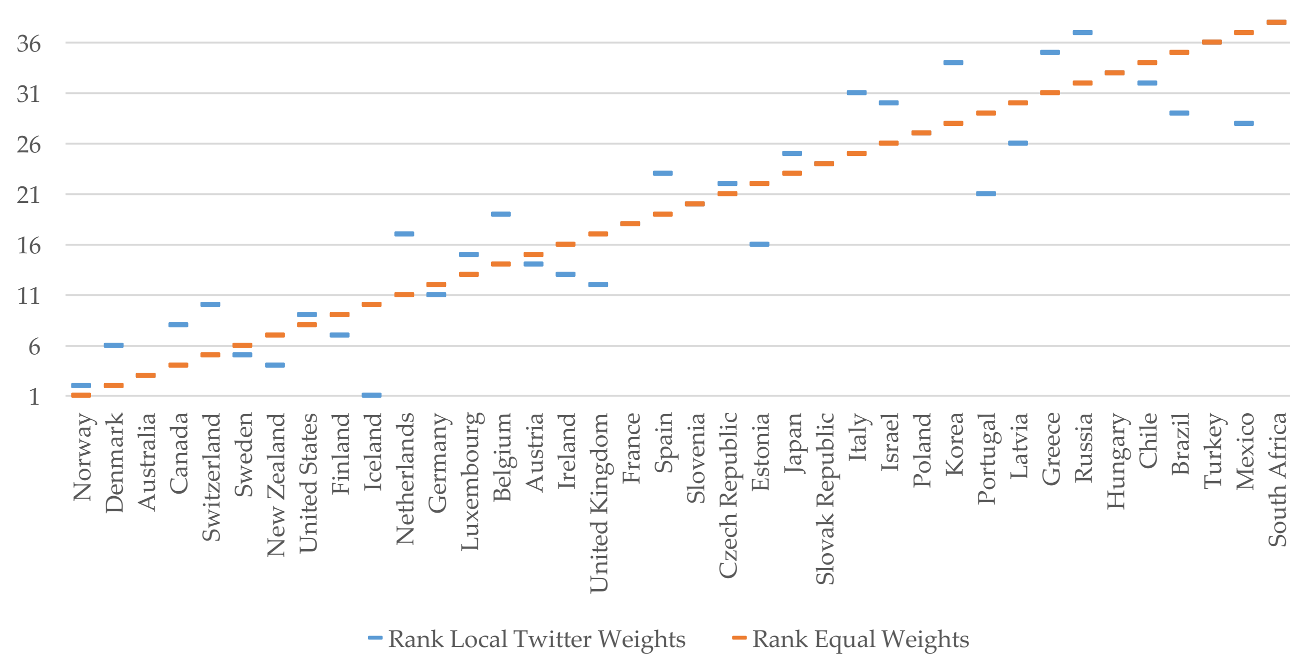


In Figure 4 we show the composite BLI using national Twitter trends as weights, and the composite BLI using equal weights. The black bar represents a gain in using Twitter frequency, whereas a white bar represents a loss. As in the case of general trends (section 4.1), there are huge differences both in the indices and in the rank that countries obtain with the different assumptions (rank correlation coefficients are in Table A.3.1 in the Appendices). Comparing the composite BLI using national Twitter trends with the composite BLI using equal weights, there emerges a huge improvement (more than 3%) in the composite BLI of Iceland, Mexico, Portugal, Norway, Estonia, Latvia, Australia, New Zealand, Brazil, United Kingdom, Turkey, Sweden, Finland, Germany, Ireland, Slovenia, Slovak Republic, Denmark, South Africa, Chile, and United States. The multi-dimensional BLI performance in these countries is in line with the order of importance given by their (Twitter community) citizens.

Five countries in Figure 4 do not show statistically significant differences between the indices (less than 1%): Poland, France, Luxembourg, Greece, and Switzerland. Relevant worsening in the composite BLI when the country-level Twitter trends are taken into account are in seven countries: Netherlands, Spain, Korea, Belgium, Israel, Italy, and Russia. The mix of well-being provided in these countries is not in line with the importance given by their (Twitter community) citizens.

In Figure 5 we show the difference in rank between composite BLI using equal weights and composite BLI using country-level Twitter frequency as weights. Comparing Figure 5 and Figure 3, it emerges that the difference between the ranks is more significant using local trends than global trends. Using local trends, Iceland, Mexico, Portugal, Brazil, Estonia, United Kingdom, Latvia, New Zealand, Ireland, Chile, Finland, Sweden, Germany, and Austria show a gain in the rank. Turkey, South Africa, Australia, France, Slovenia, Slovak Republic, Poland, and Hungary show no difference. Norway, United States, Czech Republic, Luxembourg, Japan, Denmark, Canada, Spain, Greece, Israel, Switzerland, Belgium, Russia, Netherlands, Italy, and Korea have losses in the ranking.

