IMAGINÁRIO SOCIAL ONLINE: UMA ANÁLISE GLOBAL DA DESIGUALDADE DE GÊNERO E VALORES SOCIAIS NA INTERNET

GABRIEL MAGNO DE OLIVEIRA SILVA

IMAGINÁRIO SOCIAL ONLINE: UMA ANÁLISE GLOBAL DA DESIGUALDADE DE GÊNERO E VALORES SOCIAIS NA INTERNET

Tese apresentada ao Programa de Pós-Graduação em Ciência da Computação do Instituto de Ciências Exatas da Universidade Federal de Minas Gerais como requisito parcial para a obtenção do grau de Doutor em Ciência da Computação.

Orientador: Virgílio Augusto Fernandes Almeida

Belo Horizonte

Dezembro de 2020

GABRIEL MAGNO DE OLIVEIRA SILVA

ONLINE SOCIAL IMAGINARY: A GLOBAL ANALYSIS OF GENDER GAP AND SOCIAL VALUES IN THE INTERNET

Thesis presented to the Graduate Program in Computer Science of the Universidade Federal de Minas Gerais in partial fulfillment of the requirements for the degree of Doctor in Computer Science.

Advisor: Virgílio Augusto Fernandes Almeida

Belo Horizonte

December 2020

Silva, Gabriel Magno de Oliveira.

S586o Online social imaginary [manuscrito]: a global analysis of gender gap and social values in the internet / Gabriel Magno de Oliveira Silva.- 2020.
xv, 99 f. il.

Orientador: Virgílio Augusto Fernandes Almeida.
Tese (Doutorado) - Universidade Federal de Minas Gerais, Instituto de Ciências Exatas, Departamento de Ciência da Computação.
Referências: f.87-99.

1. Computação – Teses. 2. Análise de sentimentos – Teses. 3. Redes complexas – Teses. 4. Redes sociais on-line – Teses. 5. Vetorização de palavras – Teses. 6. Valores humanos – Teses. 7. Relações de gênero – Teses. I.Almeida, Virgílio Augusto Fernandes. II Universidade Federal de Minas Gerais, Instituto de Ciências Exatas, Departamento de Ciência da Computação. V.Título.

CDU 519.6*04(043)

Ficha catalográfica elaborada pela bibliotecária Belkiz Inez Rezende Costa
CRB 6ª Região nº 1510

UNIVERSIDADE FEDERAL DE MINAS GERAIS
INSTITUTO DE CIÊNCIAS EXATAS
PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIA DA COMPUTAÇÃO

FOLHA DE APROVAÇÃO

Online Social Imaginary: a global analysis of gender gap and social values in the Internet

GABRIEL MAGNO DE OLIVEIRA SILVA

Tese defendida e aprovada pela banca examinadora constituída pelos Senhores:

PROF. VIRGÍLIO AUGUSTO FERNANDES ALMEIDA - Orientador
Departamento de Ciência da Computação - UFMG

PROFA. JUSSARA MARQUES DE ALMEIDA GONÇALVES
Departamento de Ciência da Computação - UFMG

PROF. PEDRO OLMO STANCIOLI VAZ DE MELO
Departamento de Ciência da Computação - UFMG

PROF. HUMBERTO TORRES MARQUES NETO
Departamento de Ciência da Computação - PUC Minas

PROF. EDUARDO RIOS NETO
Departamento de Ciências Econômicas - UFMG

DR. INGMAR WEBER
Social Computing - Qatar Computing Research Institute

Belo Horizonte, 14 de Dezembro de 2020.

Aos meus pais, Márcia e Salvador, por todo amor e carinho.

À minha irmã Águata, por toda a amizade e companheirismo.

À minha companheira Flávia, por toda a motivação e inspiração para que eu seja uma pessoa melhor todos os dias.

Agradecimentos

Gostaria de agradecer à minha família (Márcia, Salvador e Ághata), que sempre me apoiou e me deu todo o suporte necessário para que eu pudesse me dedicar aos estudos e poder seguir a carreira acadêmica.

Também sou grato a meu orientador, Virgilio, que acreditou e investiu no meu potencial como pesquisador. Além de todo o conhecimento compartilhado, sempre foi uma pessoa muito compreensiva. Agradeço também ao professor Eduardo Rios-Neto, que nos apresentou diversas referências sobre valores humanos e também nos deu orientações fundamentais para a existência deste trabalho.

Agradeço ainda a meus amigos do laboratório CAMPS (Douglas, Evandro, Josemar, Matheus, Raphael e Yuri), que tornaram a vida acadêmica muito mais leve, divertida, inteligente e interessante. Seja em momentos de extrema concentração ou de procrastinação, eles sempre estavam lá para alegrar o dia.

Por fim, agradeço à Flávia, que, além de me ensinar várias coisas sobre Sociologia, foi meu porto seguro e sempre me motivou a percorrer os melhores caminhos. Ela tem um coração enorme, e me faz feliz todos os dias. É uma companheira incrível que tenho a sorte de ter a meu lado.

