

Using textual data to measure Social Capital

RESEARCH STILL IN PROGRESS

ALL SUGGESTIONS ARE WELCOME!

Challenge 1

Social Capital is difficult to define and measure. According to literature it has several dimensions and possible indicators. As a consequence the researcher has to **CHOOSE A DEFINITION** and select carefully his main aim, because...

the more **COMPLEX** is the concept to find
the more **COMPLEX** is find out
HOW TO MAP it in texts

Advantages

International Organizations (IOs) and Non Governmental Organizations (NGOs) are constantly working for a better world, but they always need to test their impact on communities' life. Measuring Social Capital is generally time consuming and data demanding. IOs and NGOs produce lot of texts for their ordinary activities (reports, policy papers, interviews...) As a consequence there are a lot of...

AVAILABLE DATA

Challenge 2

Usually these organizations work worldwide and official documents are written by officers with different backgrounds and language proficiency (most of them is in English used as Lingua Franca). As a consequence there is a lot of...

usually **GIVEN** and generally **not perfect**, full of Sociolinguistic variations.
DATA

Methodological TIP

ALWAYS CONSIDER SOCIOLINGUISTIC VARIATIONS and KNOW YOUR DATA AS BEST AS YOU CAN

OUR DEFINITION

Social Capital indicates institutions, relationships and norms that improve the quality and quantity of society's social interactions and facilitate collaboration among actors.

Methodological TIP

Do NOT TRUST blindly your **METHODS** and... try to **MIX**

OUR CASE STUDY

Stop TB Partnership

Our case study is the initiative Stop Tuberculosis (TB) Partnership (Stop TB), implemented by the World Health Organization (WHO) from 2010 to 2014. Our corpus includes 13 interviews referring to the Stop TB Partnership experience. The interviews were collected in 2012 by the central Stop TB Partnership Secretariat on the basis of an interview guide. All the interviews were recorded and fully transcribed by the same WHO officer.

CORPUS FEATURES
V = 2 324 type N = 22 221 tokens
Type-Token Ratio = 10.5%
Hapax legomena 41.2%
HUGE Sociolinguistic Variations

SOFTWARE
The corpus was pre-processed by means of the software Taltac2 (Bolasco, 2010), the textual data were processed by Iramuteq (Retinaud and De'jean, 2009).

OUR METHODOLOGY

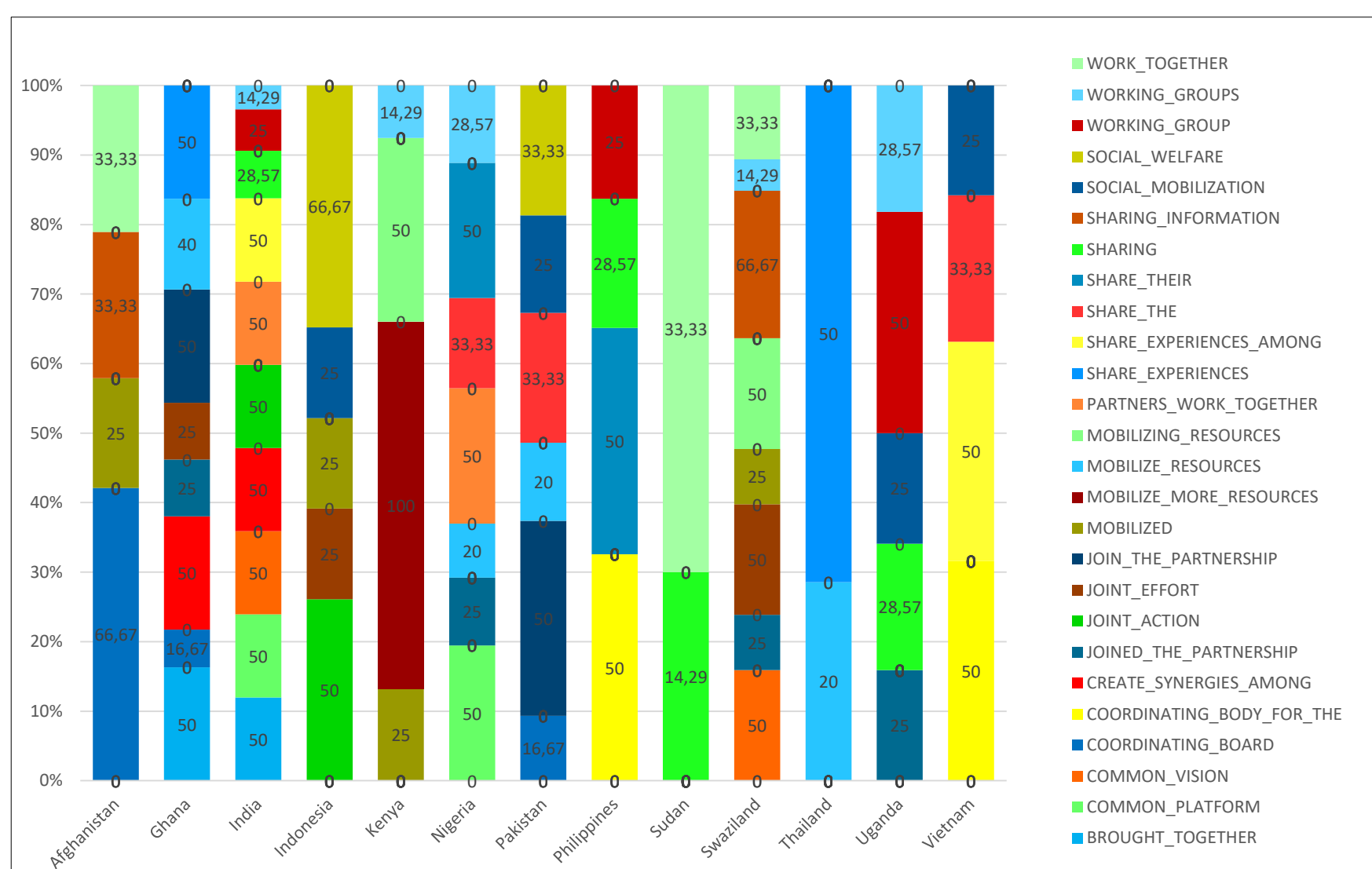
Considering the lack of homogeneity of our corpus we combine quantitative and qualitative techniques.