Figure 5 Rank in Composite BLI using local Twitter trends as Weights and using Equal Weights



Once again, these results confirm that the Twitter trends make a significant difference to the composite BLI. The directions of this differences can be used to compare the internal relative appreciation of the eleven topics of BLI, and the multi-dimensional performance of the country. When there is a gain in composite BLI, the national proportions among the performances of the eleven dimensions of the BLI are in line with the internal relative appreciation. When there is a loss, the national performance mix does not reflect the internal relative appreciation of the BLI.

5. Conclusions and discussion

This paper has proposed a composite Better Life Index by weighting the national performance in each topic according to their relative trends on Twitter. The idea behind this analysis is that the aggregate of millions of tweets submitted to Twitter may provide a representation of the priorities (the relative appreciations) among the eleven topics of the BLI. Since the relative appreciations are themselves part of multi-dimensional well-being, using them as weights can help us to design Composite Indices in a more targeted way than previous proposals.

The majority of analytics tools used in this paper are based on freely available open source NLP and social media analysis tools developed by the University of Sheffield, and thus are easily reusable by the community. The analysis has three main stages: in the first, we identify topics and sub-topics related to the eleven dimensions of the BLI; in the second, we collect tweets containing the topics and related sub-topics selected in the first stage; and in the last stage, we analyze the tweets and estimate two new Composite Indexes of Well-Being using the Twitter trends as weights.

The Twitter trends are used to build two different composite BLIs. The first is based on the general trends on the whole dataset, and reflects the relative priorities of the whole Twitter community. The second is based on the national Twitter trends, estimated using the geo-localizations associated with tweets, and reflects the relative appreciations of users located in the country.

The main results show that both the general and national Twitter trends make a difference to the composite BLI. Measuring performance taking into account Twitter trends affects both the level of Composite Indices and the rank obtained by countries. The directions of these differences can provide information about the relationship between relative performances and the relative trends in the Twitter community. When there is a gain in the composite BLI, the country mix of BLI is in line with the Twitter community's relative appreciation. When there is a loss, the country performance mix does not reflect the relative priorities given by the Twitter community. In general, relative gains when Twitter trends are taken into account are in the composite BLI of Scandinavian countries, South America, Australia, New Zealand, and Turkey. On the opposite side, Russia, Greece and Spain show losses in the composite BLI when Twitter trends are considered. These results are confirmed with both global and local trends.

The pilot study we have reported has a number of issues which could be improved in future work. First, there are language problems: the OECD website dedicated to the BLI has only seven official translations, which cover only 21 of the 37 countries in which the BLI is measured. This means that part of the OECD population may not be able to understand, and therefore cannot vote on the OECD website. Consequently, studies like this have no baseline in all the languages. An extension to this research requires translations of topics into the other languages.

Second, the size of our collection is limited, since we only collected one month of tweets, which makes our analysis sensitive to temporary trends caused by recent events in that time period. The high interest in the topic Environment, due to the timing of our collection which coincided with Trump's decision to leave the Paris agreements, is a clear signal of this. Longer collections would help smooth these temporal outliers.

Third, there is a significant number of vacancies and job offers that our model recognizes as people's interest in #jobs. Machine learning approaches could be used to classify (and exclude) such tweets, as well as an author categorization tool to separate tweets authored by people from those authored by organizations (Fernandez *et al.* 2016).

Fourth, there is only a small set of geo-located tweets in our collection. Recently, methods have been proposed for estimating the geo-location when it is not reported, mainly using follower networks (for a review see Jurgens *et al.* 2015). These could be used to expand the geo-located tweet collection.

Finally, with the 770 selected keywords and related hashtags, there are clearly missing relevant tweets. The Twitter collection should ideally be expanded with more hashtags and potentially other keywords in the body of the tweet (for example, people who talk about work-life balance do not always use that hashtag). The identification of "interest in a topic" could thus be made more complex than just counting tweets containing relevant hashtags, although we believe this is a good baseline. Despite these flaws, we believe that this analysis clearly demonstrates how Twitter trends can be used to approximate the relative appreciations and values of people, which in turn can be represented in the Composite Index.

Future research in economics should manage the multi-dimensionality of the phenomena interacting with societal behavior in a holistic approach. The UN Sustainable Development Goals (SDGs) clearly go in this direction (Costanza et al. 2016). This pilot study has shown promising results, which can be easily replicated using the framework proposed, and has demonstrated the benefit of big data as a support for these objectives.

Acknowledgements

This work was partially supported by the European Union's Horizon 2020 programme under grant agreement No. 654024 SoBigData. The primary author is grateful for the support and advice of the GATE team at the University of Sheffield, and their hosting of a SoBigData TransNational Access research visit which made this work possible.