"It turns out that an eerie type of chaos can lurk just behind a facade of order - and yet, deep inside the chaos lurks an even eerier type of order."
(Douglas Hofstadter in 'Metamagical Themas: Questing for the Essence of Mind and Pattern')

Resumo

A medida que a Internet cresce em termos do número de usuários e em sua diversidade de serviços, ela se torna mais influente na vida das pessoas. Ela tem o potencial de construir ou modificar a opinião, a percepção mental e os valores dos indivíduos. O que é criado e publicado online é um reflexo dos valores e crenças das pessoas. Sendo uma plataforma global, a Internet é uma ótima fonte de informação para pesquisar a cultura online de países diferentes.

Neste trabalho nós desenvolvemos métodos para coletar e medir dados de diferentes fontes online para criar métricas digitais que capturam traços culturais e valores de diversos países. Essas métricas online são comparadas com outros indicadores socioeconômicos offline, para que possamos avaliar o que influencia o fenômeno online, e também medir a correlação.

Nós estudamos dois fenômenos: desigualdade de gênero e valores sociais. Na primeira parte, usamos o grafo social do Google+ para calcular métricas de redes complexas e compará-las entre mulheres e homens. Na segunda parte, desenvolvemos uma metodologia para calcular valores usando textos do Twitter com *word embeddings*. Mostramos que nossas duas abordagens são capazes de medir o relacionamento entre o online e o offline para dados internacionais.

Palavras-chave: Internet, Análise de sentimentos, Redes complexas, Redes sociais on-line, Vetorização de palavras, Cultura, Valores humanos, Relações de gênero.

Abstract

As the Internet grows in number of users and in the diversity of services, it becomes more influential on peoples lives. It has the potential of constructing or modifying the opinion, the mental perception, and the values of individuals. What is being created and published online is a reflection of people's values and beliefs. As a global platform, the Internet is a great source of information for researching the online culture of many different countries.

In this work we develop methods for collecting and measuring data from different online sources to create digital metrics that capture cultural traits and values for several countries. These online metrics are compared with other offline socioeconomic indicators, so that we can evaluate what drives the online phenomena and to also measure their correlation.

We study two phenomena: gender inequality and social values. In the first part, we use the social graph from Google+ to measure complex network metrics and compare between woman and man. In the second part, we develop a methodology to measure values using online written-text from Twitter utilizing word embeddings. We show that our two approaches are capable of measuring the relationship between online and offline for international data.

Keywords: Internet, Sentiment analysis, Complex networks, Online social networks, Word embeddings, Culture, Human values, Gender.

List of Figures

List of Tables

Contents

Chapter 1

Introduction

John Thompson defines the *social imaginary* as "the creative and symbolic dimension of the social world, the dimension through which human beings create their ways of living together and their ways of representing their collective life"[1]. Closely related to the social imaginary is the concept of *values*, which are one of the aspects that compose the culture of social groups, influencing their actions, modes of conduct and the way of thinking.

Besides the well-established cultural bounds of a society, another source of influence for the construction or modification of the mental perception, opinion and values of an individual is the Internet. What someone reads in the Facebook feed or the results she gets when querying Google search might heavily shape her decisions, from where she will have dinner to whom she will vote for.

In another perspective, looking at Internet users as producers instead of consumers, we believe that what people create and publish in the online world is a reflection of their values. In this work we study methods of capturing cultural traits and values from different online sources. We take advantage of the global aspect of the Internet and collect information from websites, profiles and publications from various countries, allowing us to investigate the online culture of several regions of the world.

After having a methodology to measure cultural traits in the online environment, we compare our metrics with offline indexes, such as the socioeconomic indicators from the World Bank, aiming at identifying what are the main characteristics that correlates with the online cultural traits. For the cases of high correlation, we could argue that our methodology would allow the Internet to be an alternative or extra source of information for building social indexes.

It is common to use the expression "In real life", or the equivalent acronym "IRL", to differentiate the activities in the physical and "real" world from actions in the "online

world" (e.g.: people chatting in a restaurant are chatting in real life, while people chatting in WhatsApp are chatting online). This division between online and offline is misleading. When people use the Internet, they are geographically located somewhere, and their online activities are potentially influenced by many aspects, like personality, culture, mood, etc.

The critic to the separation between the online and the offline has been discussed by some authors. With a geography-based approach, Graham argues that there is not an "online space" where one can transport into, what is happening is that we are mediating our actions with digital tools, while being influenced by algorithms and data, which will *augment* the world we are living [50]. In the area of philosophy, this topic has been discussed by Floridi, which argues that the online and offline spaces are becoming harder to distinguish, and even creates the term *onlife*:

> [...] the boundaries between the online and offline are disappearing, the appearance of the *onlife* experience, and hence the fact that the virtual infosphere can affect politically the physical space, that reinforces the sense of the political MAS [1] as a real agent. [40]

The entanglement between online and offline is so strong and evident these days, that companies and governments are using digital data to aid their actions. The MIT founded Thasos Group is collecting, processing and analyzing mobility data from smartphones to measure economic activities, planning to create an engine capable of comparing several countries [81].

