Understanding Societal Well-Being Through the Eyes of the News World Media

Vasiliki Voukelatou
Scuola Normale Superiore and
Institute of Information Science and
Technologies (ISTI)-CNR
56126-56127, Pisa, Italy
Email: vasiliki.voukelatou@sns.it

Luca Pappalardo
Institute of Information Science and
Technologies (ISTI)-CNR
56127, Pisa, Italy
Email: luca.pappalardo@isti.cnr.it

Fosca Giannotti
Institute of Information Science and
Technologies (ISTI)-CNR
56127, Pisa, Italy
Email: fosca.giannotti@isti.cnr.it

Abstract—**Societal well-being is an important value for people's lives and it contributes considerably to the societal progress. It has been traditionally captured with surveys and during the last decades innovative approaches have been applied for its measurement. Global Peace Index is an indicator, which measures well-being in terms of peace and safety. This study suggests the potential measurement of this index through a novel automatic methodology.**

Keywords–GDELT; Well-being; Global Peace Index; Big Data

I. INTRODUCTION

Defining well-being has been always considered a challenge. Therefore, researchers have expressed its nature by focusing on the dimensions of well-being, rather than on definition[1]. In fact, well-being encompasses a set of health, socio-economic (such as unemployment) and political dimensions (such as peace and safety) [2], [3], [4]. Therefore, monitoring it is one of the main concerns of policy-makers. In fact, when a new policy is applied or an unexpected event occurs, what policy-makers focus on is eventual consequences on humans' well-being.

Researchers have traditionally considered Gross Domestic Product (GDP) as a good indicator of well-being in society. The reason for which it has been considered as a suitable indicator for measuring well-being is that it is strongly linked with the standard of living indicators [5]. However, GDP has been criticized as a weak indicator of well-being and therefore a misleading tool for public policies [6]. In fact, Stiglitz Commission [7] in 2009 observed that there could be used other statistical tools, complementary to GDP, for the well-being measurement. Following this direction, researchers have created various indexes, for the measurement of well-being, for many purposes and for capturing a variety of its dimensions [6]. Some important examples are the Human Development Index (HDI), created by the United Nations Development Programme (UNDP) [8], the Better Life Index (BLI), created by the Organisation for Economic Co-operation and Development (OECD) [4] and the Sustainable Well-Being Index (Benessere Equo Sostenibile-BES), created by Italian National Statistics Istitute (ISTAT) [9].

Traditionally, well-being and territorial socio-economic development are measured through surveys of household income and consumption [10]. Nevertheless, surveys are usually very costly, making it difficult for many developing countries to update their estimates frequently. During the last decades, with the proliferation of technology, researchers are inclined to use more innovative and cost-effective approaches complementing the traditional measurement of well-being. In fact, over the last years, researchers have frequently used Big Data sources, which seem to offer new opportunities to study the well-being dimensions and to circumvent the limitations carried from traditional methodologies. For this research purpose, several data sources have been used, with the most important ones being Twitter (see e.g. [11], [12]), Call Detail Record data (CDRs) (see e.g. [13], [14], GPS and transportation data (see e.g. [15], [16], [17]), and a variety of approaches have been applied, such as sentiment analysis, face recognition, network analysis and others.

II. METHODOLOGY AND RESULTS

For the purposes of the current study, nowcasting well-being is the main task to be realized. In particular, well-being is explored in terms of safety, which is one of the well-being dimensions, as defined by OECD [4], and in terms of peace, which is one of the sustainable development goals, as defined by United Nations [18]. *Global Peace Index (GPI)* [19], created by the Institute for Economics and Peace, captures the peacefulness of continents and ranks 163 independent states and territories . GPI is traditionally measured by institutional surveys and governmental data. Therefore, an innovative data source is suggested, which could capture the GPI score automatically, complementing the traditional methodology.

GDELT[20], yet a barely explored data source, is validated whether it could satisfy the aforementioned needs. It is a publicly available event database supported by Google Jigsaw. contains data based on international English-language news, such as AfricaNews, Agence France Presse, Associated Press, Associated Press Online, Associated Press Worldstream, BBC Monitoring, Christian Science Monitor, Facts on File, Foreign Broadcast Information Service, The New York Times, United Press International and The Washington Post etc. In particular, Tabari system extracts the events from each article and stores

them in an expanded version of the dyadic CAMEO format, a conflict and mediation event taxonomy [21]. Examples of identified events are protests, conflicts, peace appeals, terrorist attacks, violence, etc. Additional software identifies the location of each event, with a similar approach used to map wikipedia [22], and the tone, using the tonal algorithm from Shook et al. [23]). Multiple references to the same event across one or more articles from the same newswire are collapsed into a single event record, but are not deduplicated across newswires. Data are updated daily and historical data, since 1979, are also provided (see Leetaru et al. [24] for more details on GDELT).

the study presented here, the measurement of GPI is by the creation of new variables extracted from 0 GDELT event . In particular, official GPI variables are recreated by mapping them with GDELT data, which provide similar information. For instance, "Number of jailed population per 100,000 people" official GPI variable is recreated by "Arrest, detain; legal or extrajudicial arrests, detentions, or imprisonments" and "Threaten with repression" GDELT event categories and is simply called "jailed" for convenience. After a careful mapping, 9 new variables are extracted from the GDELT event database, as a count of events associated with at a country and year level, normalized to the total number of events, at a country and year level.

order to evaluate the new variables on their relationship with the GPI official score, correlation analysis is conducted. Preliminary analysis is done without distinguishing between countries and years. Results show noticeable correlations between the created GDELT variables and the official GPI score. particular, the simply called "conventional weapons" variable and GPI, as well as "impact of terrorism" variable and GPI show Pearson's correlation coefficients r=0.41 and r=0.35 respectively.