Appendices

A.1 Selected hashtags

A. 1.1 Baseline hashtags (from words used in OECD website dedicated to BLI)

#housing; #income; #jobs; #community; #education; #environment; #civicengagement; #health; #lifesatisfaction; #safety; #worklifebalance; #vivienda; #ingresos; #empleo; #comunidad; #educación; #medioambiente; #compromisocívico; #salud; #satisfacción; #seguridad; #balancevidatrabajo; #logement; #revenu; #emploi; #lienssociaux; #éducation; #environnement; #engagementcivique; #santé; #satisfaction; #sécurité; #équilibretravailvie; #жилищныеусловия; #доход ; #работа; #общество; #образование; #экология; #гражданскиеправа; #здоровье; #удовлетворенность; #безопасность; #работаотдых; #wohnverhältnisse; #einkommen; #beschäftigung; #gemeinsinn; #bildung; #umwelt; #zivilengagement; #gesundheit; #lebenszufriedenheit; #sicherheit; #moradia; #renda; #empregos; #comunidade; #educação; #meioambiente; #engajamentocívico; #saúde; #satisfaçãopessoal; #segurança; #vidatrabalho; #abitazione; #reddito; #occupazione; #relazionisociali; #istruzione; #ambiente; #impegnocivile; #salute; #soddisfazione; #sicurezza; #equilibriolavorovita.

A. 1.2 Hashtags related to Civic Engagement

From Maynard (2017) lists:

#right; #rights; #citizenship; #elections; #party; #parliament; #policy; #president; #minister; #derecho; #derechos; #ciudadanía; #elecciones; #partido; #parlamento; #política; #presidente; #ministro; #droite; #droits; #citoyenneté; #élections; #partie; #parlement; #politique; #président; #ministre; #правый; #права; #гражданство; #выборы; #партия; #парламент; #политика; #президент; #министр; #recht; #rechte; #staatsangehörigkeit; #wahlen; #parlament; #politik; #präsident; #direito; #direitos; #cidadania; #eleições; #diritto; #diritti; #cittadinanza; #elezione; #partito; #politica.

From Word2vec estimates (Bojanowski et al 2016):

#volunteerism; #politics; #participation; #fundraising; #solidaridad; #participativo; #cívico; #civisme; #volontariat; #solidarité; #гражданской; #конституционные; #неимущественные; #sozialengagement; #friedensengagement; #zivilem; #voluntariado; #solidarismo; #cívicos; #civile; #solidarietà; #volontariato.

A. 1.3 Hashtags related to Community

From Maynard (2017) lists:

#communities; #council; #local; #religion; #social; #islam; #muslim; #romancatholics; #neighborhood; #comunidades; #consejo; #religión; #musulmán; #católicosromanos; #barrio; #communautés; #conseil; #musulman; #catholiquesromains; #quartier; #сообщества; #совет; #местный; #религия; #социальное; #ислам; #мусульманка; #римскиекатолики; #окрестности; #gemeinschaften; #rat; #lokal; #soziale; #islamismo; #römischkatholisch; #nachbarschaft; #conselho; #religião; #muçulmano; #bairro; #comunità; #consiglio; #locale; #religione; #sociale; #musulmano; #cattolici; #quartiere.

From Word2vec estimates (Bojanowski et al 2016):

#including; #incluyendo; #intégratifs; #товарищество; #pluralität; #comunitaria; #multiculturalità.

A. 1.4 Hashtags related to Education

From Maynard (2017) lists:

#literacy; #illiteracy; #schools; #teaching; #qualification; #qualifications; #teacher; #school; #pupils; #alfabetismo; #analfabetismo; #escuelas; #enseñando; #calificación; #calificaciones; #profesor; #colegio; #alumnos; #alphabétisation; #analphabétisme; #écoles; #enseignement; #prof; #école; #élèves; #грамотность; #неграмотность; #школы; #обучение; #квалификация; #квалификации; #учитель; #школа; #зрачки; #alphabetisierung; #analphabetentum; #schulen; #lehren; #qualifikation; #qualifikationen; #lehrer; #schule; #schüler ; #alfabetização; #escolas; #ensino; #qualificação; #qualificações; #professor; #escola; #alunos; #alfabetizzazione; #scuole; #insegnamento; #qualificazione; #titolidistudio; #insegnante; #scuola; #alunni.

A. 1.5 Hashtags related to Environment

From Maynard (2017) lists:

#landfill; #green; #sustainability; #nature; #natural; #power; #energy; #forest; #sea; #vertedero; #verde; #sostenibilidad; #naturaleza; #poder; #energía; #bosque; #mar; #décharge; #vert; #durabilité; #lanature;

#naturel; #puissance; #énergie; #forêt; #mer; #полигон; #зеленый; #устойчивость; #природа; #натуральный; #мощность; #энергия; #лес; #море; #deponie; #grün; #nachhaltigkeit; #natur; #natürlich; #leistung; #energie; #wald; #meer; #aterro; #sustentabilidade; #natureza; #energia; #floresta; #discarica; #sostenibilità; #natura; #naturale; #foresta; #mare.