Governments can also benefit from using online data to guide urban planning and to improve government projects, having impact on several areas such as transportation, energy consumption and others. The *smart city* can be a role model, where a plethora of connected systems and devices collect and share massive data about residents and the environment, which helps us to understand the complex networks of urban dynamics, to improve life and to create solutions for societal problems [41].

There is also evidence of internet technology shaping the culture. For instance, it is being argued that video streaming services, like Netflix, are making people return "[...] to the cultural era that predated radio and TV, an era in which entertainment was fragmented and bespoke, and satisfying a niche was a greater economic imperative than entertaining the mainstream" [79]. There are other arguments on the influence of digital

[1] MAS: Multi-agent System

technologies, claiming how the Internet is changing our way to communicate [112], eat [113], dress [23], and other facets of our life.

Assuming, therefore, that the online and offline worlds are strongly interconnected, we compare and contrast data obtained from offline sources with data from online platforms. We believe that the "online space" is an extension of what people do and think in their lives. What we advocate is that it is possible to capture trends and signals from the online environment and show rather they are noticeable in the offline.

In this work we explore the linkage of online and offline in a unique way. While there are research on how digital data can be used to capture offline characteristics (and vice-versa), most of these studies relies on comparing different aspects in each world. In our case, we focus on comparing the online measurement proposed by us, with the equivalent offline measurement of the *same phenomena*. Further, we have an *international approach* to cover several countries in the world, so that we can investigate different cultures.

First, we investigate the relationship between online measurements with an specific offline social index. The Global Gender Gap report gives a rank for countries related to the inequality between women and men. We gather a large dataset from Google+ and measure complex network metrics of its social graph, contrasting female with male users. We then calculate an online gender gap score for these metrics for several countries, then compare with the actual offline gender gap. Among other findings, we show that while countries with more men online have higher gender inequality offline, the opposite holds for popularity metrics such as in-degree and PageRank: in countries with higher offline inequality, woman are actually followed more.

In the second part of our work, we study human values, a key characteristic of the culture of a social group. Values are traditionally measured with a survey, with questions regarding actions, opinions and habits of the person related to a multitude of topics, such as religion and science. We propose and evaluate a methodology to measure values online by using word embeddings. The technique relies on an association test applied to the vectorial space and the distance between words that are related and expected to capture a specific question of the World Values Survey. We collect a huge collection of tweets and create a word embedding model for each country, and test our method in this dataset. By correlating the offline value of a question with the online value of our association test, we observe that there's indeed a link between the online and the offline.

These are our hypotheses:

- **H1**: Gender ratios of online social network metrics are different among different

countries;

- **H2**: Gender ratios of online social network metrics of countries are correlated with their corresponding Global Gender Gap score (offline);
- **H3**: Online human values measured with word embeddings have different signals (positive or negative) according to the target human value (e.g. religion, family, science).
- **H4**: Online human values measured with word embeddings are different among countries.
- **H5**: Online human values measured with word embeddings are correlated with their corresponding World Values Survey score (offline).

We believe that our two approaches are novel in measuring the relationship between online and offline for international data. We propose and describe the algorithm to measure the phenomena with online data, and evaluate it by comparing with robust and independent offline data. Interestingly, we show that both the online social graph and the online text are valuable sources for capturing social traits.

We highlight here the main contributions of this work in relation to new methodologies and also the application of these methodologies:

- A methodology (GR - Gender Ratio) that allows measuring differences (gaps) between women and men in relation to online social network metrics for different social groups (e.g. countries);
- A list of complex network metrics (e.g. in-degree, recriprocity, assortativity, etc) useful for measuring and analyzing gender gaps in online social networks;
- A methodology to identify the country of a tweet, considering the self-declared free-text location provided by the author of the tweet;
- A methodology (OVI - Online Values Inquiry) that allows measuring human values from textual data using word embedding models, focused on (but not limited to) online texts published online;
- A methodology that aggregates Twitter and Wikipedia data to create different types of word embedding models for different countries and languages;
- A list of 24 OVIs (Online Values Inquiry), which is a list of sets of words inspired by questions from the World Values Survey, that can be used to measure certain human values using our OVI methodology.