III. DISCUSSION AND FUTURE STEPS

The next step of the study is the creation of odel, that given the variables created, will nowcast/predict the official GPI score a country and year level. The model will be built with the new variables and, potentially, with the Scale ranking data provided by GDELT, capturing the potential impact each event might have on the stability of a country. model is expected to provide a GPI trend pattern similar to the one provided by the official model. Such a result would contribute to the official GPI yearly study, since GDELT data are updated daily, making it possible to provide policy researchers with GPI variations throughout the year. In addition, in case of a new instability in a country, such a model could provide predictions for the variations of the GPI score of the country for the upcoming year.

REFERENCES

[1] R. Dodge, A. P. Daly, J. Huyton, and L. D. Sanders, "The challenge of defining wellbeing," International journal of wellbeing, vol. 2, no. 3, 2012.

[2] B. Rapporto, "Il benessere equo e sostenibile in italia," 2015.

[3] W. H. Organization, World health statistics 2016: monitoring health for the SDGs sustainable development goals.   World Health Organization, 2016.

[4] O. for Economic Co-operation and Development, How's life?: measuring well-being.   Oecd Paris, 2011.

[5] C. M. Reinhart and V. R. Reinhart, "After the fall," National Bureau of Economic Research, Tech. Rep., 2010.

[6] M. Fleurbaey, "Beyond gdp: The quest for a measure of social welfare," Journal of Economic literature, vol. 47, no. 4, 2009, pp. 1029–75.

[7] J. E. Stiglitz, A. Sen, and J.-P. Fitoussi, "Report by the commission on the measurement of economic performance and social progress," 2017.

[8] S. Anand and A. Sen, "Human development index: Methodology and measurement." 1994.

[9] C. Istat, "Il benessere equo e sostenibile in italia," Roma, Istat, 2013.

[10] A. Deaton, The analysis of household surveys: a microeconometric approach to development policy.   The World Bank, 1997.

[11] J. C. Eichstaedt, H. A. Schwartz, M. L. Kern, G. Park, D. R. Labarthe, R. M. Merchant, S. Jha, M. Agrawal, L. A. Dziurzynski, M. Sap et al., "Psychological language on twitter predicts county-level heart disease mortality," Psychological science, vol. 26, no. 2, 2015, pp. 159–169.

[12] L. Dong, S. Chen, Y. Cheng, Z. Wu, C. Li, and H. Wu, "Measuring economic activity in china with mobile big data," EPJ Data Science, vol. 6, no. 1, 2017, p. 29.

[13] P. Sundsøy, J. Bjelland, B.-A. Reme, E. Jahani, E. Wetter, and L. Bengtsson, "Towards real-time prediction of unemployment and profession," in International Conference on Social Informatics.   Springer, 2017, pp. 14–23.

[14] F. Finger, T. Genolet, L. Mari, G. C. de Magny, N. M. Manga, A. Rinaldo, and E. Bertuzzo, "Mobile phone data highlights the role of mass gatherings in the spreading of cholera outbreaks," Proceedings of the National Academy of Sciences, vol. 113, no. 23, 2016, pp. 6421–6426.

[15] C. Smith, D. Quercia, and L. Capra, "Finger on the pulse: identifying deprivation using transit flow analysis," in Proceedings of the 2013 conference on Computer supported cooperative work.   ACM, 2013, pp. 683–692.

[16] A. I. Robinson, F. Carnes, and N. M. Oreskovic, "Spatial analysis of crime incidence and adolescent physical activity," Preventive medicine, vol. 85, 2016, pp. 74–77.

[17] B. Ariel and H. Partridge, "Predictable policing: Measuring the crime control benefits of hotspots policing at bus stops," Journal of Quantitative Criminology, vol. 33, no. 4, 2017, pp. 809–833.

[18] G. Assembly, "Sustainable development goals," SDGs), Transforming our world: the, vol. 2030, 2015.

[19] The Institute for Economics and Peace, "VISION of HUMANITY," http://visionofhumanity.org//, 2017, (Online; accessed 05-June-2010).

[20] Kalev Leetaru, "The GDELT Project," https://www.gdeltproject.org/, 2013, (Online; accessed 05-June-2010).

[21] D. J. Gerner, P. A. Schrodt, O. Yilmaz, and R. Abu-Jabr, "Conflict and mediation event observations (cameo): A new event data framework for the analysis of foreign policy interactions," International Studies Association, New Orleans, 2002.

[22] D.-L. Magazine, "Fulltext geocoding versus spatial metadata for large text archives: Towards a geographically enriched wikipedia," D-Lib Magazine, vol. 18, no. 9/10, 2012.

[23] E. Shook, K. Leetaru, G. Cao, A. Padmanabhan, and S. Wang, "Happy or not: Generating topic-based emotional heatmaps for culturomics using cybergis," in 2012 IEEE 8th International Conference on E-Science.   IEEE, 2012, pp. 1–6.

[24] K. Leetaru and P. A. Schrodt, "Gdelt: Global data on events, location, and tone, 1979–2012," in ISA annual convention, vol. 2, no. 4.   Citeseer, 2013, pp. 1–49.