From Word2vec estimates (Bojanowski et al 2016):

#climatechange; #ecosystems; #ecology; #cambioclimático; #ecológico; #ecosistema; #changementclimatique; #écosystème; #écologie; #изменениеклимата; #экологический; #экосистема; #klimawandel; #ökosystemgesundheit; #lebensumwelt; #alteraçõesclimáticas; #ecologias; #cambiamentoclimatico; #ecologico.

A. 1.6 Hashtags related to Health

From Maynard (2017) lists:

#doctor; #doctors; #healthcare; #nurse; #nurses; #ambulance; #hospital; #emergency; #lifeexpectancy; #doctores; #sanitario; #enfermera; #enfermeras; #ambulancia; #emergencia; #esperanzadevida; #docteur; #médecins; #soinsdesanté; #infirmière; #infirmières; #hôpital; #urgence; #espérancedevie; #врач; #врачи; #здравоохранение; #медсестра; #медсестры; #скораяпомощь; #больница; #крайняянеобходимость; #продолжительностьжизни; #arzt; #ärzte; #gesundheitswesen; #krankenschwester; #krankenschwestern; #krankenwagen; #krankenhaus; #notfall; #lebenserwartung; #médico; #médicos; #cuidadosdesaúde; #enfermeira; #enfermeiras; #ambulância; #emergência; #expectativadevida; #medico; #medici; #sanità; #infermiera; #infermieri; #ambulanza; #ospedale; #emergenza; #speranzadivita.

From Word2vec estimates (Bojanowski et al 2016):

#illness; #enfermedad; #maladie; #болезнь; #krankheit; #doença; #malattia.

A. 1.7 Hashtags related to Housing

From Maynard (2017) lists:

#house; #home; #landlord; #landlords; #rent; #mortgage; #homebuilding; #socialhousing; #casa; #hogar; #dueño; #propietarios; #alquilar; #hipoteca; #construccióndeviviendas; #viviendasocial; #alojamiento; #maison; #domicile; #propriétaire; #propriétairesfonciers; #location; #hypothèque; #constructiondemaisons; #logementsocial; #дом; #главная; #арендодатель; #помещики; #аренда; #ипотека; #строительстводома; #социальногожилья; #корпус; #haus; #zuhause; #vermieter; #miete; #hypothek; #hausbau; #sozialwohnungen; #gehäuse; #lar ; #senhorio; #senhorios; #aluguel; #construçãodecasas; #habitaçãosocial; #habitação; #proprietario; #padronidicasa; #affitto; #mutuo; #costruttore; #alloggisociali; #alloggio

A. 1.8 Hashtags related to Income

From Maynard (2017) lists:

#austerity; #budget; #debt; #debts; #economic; #economy; #prices; #wealth; #spending; #austeridad; #presupuesto; #deuda; #deudas; #económico; #economía; #precios; #riqueza; #gasto; #austérité; #dette; #dettes; #économique; #économie; #desprix; #richesse; #dépenses; #строгость; #бюджет; #долг; #долги; #экономической; #экономика; #цены; #богатство; #расходы; #strenge; #schuld; #schulden; #wirtschaftlich; #wirtschaft; #preise; #reichtum; #ausgaben; #austeridade; #orçamento; #dívida; #dívidas; #econômico; #economia; #preços; #gastos; #austerità; #bilancio; #debito; #debiti; #economico; #prezzi; #ricchezza; #spesa.

From Word2vec estimates (Bojanowski et al 2016):

#poverty; #earners; #salarios; #subsidios; #pauvreté; #salariés; #бедность; #добытчики; #armut; #verdiener; #pobreza; #assalariados; #povertà; #percettoridireddito.

A. 1.9 Hashtags related to Jobs

From Maynard (2017) lists:

#unemployment; #apprenticeship; #wage; #wages; #unemployed; #employment; #employees; #work; #worker; #desempleo; #aprendizaje; #salario; #desempleados; #empleados; #trabajo; #obrero; #chômage; #apprentissage; #salaire; #lessalaires; #sansemploi; #employés; #travail; #ouvrier; #безработица; #ученичество; #заработнаяплата; #безработные; #занятость; #сотрудников; #работник; #arbeitslosigkeit; #lehre; #führen; #lohn; #arbeitslos; #angestellte; #arbeit; #arbeitnehmer; #desemprego; #aprendizagem; #salário; #salários; #desempregado; #emprego; #empregados; #trabalhos; #trabalhador; #disoccupazione; #apprendistato; #salari; #disoccupato; #dipendenti; #lavoro; #lavoratore.