Further, we apply the aforementioned techniques and methodologies in the two previously mentioned contexts (Google+ for online gender gaps, and Twitter for online human values) and discover some interesting findings, such as:

- Countries with more men than women online are countries with more pronounced gender inequality;
- Women are more tightly cliqued and their links are more reciprocated;
- In countries with higher offline inequality women are, suprisingly, followed more than men. This result holds both using the mean and the median, and it holds for other "status" metrics such as PageRank;
- Countries with a larger fraction of within-gender social links, rather than across-gender, are countries with *smaller* offline gender inequalities;
- Countries with larger offline gender inequalities have a larger "differential assortativity" where women have a stronger preference for within-gender links than men;
- Applying existing gap-based methodology to online data yields a strong negative correlation, up to $r = -0.76$ (p-value < 0.05), with existing offline measures;
- Different online human values have different patterns: some are predominantly positive, others predominantly negative, and also diverse (both positive and negative, depending on the country);
- The OVI methodology is capable of capturing differences between countries in relation to online human values, presenting a diversity of scores between the countries for target values;
- When comparing the four types of model, we observe that using the same language for all the countries might be a good compromise in scenarios where creating a multi-language inquiries is infeasible;
- By using a factor analysis approach, it is possible to cluster the countries in relation to their intrinsic similarities of the online values, creating an online cultural map of the countries;
- The *signal* of the online values has a relatively high match with its corresponding offline value, meaning that an offline overall positive agreement for a certain value will also have an overall positive score online;

- There is a strong positive correlation between the online and the offline for *some* human values, specially for the inquiries related to religion;
- By analyzing several socioeconomic factors, we observe that geographical location and the digital infrastructure of the countries are strong characteristics related with the online values.

Our goal is not to construct and publish online indexes ourselves, but to provide the methodology that allows one to measure social online characteristics. We apply and validate our methods with online data from specific sources (Google+ and Twitter), but they are generic sufficiently to be used in different contexts (i.e. other online social networks and general Internet sources). Natural Language Processing (NLP) researchers could use it to measure and evaluate intrinsic values present in their text corpora and model.

We believe our work is specially useful for social sciences specialists, such as demographers and sociologists, that can use their domain knowledge and expertise to create their own analyzes, allowing them to investigate gender gaps and human values in the online environment.

The rest of this dissertation is organized as following. On Chapter 2 we present and discuss several related work. Chapter 3 presents our methodology for measuring gender gap with online social graphs. Next, on Chapter 4, the study of measuring online social values with word embeddings is shown. Finally, in Chapter 5 we conclude our work and discuss its implications.

Chapter 2

Related Work

In this section we will present several research papers that are related to ours in different aspects. Some authors compare offline and online data from different sources, others focus specifically on measuring values online, and there is also research on creating online indexes.

As far as we are aware, this is the first study that links online gender differences in dozens of countries to existing quantitative offline indicators. However, lots of valuable research has been done looking at gender differences and gender inequality offline and online separately and such work has considered various psychological, sociological and economical differences. It is not within this work's scope to serve as a complete review of literature in gender studies but, rather, it should give the reader a good overview of aspects that have been investigated.

2.1 Gender in the Offline environment

Feingold conducted a meta-analysis to investigate differences in personal traits between genders as reported in literature [37]. For some traits such as extroversion, anxiety and tender-mindedness, women were higher, while for others such as assertiveness and self-esteem, men had higher scores. And, as one might hope, there are also traits with no observed gender differences such as social anxiety and impulsiveness.

Pratto et al. studied gender differences in political attitudes [97]. By analyzing a sample of US college students, they found that men tend to support more conservative ideology, military programs, and punitive policies, while women tend to support more equal rights and social programmes. They also show that males were in general more social dominance oriented than females.

Costa et al. [31] aggregated results of psychological tests from different countries for the so-called "Big Five" basic factors of personality: Neuroticism, Extroversion, Openness to Experience, Agreeableness, and Conscientiousness [93]. They observed that, contrary to predictions from the social role model, gender differences concerning personality were most pronounced in western cultures, in which traditional sex roles are comparatively weak compared to more traditional cultures. In a similar line of work, Schmitt et al. [105] conceived the General Sex Difference Index and observed that sex differences are higher in Western cultures compared to non-Western ones.

Hyde performed a meta-analysis on psychological gender differences to show that, according to the gender similarities hypothesis, males and females are alike on most psychological variables, contrasting the differences model that states that men and women are vastly different psychologically [59].

2.2 Gender in the Online environment

2.2.1 Gender gap

Bimber analyzed data from surveys in the United States, in which people were asked about Internet access and frequency of utilization [17]. His analysis showed that there is a gap in access regarding the gender, but that this gap is not related to the gender itself, but rather to socioeconomic factors, such as education and income.

Collier and Bear investigated the low participation of women in terms of contributions to Wikipedia [30]. They found strong support that the gender gap is due to the high levels of conflict in discussions, and also due to a lack of self-confidence in editing others' work.

Iosub et al. investigated the communication between editors in Wikipedia and observed that female editors communicate in a way that develop social affiliation [64]. In terms of online social network usage in the US in 2013, women had higher rates of users for Facebook, Pinterest or Instagram, whereas usage was similar for both genders for Twitter and Tumblr [25]. In our data for the US, we have more male users. A possible explanation for this is an increased concern for privacy with a corresponding choice to reveal less information about themselves. See related work further down on this subject.