We analyze common, most frequent and most relevant (TFIDF-based evaluation) words and meaningful units. We built and map the distribution of two **SEMANTIC FIELDS** including words and meaningful units related to Social Capital.

The aim of our qualitative analysis was to **detect if the quantitative results were reliable**. We categorized every sentence of our corpus using an ex ante methodology, based on two questions:

- Do our speakers tell us **different stories** or they just use **different words**?
- Are semantic fields the right tool to measure Social Capital?

some QUANTITATIVE RESULTS

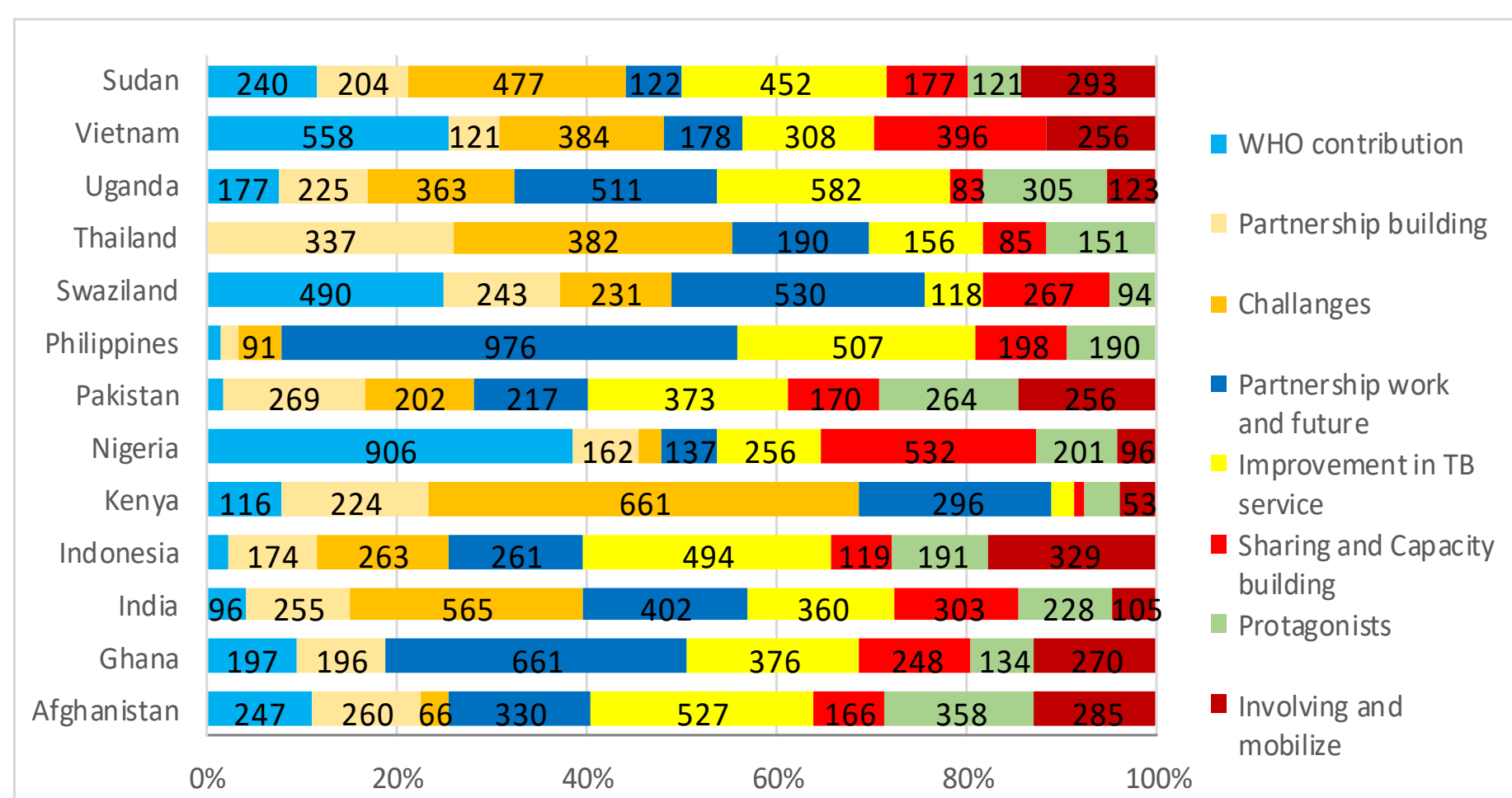


Distribution of words included in the semantic field *Working together*. Percentages are related to total occurrences of single expressions.

In progress...

- Stylometry
- POS Analysis
- Co-occurrences qualitative analysis of some keywords
- Co-occurrence networks

some QUALITATIVE RESULTS



Words count per codes. Percentages are related to total words of single interview. We underline descriptions codes in blue shades and Social Capital codes in red shades. Bars related to improvement in TB service were colored in yellow to facilitate the difference between sanitary improvements (included in partnership experience but not related to social impact) and Social Capital related codes.

Some References

Berruto G. (1995). Fondamenti di sociolinguistica. Editori Laterza
Bourdieu P. (1980). Le capital social. Actes de la recherche en sciences sociales. vol.(31): 25-30.