From Word2vec estimates (Bojanowski et al 2016):

#layoffs; #despidos; #entlassungen; #увольнения; #demissões; #licenziamenti.

A. 1.10 Hashtags related to Life Satisfaction

From Word2vec estimates (Bojanowski et al 2016):

#happiness; #enjoy; #glee; #joy; #prosperity; #pleasure; #felicidad; #disfrutar; #alegría; #prosperidad; #placer; #bonheur; #prendreplaisir; #joie; #prospérité; #plaisir; #приподнятоенастроение; #наслаждение; #спокойствиедуха; #удовольствие; #процветание; #glück; #genießen; #freude; #wohlstand; #vergnügen; #felicidade; #apreciar; #alegria; #prosperidade; #prazer; #felicità; #godere; #allegria; #gioia; #prosperità; #piacere; #satisfactions; #dissatisfaction; #unsatisfaction; #pleasures; #friendship; #satisfacciones; #insatisfacción; #insatisfactions; #placeres; #desamparo; #insatisfaction; #insatisfacciones; #amitié; #сатисфакции; #неудовлетворенность; #неудовлетворенности; #удовольствиядуха; #дружба; #befriedigungen; #unzufriedenheit; #unzufriedene; #freuden; #freunde; #satisfações; #insatisfação; #insatisfações; #prazeres; #amigos; #soddisfazioni; #insoddisfazione; #insoddisfazioni; #piaceri; #amicizia

A. 1.11 Hashtags related to Safety

From Maynard (2017) lists:

#crime; #crimes; #prison; #prisons; #drugs; #police; #terrorist; #terrorism; #policemen; #crimen; #crímenes; #prisión; #cárceles; #drogas; #policía; #terrorista; #terrorismo; #policías; #criminalité; #drogues; #terroriste; #terrorisme; #policiers; #преступление; #преступления; #тюрьма; #тюрьмы; #наркотики; #полиция; #террорист; #терроризм; #полицейские; #kriminalität; #verbrechen; #gefängnis; #gefängnisse; #drogen; #polizei; #terrorismus ; #polizisten; #prisão; #prisões; #polícia; #policiais; #crimine; #crimini; #prigione; #carceri; #droghe; #polizia; #poliziotti; #security.

A. 1.12 Hashtags related to work Life Balance

From Word2vec estimates (Bojanowski et al 2016):

#freetime; #holiday; #holidays; #fun; #hobby; #dayoff; #weekend; #family; #rest; #ocio; #fiesta; #vacaciones; #diversión; #díadedescanso; #findesemana; #familia; #resto; #loisir; #vacances; #fêtes; #amusement; #jourdecongé; #famille; #repos; #досуг; #праздник; #каникулы; #веселье; #хобби; #выходнойдень; #уик-энд; #семья; #отдых; #freizeit; #urlaub; #ferien; #spaß; #ruhetag; #wochenende; #familie; #lazer; #feriado; #feriados; #diversão; #passatempo; #diadefolga; #fimdesemana; #família; #descanso; #tempolibero; #vacanza; #ferie; #svago; #giornolibero; #finesettimana; #famiglia; #riposo; #lifetime; #vidas; #duréedevie; #lebenszeit; #tempodevida; #tuttalavita

A.2 Preferences on the BLI dedicated website

Table A.2.1 Relative Appreciations of Better Life Index on OECD dedicated website

	Average	Median
Health	3.73	4.00
Life Satisfaction	3.69	4.00
Education	3.54	4.00
Work-Life Balance	3.35	4.00
Safety	3.32	4.00
Environment	3.26	3.00
Jobs	3.19	3.00
Housing	3.17	3.00
Income	3.10	3.00
Community	2.90	3.00
Civic Engagement	2.41	2.00

Author's elaboration on Data downloaded on 23 July 2017

A.3 Correlations among composite BLI

Table A3.1 Rank Correlations (95 % bootstrap upper and lower bounds)

	C BLI global trends	C BLI local trends	C BLI equal weights
C BLI global trends	1.000		
LB	0.898		
C BLI local trends	0.958	1.000	

UB	0.975		
LB	0.930	0.851	
C BLI equal weights	0.973	0.927	1.000
UB	0.985	0.953	

Note: LU= lower bound; Ub=Upper Boung; Bootstrap with 1000 replicates using R package by Herve´ (2015)

A.4 Variables and Topics in Better Life Index

Table A4.1 Dimensions and related variables of the BLI

Topics	Related variables
Housing	Dwellings without basic facilities Housing expenditure Rooms per person
Income	Household net adjusted disposable income Household net financial wealth
Jobs	Employment rate Job security Long-term unemployment rate Personal earnings
Community	Quality of support network
Education	Educational attainment Student skills Years in education
Environment	Air pollution Water quality
Civic engagement	Consultation on rule-making Voter turnout
Health	Life expectancy Self-reported health
Life Satisfaction	Life satisfaction