Most of our gender gap study from Chapter 3 was presented and published in SocInfo 2014 [78]. After that, the research on gender gap in the Online environment developed, and we highlight two papers. Fatehkia et al. drives an international approach to investigate the correlation and prediction potential of gender gap index (GGI) using

Facebook advertisement data (country, gender, age and phone model) [36]. They compare the Facebook online data with with ITU Internet and GSMA Mobile Phone , as well as with the Global Gender Gap Report (GGGR) scores. Facebook GGI are highly correlated with the ground truth variables (up to 0.834). Also, the regression model combining online (Facebook) and offline (GGGR) has good fit quality scores (Adjusted R-squared of 0.791).

Another paper studying gender inequality using advertisement data from Facebok is the work from Garcia et al. [42]. They propose the FGD (Facebook Gender Divide) metric, which is the logarithm of the ration between activity ratios for men and women. They evaluate the FGD using a linear regression with gender equality indices measure by the World Economic FOrum and other control variables (Internet penetration, population, economic inequality). The model had good fit (R-squared of 0.74) and consistent with survey samples.

2.2.2 Privacy and interests

Researchers investigated whether there is a difference between genders regarding the kind and amount of information shared online. Thelwall conducted a demographic study of MySpace members, and observed that male users are more interested in dating, while female users are more interested in friendship, and also tend to have more friends [115]. When analyzing the privacy behavior, women were found to be more likely to have a private profile. Joinson analyzed reports on motivation to utilize Facebook [66]. He found that female users are more likely to use Facebook for social connections, status updates and photographs than male users. Also, female users are more prone to make an effort to make their profile private.

Bond conducted a survey among undergraduate students regarding their utilization in OSNs and found that female participants disclose more images and information on OSN profiles than male participants [21]. They also observed that the kind of content shared between genders are different. For instance, female users tend to share more content about friends, family, significant others, and holidays, while male users are more likely to post content related to sports. Other works also investigated the vocabulary used by users in OSNs, and found that there are differences regarding the semantic category of words between women and men [92, 33].

Quercia et al. studied the relationship between information disclosure and personality by using information from personality tests done by Facebook users, and found out that women are less likely than men to publicly share privacy-sensitive fields [100].

2.2.3 Network

Szell and Thurner analyzed the interactions between players of a massive multiplayer online game [114]. They constructed the interactions graphs and observed that there are differences between male players and female players for all kinds of connections. For instance, females have higher degrees, clustering coefficient and reciprocity values, while males tend to connect to players with higher degree values. Ottoni et al. also investigated the friendship connections of the users in Pinterest and observed that females are more reciprocal than males [92]. In our analysis , we also found women to have a higher clustering coefficient and a larger fraction of reciprocated friendship links on Google+. Heil et al. analyzed Twitter data from 300 thousand users, and found that males have 15% more followers than women. When looking at homophily, they found that on average men are almost twice as likely to follow other men than women, and, surprisingly, women are also more likely to follow men [57, 86]. In our analysis, we observed homophily for both genders in Google+, i.e. females tend to follow more females and males to follow more males. Recent work has also looked at generalizing concepts from the "Bechdel Test"[1] to Twitter [43]. The authors look at tweets from the US for users sharing movie trailers, which are then linked to Bechdel Test scores, and they find larger gender independence for urban users in comparison to rural ones, as well as other relations with socio-economic indicators.

2.3 Comparing online and offline data

Here we show papers that, like ours, compare online behaviour with other offline information. Most of them are focused on using the geo-location as the online source, while comparing it with different offline information such as activity inequality, migration, personal interests, political opinion and language patterns.

Garcia-Gavilanes et al. studied the link between actions of people on Twitter and their respective culture (country cultural traits) [45]. They collect a random sample of Twitter users tweeting in March 2011, also collecting their followees (out links), totalizing 2.34 million users with their country identified looking into the 'location' field (similar to our methodology used in the values study presented in this work). Three cultural traits area analyzed, by measuring an online behavioral metric in twitter and calculating the correlation with a proper behavioral index of the country. First, they compare the Pace of Life of a country and the temporal predictability of mentioning users in Twitter, and found out that countries with higher the paces of life are easier

[1] http://en.wikipedia.org/wiki/Bechdel_test

to predict (correlations of $r = -0.62$, $r = -0.68$ and -0.58 for tweet posting, user mentioning and tweeting in working hours), a consistent result with findings for offline behaviour. Secondly, they compare the Individualism (from Hofstede's cultural dimensions[58]) with interaction level in Twitter, and conclude that countries that are more individual tend to mention each other less ($r = -0.55$). Finally, they analyze whether users from countries with high power-distance (Hofstede's dimension measuring how "comfortable" people are with inequality of power) will prefer to interact (follow, recommend and accept recommendation) with more popular users, concluding that it is indeed the observed pattern (correlations of $r = 0.62$, $r = 0.33$ and 0.42 for following, recommending and accepting recommendation respectively). Their results reinforces the argument that cultural differences can be observed (and measured) with online data. Garcia Gavilanes also discusses and proposes a general methodology to measure cultural traits in online social media, with the goal of representing Hofstede cultural dimensions with online characteristics [46].