Coleman J.S. (1988). Social capital in the creation of human capital. American Journal of Sociology. vol.(5):35-42.
Cresswell J.W. and Plano Clark V.I. (2007). Designing and conducting mixed methods research. Sage Publications
Putnam R. D. (1995). Bowling Alone: America's Declining Social Capital. Journal of Democracy. vol.(6):65-78.
Sbalchiero S. and Tuzzi A. (2015). Scientists' spirituality in scientists' words. Assessing and enriching the results of a qualitative analysis of in depth interviews by means of quantitative approaches. *Quality and Quantity*. vol.(04):1-16.
Tuzzi A. (2003) L'analisi del contenuto. Introduzione ai metodi ed alle tecniche di ricerca, Carocci, Roma

Acknowledgements

I would like to express my deep appreciation to Professor Ajuna Tuzzi (FISSPA, Università degli studi di Padova), Professor Erika Teodora Uberti (CSCC, Università Cattolica del Sacro Cuore) and Dr. Stefano Sbalchiero (SPGI, Università degli Studi di Padova) for their valuable and constructive suggestions during the planning and development of this research work.

My sincere thanks also goes to Professor Simona Beretta (CSCC, Università Cattolica del Sacro Cuore) e Dr Giuliano Gargioni (WHO) who provided me the opportunity to join their team and who gave me access to Stop TB documents. Without their precious support it would not be possible to conduct this research.

Irene Saonara

Institutions and Policies PhD School

Eupolis Scholarship for Policy Analysis

Università Cattolica del Sacro Cuore (Milano)

Irene.saonara@unicatt.it