Safety	Assault rate Homicide rate
Work-Life Balance	Employees working very long hours Time devoted to leisure and personal care

References

A. Alesina, P. Giuliano, Preferences for redistribution No. w14825, National Bureau of Economic Research (2009).

B. Bleys, Beyond GDP: Classifying alternative measures for progress, Social Indicators Research 109.3 (2012): 355 376.

J. Blumenstock, G. Cadamuro, R. On, Predicting poverty and wealth from mobile phone metadata, Science 350.6264 (2015): 1073 1076.

P. Bojanowski, E. Grave, A. Joulin, T. Mikolov, Enriching word vectors with subword information, arXiv preprint arXiv:1607.04606 (2016).

J. Bollen, H. Mao, X. Zeng, Twitter mood predicts the stock market, Journal of computational science 2.1 (2011): 1 8.

R. M. Bond, C. J. Fariss, J. J. Jones, A. D. Kramer, C. Marlow, J. E. Settle, J. H. Fowler, A 61-million-person experiment in social influence and political mobilization, Nature, 489(7415) (2012): 295 298.

D. Boyd, K, Crawford, Critical questions for big data: Provocations for a cultural, technological, and scholarly phenomenon, Information, communication & society 15.5 (2012): 662 679.

A. Cavallo, Online and official price indexes: Measuring Argentina's inflation, Journal of Monetary Economics, 60(2) (2013): 152-165.

A. Cavallo, Are online and offline prices similar? evidence from large multi-channel retailers, American Economic Review, 107(1) (2017): 283-303.

A. Cavallo, R. Rigobon, The billion prices project: Using online prices for measurement and research, Journal of Economic Perspectives, 30(2) (2016): 151-78.

C. C. Chang, K. H. Chu, A recommender system combining social networks for tourist attractions, Computational Intelligence, Communication Systems and Networks (CICSyN), 2013 Fifth International Conference on. IEEE, (2013): 42 47.

L. Cherchye, W. Moesen, N. Rogge, T. Van Puyenbroeck, An introduction to ‘benefit of the doubt’ composite indicators, Social Indicators Research, 82(1) (2007): 111 145.

H. Choi, H Varian, Predicting the present with Google Trends, *Economic Record*, 88(s1) (2012): 2-9.

R. Costanza, L. Daly, L. Fioramonti, E. Giovannini, I. Kubiszewski, L. F. Mortensen, ..., R. Wilkinson, Modelling and measuring sustainable wellbeing in connection with the UN Sustainable Development Goals, Ecological Economics, 130 (2016): 350 355.

D. Coyle, GDP: A brief but affectionate history, Princeton University Press, Princeton 2014.

H. Cunningham, D. Maynard, K. Bontcheva, V. Tablan, GATE: A Framework and Graphical Development Environment for Robust NLP Tools and Applications, Proceedings of the 40th Anniversary Meeting of the Association for Computational Linguistics (ACL'02), Philadelphia, July 2002.

C. De Beukelaer, Gross Domestic Problem: The Politics Behind the World's Most Powerful Number, Journal of Human Development and Capabilities 15.2-3 (2014): 290 291.

E. di Bella, L. Leporatti, F. Maggino, Big Data and Social Indicators: Actual Trends and New Perspectives, Social Indicators Research, (2016): 1 10.

A. Drazen, Political economy in macroeconomics, Orient Blackswan, 2004.

R. Dutta, A. Morshed, J. Aryal, C. D'este, A. Das, Development of an intelligent environmental knowledge system for sustainable agricultural decision support, Environmental Modelling & Software, 52 (2014): 264 272.

R. A. Easterlin, Does economic growth improve the human lot? Some empirical evidence, Nations and households in economic growth, 89 (1974): 89 125.

L. Einav, D. Knoepfle, J. Levin, N. Sundaresan, Sales taxes and internet commerce, *American Economic Review*, 104(1) (2014): 1-26.

L. Einav, T. Kuchler, J. D. Levin, N. Sundaresan, Learning from seller experiments in online markets, *NBER Working Paper* no. 17385, Cambridge: National Bureau of Economic Research (2013).

L. Einav, J. Levin, The data revolution and economic analysis, *Innovation Policy and the Economy*, 14(1) (2014): 1-24.

C. D. Elvidge, P. C. Sutton, T. Ghosh, B. T. Tuttle, K. E. Baugh, B. Bhaduri, E. Bright, A global poverty map derived from satellite data, Computers & Geosciences, 35(8) (2009): 1652 1660.