García-Gavilanes et al. focus on investigating the cross-country communication between users in Twitter [44]. A similar dataset and data collection methodology used in the other previously cited Garcia paper [45] was used, totalizing 13 million geo-located users. It was only considered countries with at least 1,000 users in their sample, totalizing 111 countries. In the first analysis, they calculate the correlation between the actual physical distance of countries (using the gravity model and Haversine distance) and volume of communication, and conclude that the number of mentions and retweets are moderately correlated ($r = 0.68$ and $r = 0.66$ respectively). The next step was to build a regression model using more variables besides the gravity model to predict the number of unique mentions between countries. They use three categories of variables: economic (income, exports, trade intensity, and trade market share), social (routes, emigration, migration, migration rate), and cultural (language, intolerance, and the Hofstede's cultural dimensions). For measuring intolerance they use the World Values Survey question regarding rather someone would want neighbors from a different race. Their regression model performed well (R-squared of 0.80), indicating that social economic and cultural characteristics of the countries are important to explain the communication in the online environment.

Silva et al. proposed a methodology to measure similarities and identify boundaries between people from different populations related to food and drink consumption [108]. They collect Foursquare check-ins through Twitter, covering a single week of April 2012. The check-ins are categorized in 101 sub-categories (such as 'Bar', 'Breakfast' and 'Gastropub'), which are then grouped in three classses (Drink, Fast Food and Slow Food). By aggregating the number of checkins among the sub-categories in

a particular geographical area, they extract a *cultural signature* of that location, and by clustering locations with similar cultural signatures it is possible to show a cultural map of countries (and also cities and regions) with similar food consumption culture. Finally, they compare their results with the cultural map of the world given by the World Values Survey by computing the Spearman's rank correlation between their approach and the WVS one, concluding that there is indeed high similarity. This work is similar to ours in the sense of comparing online-measured cultural traits with offline survey data from World Values Survey, but it is different in the sense that we measure values using text data from Twitter, and they measure food consumption using check-ins from Foursquare.

Althoff et al. studied the physical activity of people from several countries in the world [2]. They gathered a dataset from a smartphone software company consisting of step counts from over 700 thousand people. They aggregate the data in the country level and create an "activity inequality" index, consisting of the gini coefficient of the population activity distribution. Having this index, they correlate with other data. For instance, they found out that activity inequality is positively correlated with obesity levels. They also observe that a higher activity inequality is associated with a higher gender gap of activity in the countries. This work resembles ours in the sense that it is calculating an index in the country level and correlating it with other indexes, but its important to notice that their index does not use data from online activity, even though being collected from an online application. The data actually represents *step counts*, which is an offline action.

Fiorio et al. investigated migration patterns using a sample of geo-referenced tweets located in United States [38]. By using *migration curves* as a theoretical framework, they categorize and aggregate twitter users based on time and location. Among other results they show that there is a negative relationship between migration rate and the duration, and a positive relationship between migration rate and the interval. The authors argue that their methodology could be a faster and more precise solution compared to the conventional methodology of using surveys. This work uses online data (tweets) to measure and analyze patterns of an offline activity (migration), but it doesn't actually compare them directly. Also, even though the methodology could apply for other regions besides U.S., it does not compare migration patterns between countries.

Guo et al. developed a probabilistic framework that identifies the interests of the users relying on their physical movement (footprint GPS information) [54]. The concept of *personal interests* used in the paper is a generic set of textual topics or tags associated to the user. Their method explores the relationship between the topics associated to

partitioned regions of the city, and then uses the transitions and movement patterns of the users between these regions to infer their interests. Their methodology is generic enough to use any source of data. Footprints are essentially offline (even though could be gathered via online sources), and "interests" can be seen as a personal trait, which in their case are expressed and published in the online world, but could have also been collected from offline sources (e.g. survey).

A recent paper from Bastos et al. investigated whether echo-chamber communication in Twitter derives from offline location clustering, in the context of the Brexit campaign [13]. The authors collected tweets using a set of keywords and hashtags related to Brexit, then classify the users as "Remainer" or "Leaver" based on highly-charged hashtags related to both positions. After that, they analyze the interactions (retweet and mention) between users and their respective locations (extracted from tweet information). Among other interesting results, they show that in-bubble communication (echo-chamber) is associated with the geographic distance. In this work authors analyze the relationship between online activity (users interaction) and offline information (physical location), both information being extracted from an online source (Twitter).

Abitbol et al. looked into the variability of linguistic patterns in Twitter compared to other external social factors [1]. A twitter dataset was created, consisting of 170 million tweets written in French, containing the (preprocessed) text of the tweet, the social network of mentions, and the geolocation position. A second dataset was gathered and combined with the previous one, which contain sociodemographic indicators from geographic locations in France. Three linguistic variables related to the French language were calculated for the users. By using a regression analysis, they found out that people of higher socioeconomic status, and people from the southern part of the country, use a more standard language, and people that interact with each other are also closer in terms of linguistic similarity. In this study, an online activity (linguistic characteristics of text written in the Internet) was compared with offline information (socioeconomic status and geographic position).

