

SHORT TERM SCIENTIFIC MISSION (STSM) SCIENTIFIC REPORT

This report is submitted for approval by the STSM applicant to the STSM coordinator

Action number: CA18209 - European network for Web-centred linguistic data science

STSM title: Circular Economy Indicators Alignment Framework

STSM start and end date: 04/03/2020 to 13/03/2020

Grantee name: Konstantinos P. Tsagarakis

PURPOSE OF THE STSM:

(max.200 words)

This STSM aimed at the collaboration between the grantee, Konstantinos P. Tsagarakis, Democritus University of Thrace, Greece with the host, Prof. Dimitar Trajanov, Ss. Cyril and Methodius University, North Macedonia. According to the work plan, the aim was to consider several indicators developed by international organizations and examine their relationships with circular economy indicators. Towards this direction, the following two were combined: a) the Host's expertise deriving from similar work on UN Global Sustainable Development Goal Indicators with other sets of indicators from various domains and b) the grantee's expertise in circular economy, making this collaboration of mutual benefit. This STSM falls under Task 4.4., according to the MoU, i.e., "Use cases in life sciences: With the use of text analytics techniques currently applied in biomedicine and other life sciences can be expanded with the outcomes of WG3 in order to fully exploit structured linguistic information in the process of discovering hidden facts out of textual data."

DESCRIPTION OF WORK CARRIED OUT DURING THE STSMS

(max.500 words)

During the STSM, the grantee had the opportunity to meet and interact, apart from Prof. Dimitar Trajanov, with researchers from the host University including: Aleksandar Petrushev, Andrej Milovanovikj, Boris Tancovski, and Kostadin Mishev. Quantitative data about circular economy and other indicators had been retrieved and treated before analyzing. The main data sources include:

<http://hdr.undp.org/en/content/human-development-index-hdi>

<http://info.worldbank.org/governance/wgi/>

https://appsso.eurostat.ec.europa.eu/nui/show.do?dataset=env_ac_cur&lang=en

<https://databank.worldbank.org/source/world-development-indicators>

https://ec.europa.eu/eurostat/databrowser/view/cei_wm011/default/table?lang=en

<https://www.doingbusiness.org/en/data/doing-business-score?topic=>

<https://www.heritage.org/index/explore>

The analysis had two main directions; one quantitative (based on the indicators' values), and one qualitative (based on textual metadata). For both directions, we investigated correlations and causality in order to advance pre-established statistical panel data analysis with same sources of data. Data consisted of annual values of selected indicators at country (macro) level. The data analysis focused on EU

countries, as they comply to the same environmental and economic policies. The python "matplotlib" and "DoWhy" libraries for similarity and causality analysis were used.

DESCRIPTION OF THE MAIN RESULTS OBTAINED

500 word summary here

For the quantitative data analysis indicators were classified to groups to explain circularity indicators. The groups are related to Development, Corruption, Freedom, Size, Inequality, Energy, Government Quality, and Business. Analysis involved investigation for both direct effects of individual indicators, but also factors based on said groups. An example of a quick assessment of the circularity (ci) with the business (bj) indicators is presented in Figure 1. It is evident that, based on the values of the indicators, the most strong relations are positive for c12 with b2 and b9, and negative with c11 and b5. This is a starting point for the analysis which can be further elicited with a combination of qualitative analysis.

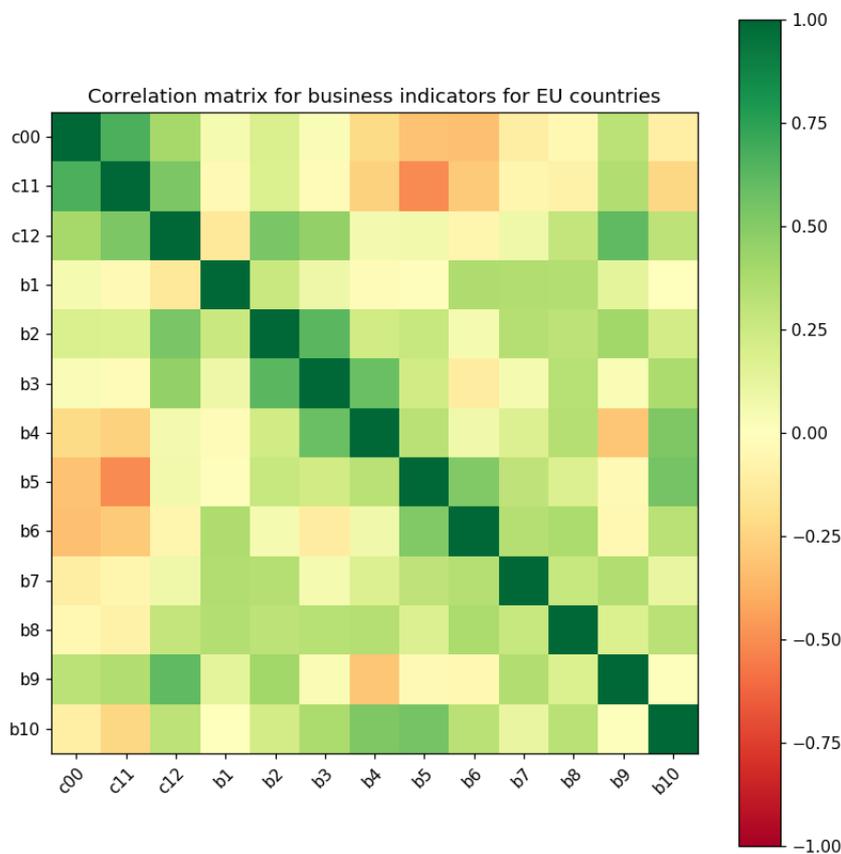


Figure 1. Example correlation of circularity with business related indicators.