Ethnologue, Languages of the World, Summary by language size (accessed 22 June 2017**), link: https://www.ethnologue.com/statistics/size**

Eurobarometer, Standard, Public opinion in the European Union, (2012).

M. Fernandez, H. Alani, L. Piccolo, C. Meili, D. Maynard, M. Wippoo, Talking Climate Change via Social Media: Communication, Engagement and Behaviour, Proc. of WebSci, May 22-25 2016, Hannover, Germany.

L. Fioramonti, Gross domestic problem: The politics behind the world's most powerful number, Zed Books, 2013.

M. Fleurbaey, Beyond GDP: The quest for a measure of social welfare, Journal of Economic literature 47.4 (2009): 1029 1075.

M. Floridi, S. Pagni, S. Falorni, T. Luzzati, An exercise in composite indicators construction: Assessing the sustainability of Italian regions, Ecological economics, 70(8) (2011): 1440 1447.

B. S. Frey, A. Stutzer, Happiness and public choice, Public Choice, 144(3-4) (2010): 557 573.

M. Garcia-Herranz, E. Moro, M. Cebrian, N. A. Christakis, J. H. Fowler, Using friends as sensors to detect global-scale contagious outbreaks, PloS one, 9(4) (2014): e92413.

D. Giannone, L. Reichlin, D. Small, Nowcasting: The real-time informational content of macroeconomic data, Journal of Monetary Economics, 55(4) (2008): 665 676.

J. Ginsberg, M. H. Mohebbi, R. S. Patel, L. Brammer, M. S. Smolinski, L. Brilliant, Detecting influenza epidemics using search engine query data, Nature, 457(7232) (2009): 1012 1014.

S. Greco A. Ishizaka B. Matarazzo G. Torrisi, Stochastic Multiattribute Acceptability Analysis: an application to the ranking of Italian regions, (forthcoming) Regional Studies (2017).

J. F. Helliwell, How's life? Combining individual and national variables to explain subjective well-being, Economic Modelling, 20(2) (2003) 331 360.

J. F. Helliwell, C. P. Barrington-Leigh, Viewpoint: Measuring and understanding subjective well-being, Canadian Journal of Economics/Revue canadienned'économique, 43(3) (2010): 729 753.

M. Herve´, RVAideMemoire: Diverse basic statistical and graphical functions, R package version0.9-52 (2015) http://CRAN.R-project.org/package=RVAideMemoire.

L. Hong, G. Convertino, H. E. Chi, Language Matters In Twitter: A Large Scale Study, ICWSM (2011).

D. Jurgens, T. Finethy, J. McCorriston, Y. T. Xu, D. Ruths, Geolocation Prediction in Twitter Using Social Networks: A Critical Analysis and Review of Current Practice, ICWSM, 15 (2015): 188 197.

Z. Karabell, The leading indicators: a short history of the numbers that rule our world, Simon and Schuster 2015.

Á. Kerényi, The better life index of the organization for economic co-operation and development, *Public Finance Quarterly*, 56(4) (2011): 518–538.

S. Kutznets, National Income, 1929-1932, NBER, 1934: 1 12.

V. Lampos, Detecting events and patterns in large-scale user generated textual streams with statistical learning methods, arXiv preprint arXiv 1208.2873 (2012).

D. Laney, 3D data management: Controlling data volume, velocity and variety, META Group Research Note, 6, 70 (2001).

T. Lansdall-Welfare, V. Lampos, N. Cristianini, Effects of the Recession on Public Mood in the UK, Proceedings of the 21st International Conference on World Wide Web ACM N. (2012): 1221 1226.

M. Lenormand, M. Picornell, O. G. Cantu-Ros, A. Tugores, T. Louail, R. Herranz, ..., J. J. Ramasco, Cross-checking different sources of mobility information, PLoS One, 9(8) (2014) e105184.

N. C. Lind, Better life index, *Encyclopedia of quality of life and well-being research*, Springer, Netherlands, (2014): 381–382.

J. Lorenz, C. Brauer, D. Lorenz, Rank-Optimal Weighting or "How to be Best in the OECD Better Life Index?", Social Indicators Research (2016): 1 18.

P. Manrique, H. Qi, A. Morgenstern, N. Velásquez, T. C. Lu, N. Johnson, Context matters: improving the uses of big data for forecasting civil unrest: emerging phenomena and big data, Intelligence and Security Informatics (ISI), 2013 IEEE International Conference on IEEE (2013): 169 172.

H. Mao, X. Shuai, Y. Y. Ahn, J. Bollen, Quantifying socio-economic indicators in developing countries from mobile phone communication data: applications to Côte d'Ivoire, EPJ Data Science, 4(1), (2015): 4 15.

M. Marković, S. Zdravković, M. Mitrović, A. Radojičić, An Iterative Multivariate Post Hoc I-Distance Approach in Evaluating OECD Better Life Index, *Social Indicators Research*, 126.1 A. (2016): 1 19.