2.4 Survey of online behaviour

With the goal of understanding how people use Internet and how online data could be used to enhance offline studies, some authors interrogated groups of people.

An article from Baghal et al. studied consent decisions in surveys in the UK, with the goal to evaluate the potential of combining Twitter data with survey data for

social studies [11]. They gathered responses from 3 surveys that had questions about revealing the twitter handle name and the consent to use this data. Their results revealed that the consent rates are relatively low (27% to 37%), older respondents are less likely to consent, and also people responding from the web, even though the later being more likely to have a Twitter account. The authors argue that even with the low consent rate the Twitter information can be used to enhance the data collected in the survey, but it is important to take care when archiving and sharing the data, making sure the users privacy or the social media platform terms are not violated.

Dutton and Reisdorf studied the phenomena of digital divides and how the attitudes of users could be used to identify cultures of the Internet [35]. The authors used data from a survey organized in Michigan (U.S. state), and in total they have 995 adult respondents. Using answers from 10 questions related to attitudes and beliefs regarding the Internet. They deploy a PCA (Principal Component Analysis) technique, where they find four components, and then conduct a cluster analyses to finally identify 5 distinct cultures: 'Digital-doubters', 'Instrumentalists', 'Cyber-wary', 'Cyber-savvy', and 'Asocials'. This work is different from ours and most of the other works, in the sense that it gathers offline data (survey) to understand online activity patterns, instead of the opposite (gathering online data to understand offline activity). It also corroborates with the idea that the Internet is an environment in its own, developing particular traits, behaviour and cultures.

2.5 Predicting Values

Our work is not the first to measure values with online data, but it is, as far as we know, the first one to use word embeddings, and the first one to apply an international approach covering several countries. The papers we will show next work in the *individual level*, while our study adopts a technique in the *aggregated level*. Another difference is that they focus on *prediction*, while we explore the *comparison* of the online data with the offline data, proposing the methodology to measure the online value.

Chen et al. conducted a study to investigate the relationship between word use in social media with human values [27]. Their approach was to interview users of Reddit through a survey that captures values according to Schwartz values framework [106], and then collect the corresponding user comments and posts in Reddit. By using the words utilized in the posts/comments, they calculate for each participant the percentage of utilization of LIWC categories. Following, the authors calculate the correlation between Schwartz' Values and LIWC categories, and also evaluate the

potential of prediction. They conclude that there is, indeed, a relationship between personal human values and word utilization for some categories, and also a considerable predictive potential. There are some important differences with our work. Besides the fact that we use Twitter and they use Reddit, we observe that their work analyze only english-speaking people and do not compare countries.

Closely related to predicting personal human values, is predicting personality traits. Youyou et al. investigated the possibility of computers evaluating the personality of humans [128]. First, they asked volunteers to complete a personality questionnaire of 100 items, totaling more than 80 thousand respondents. The answers for the questionnaire is used to calculate the self-reported personality, which is the baseline. Next, the authors ask the friends of the volunteers to classify the personality of the correspondent person, and compare with a linear regression model based on the likes of the participant in Facebook. They show that the computer predictions are more accurate than the judgment made by the friends ($r = 0.56$ and $r = 0.49$ respectively).

A study [68] from Kalimeri et al. explored the relationship between digital behaviour and demographic, psychological and human values personal information. They developed a survey that contained questions about demographic information (age, gender, location, education, health, political, etc), a psychometric questionnaire (five dimensions) and a human values questionnaire (ten dimensions). Besides the survey itself, they gather browser traffic data, either from desktop or smartphone. In total they have over 7 thousand participants. They apply a prediction experiment using the Random Forest algorithm, where they try to classify information replied on the survey based on a vector of visited domains from the traffic data. They have satisfactory results for the demographic part (accuracy up to 90%), while for moral traits and human values they obtained poor performance (accuracy around 60%). Comparing to our work, we can identify some similarities, but there is also some fundamental differences. Similar to us, they use online data (websites accessed) to compare with offline data (replies in a survey). Besides the previously mentioned fact that they do not compare countries, they rely on traffic data, while we use textual data published in Twitter.

2.6 Online Indexes

Putting aside the concrete issue of gender inequality and values, we are essentially interested in using online data as a socio-economic indicator. This idea in itself is not new and previous research has attempted to estimate things such as unemployment rates [5], consumer confidence [90], migration rates [130, 56], values of stock market

and asset values [19, 18, 131] and measures of social deprivation [101]. Work in [98] is also related as it looked at search behavior, in this case "forward looking searches" and links such queries to estimates of economic productivity around the globe.