With the use of the Sentence-BERT approach, indicator descriptions were encoded for similarity comparisons and clustering. An indicative output is a heat map for the similarities between the indicators of Figure 2. It can be noticed that 3-4 groups that contain indicators that are all similar to each other, as the indicators are partially grouped and ordered based on similarity.

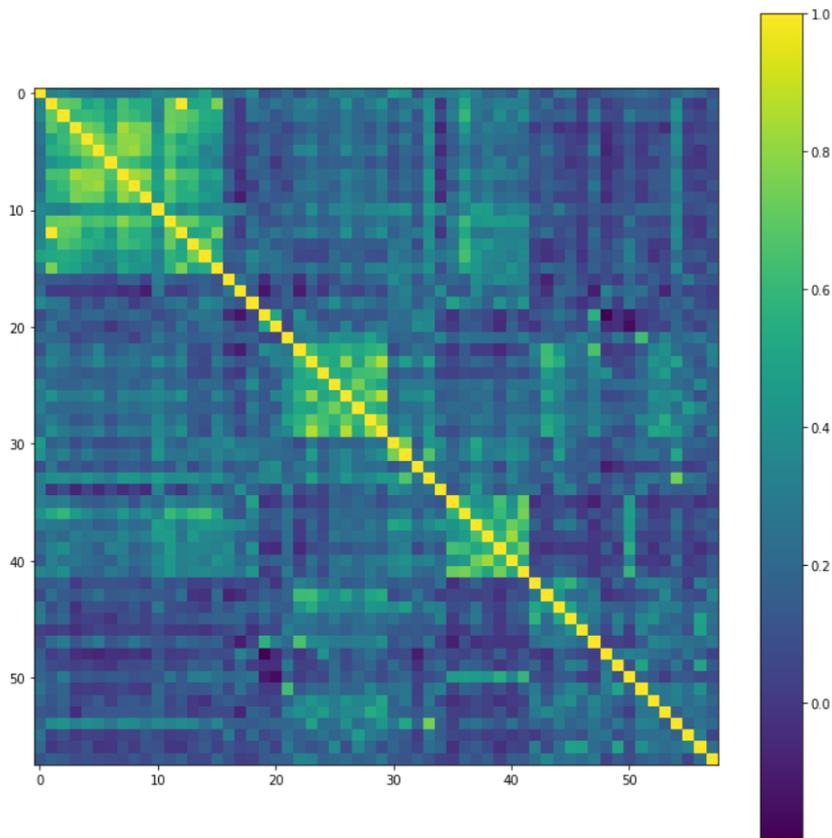


Figure 2. Heatmap for the similarities between the indicators.

An indicative example of clustering of the variables based on textual data if 5 clusters are selected classifies variables based on the following:

Cluster 1: [f2, f3, f4, f5, f6, f7, f8, f9, q2, q3, b1, b5, b6]

Cluster 2: [c11, c12, c21, c22, c23, c24, c25, c26, c27, c28, c29, c210, c31, c33]

Cluster 3: [Y, c32, g3, en1, en2, en3, en4, en5, en6, en7, b3]

Cluster 4: [g1, g2, c1, c2, s1, s2, i1, i2, q6, b7, b8, b10]

Cluster 5: [f1, i3, q1, q4, q5, b2, b4, b9]

This gives the opportunity to examine which indicators will be considered for further qualitative and quantitative analysis, by selecting the cluster that includes the independent variable (noted as Y) to be associated with variation of the remaining in-cluster exploratory indicators, and build empirical statistical models. Similarly the cluster containing "Y" if 3, 4, 6, and 10 clusters are created are presented as follows:

[Y, g1, g2, g3, c1, c2, f1, f2, f3, f4, f5, f6, f7, f8, f9, s1, s2, i1, i2, i3, q1, q2, q3, q4, q5, q6, b1, b2, b4, b5, b6, b7, b8, b9, b10]

[Y, c28, g1, g2, g3, s1, s2, i3, en1, en2, en3, en4, en5, en6, en7, b3]

[Y, c32, g3, en1, en2, en3, en4, en5, en6, en7, b3]

[Y, g1, s1, s2, i3].

FUTURE COLLABORATIONS (if applicable)

This is an ongoing co-operation with the host and his team. It is important to acknowledge, apart from Prof Trajanov, the inputs from Andrej Milovanovikj for the quantitative analysis and from Boris Tancovski for the textual analysis. We are working on two common publications to be presented in the INFER Symposium on Circular Economy and Sustainability, Alexandroupolis, Greece, 1-3 July, 2020 (<https://infer2020-thrace.gr/>). The ultimate objective is to publish this work to a scholarly journal.