N. Marks, S. Abdallah, A. Simms, S. Thompson, The (un) Happy Planet Index: An index of human well-being and ecological impact, *New Economics Foundation, London* (2006).

D. Maynard, M. A. Greenwood, Who cares about sarcastic tweets? Investigating the Impact of Sarcasm on Sentiment Analysis, *LREC* (2014): 4238 4243.

D. Maynard, I. Roberts, M. A. Greenwood, D. Rout, K. Bontcheva, A framework for real-time semantic social media analysis, *Web Semantics: Science, Services and Agents on the World Wide Web* (2017).

A. McAfee, E. Brynjolfsson, T. H. Davenport, Big data: the management revolution, *Harvard business review*, 90(10) (2012): 60-68.

T. Mikolov, K. Chen, G. Corrado, J. Dean, Efficient estimation of word representations in vector space, *arXiv preprint arXiv*:1301.3781 (2013).

H. Mizobuchi, Measuring World Better Life Frontier: A composite Indicator for OECD Better Life Index, *Social Indicators Research*, 118.3 (2014): 987 1007.

A. Monroy-Hernández, E. Kiciman, M. De Choudhury, S. Counts, The new war correspondents: The rise of civic media curation in urban warfare, *Proceedings of the 2013 conference on Computer supported cooperative work* ACM (2013): 1443 1452.

M. Nardo, M. Saisana, A. Saltelli, S. Tarantola, A. Hoffman, E. Giovannini, *Handbook on constructing composite indicators*. Paris: OECD-JRC, 2008.

OECD *How's Life? 2015 - Measuring Well-being*, Paris, 2015.

V. Patrizii, A. Pettini, G. Resce, The Cost of Well-Being, *Social Indicators Research* (2016): 1 26.

I. Permanyer Assessing the robustness of composite indices rankings, *Review of Income and Wealth*, 57(2) (2011): 306 326.

T. Persson, G. E. Tabellini, Political economics: explaining economic policy, *MIT press,* 2002.

S. Rill, D. Reinel, J. Scheidt, R. V. Zicari, Politwi: Early detection of emerging political topics on twitter and the impact on concept-level sentiment analysis, *Knowledge-Based Systems*, 69 (2014): 24 33.

M. Saisana, S. Tarantola, A. Saltelli, Uncertainty and sensitivity techniques as tools for the analysis and validation of composite indicators, *Journal of the Royal Statistical Society A*, 168(2) (2005): 307 323.

S. Seabold, A. Rutherford, O. De Backer, A. Coppola, The pulse of public opinion: using Twitter data to analyze public perception of reform in El Salvador, *Policy Research Working Paper 7399*. World Bank (2015).

A. Sharpe, Literature Review of Frameworks for Macro-indicators, *Centre for the Study of Living Standards*, Ottawa, CAN (2004).

SimilarWeb, *Top Websites Ranking, Top sites ranking for all categories in the world* (accessed 15 June 2018) link: https://www.similarweb.com/top-websites.

B. Smith-Clarke, A. Mashhadi, L. Capra, Poverty on the cheap: Estimating poverty maps using aggregated mobile communication networks, *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* ACM (2014): 511-520.

Statista, *Most famous social network sites worldwide as of April 2018, ranked by number of active users* (accessed 15 June 2018) link: https://www.statista.com/statistics/272014/global-social-networks-ranked-by-number-of-users.

J. E. Stiglitz, A. Sen, J. P. Fitoussi, Report by the commission on the measurement of economic performance and social progress, *Paris: Commission on the Measurement of Economic Performance and Social Progress* (2010).

M. A. Stoové, A. E. Pedrana, Making the most of a brave new world: Opportunities and considerations for using Twitter as a public health monitoring tool, *Preventive medicine*, 63 (2014): 109 111.

C. M. Tiebout, A pure theory of local expenditures, *Journal of political economy*, 64(5) (1956): 416 424.

UNDP *Human Development Report*, United Nation Development Programme, Nairobi, Kenia (1996).

UNICEF, Tracking anti-vaccination sentiment in eastern European social media networks, *Working paper*, UNICEF (2013).

S. Usherwood, K. A. Wright, Sticks and stones: Comparing Twitter campaigning strategies in the European Union referendum, *The British Journal of Politics and International Relations*, 19(2) (2017): 371 388.

M. Yazdani, L. Manovich, Predicting social trends from non-photographic images on Twitter. In *Big Data (Big Data), 2015 IEEE International Conference on* IEEE, (2015): 1653 1660.

E. Zagheni, V. R. K. Garimella, I. Weber, Inferring international and internal migration patterns from twitter data, *Proceedings of the 23rd International Conference on World Wide Web* ACM (2014): 439 444.