Ballatore et al. looked into the phenomena of digital information hegemony in the world [12]. They propose to study which countries produce their own representation of city-related content. To achieve that, they create an indicator of localness of search results, which given a certain keyword and a country, measures the portion of the results that are from that particular country. They observe that there is, indeed, a variance on the localness of the countries (i.e. some countries are more local than others). In a next step, they compare the localness with other "offline" indexes such as population, GDP, and tourism metrics. They found out that metrics related to scientific publications are the ones that better explain the localness of the countries. This work is very similar to our in the sense that it calculates an index with online information and compare it with other offline metrics. The difference is that it studies only one aspect (digital hegemony) and uses search results as online data, while our work studies several aspects (values and questions from WVS) and uses textual data from Twitter as source of online information.

Ojanperä et al. developed an index that measures content creation and participation in the online environment for several countries, called *Digital Knowledge Economy Index* (DKEI) [91]. To build the index three sources of online data are used: (1) number of commits in GitHub, (2) number of edits in Wikipedia and (3) number of registered domains. A normalization algorithm is applied for the three variables and then a simple average is calculated, resulting in an index for each country (relative to 2013), which is included as a new sub-index for an existing World Bank Knowledge Economy Index. They present not only the final DKEI scores, but also analyze rather the rank of the countries changed when adding the new digital participation sub-index. This work is analogous to ours in the sense that it also uses online data to build a score for several countries, but in their case they are measuring digital participation, and we are measuring values.

Chapter 3

International Gender Differences and Gaps in Online Social Networks

Gender equality and full empowerment of women remains elusive in most countries around the world. Women are often at a significant disadvantage in fields such as economic opportunities, educational attainment, political empowerment and in terms of health [55]. Reducing and ultimately erasing the "Gender Gap" in these fields is both an intrinsic, moral obligation but also a crucial ingredient for economic development [69]. By limiting women's access to education and economic opportunities an immeasurable amount of human resource is lost and huge parts of the population are not able to develop their full potential.

To quantify gender inequality around the globe and to track changes over time, for example in response to policies put in place, the World Economic Forum annually publishes "The Global Gender Gap Report" in collaboration with the Center for International Development at Harvard University and the Haas School of Business at the University of California, Berkeley. This report ranks countries according to a numerical gender gap score. These scores can be interpreted as the percentage of the inequality between women and men that has been closed and so a large gap score is desirable. In 2013 the leading country Iceland had an aggregate score of 0.87, whereas Yemen scored lowest with 0.51. Scores are based on publicly available "hard data", rather than cultural perceptions, and variables contributing include the ratio of female-to-male earned income and the ratio of women to men in terms of years in executive office (prime minister or president) for the last 50 years. The emphasis of the report is on the relative gender difference for the variables considered rather than the absolute level achieved by women.

This work contributes to this line of work by quantifying gender differences around

the globe using existing methodology and applying it to *online* data, concretely data derived from Google+ for tens of millions of users. We start our analysis by describing the absolute differences along dimensions such as the number of male vs. female users or their virtual, social ranking in terms of number of followers. Our main emphasis is on studying correlations between online indicators of inequality and existing offline indicators. We do this both for the purpose of validation, to be sure that what we measure is linked to phenomena in "the real world", and for the purpose of devising new indicators, where a seemingly important online measure does not seem to be in good agreement with existing indicators.

Our current study is deliberately done *without* doing analysis of the content shared by men and women in different countries, and we are only relying on network structure data. One reason for this choice was one of global coverage: doing any type of content analysis for languages spanning all continents and having results comparable across languages and countries remains a fundamental challenge. Doing something only for English would have beaten the purpose of measuring gender inequality online in virtually all developing countries. A second reason for our choice was the fact that current indices are based on "hard data". Whereas the number of followers is well-defined, things such as the sentiment or mood of a user are hard to measure in an objective manner and are difficult to compare across cultures.

Analyzing gender differences for 73 countries we find both expected and surprising trends. Our main findings are:

- Countries with more men than women online are countries with more pronounced gender inequality.
- Women are more tightly cliqued and their links are more reciprocated.
- In countries with higher offline inequality women are, suprisingly, followed more than men. This result holds both using the mean and the median, and it holds for other "status" metrics such as PageRank.
- Countries with a larger fraction of within-gender social links, rather than across-gender, are countries with *smaller* offline gender inequalities.
- Countries with larger offline gender inequalities have a larger "differential assortativity" where women have a stronger preference for within-gender links than men.
- Applying existing gap-based methodology to online data yields a strong negative correlation, up to $r = -0.76$ (p-value < 0.05), with existing offline measures.

Generally our analysis is more quantitative and descriptive rather than qualitative and diagnostic. Though we describe the gender differences we find and comment on whether they agree with (at least our) expectations, we do not attempt to give explanations. We hope that experts in domains such as gender studies or social psychology will find our analysis useful and that it can save as a starting point for more in-depth studies focused at the root causes of what we observe.

As more and more economic activity becomes digital and moves online, as more and more education happens online through MOOCs and other initiatives, and as more and more political engagement happens online, we are convinced that, ultimately, quantifying gender inequality also has to crucially take into account online activity.

3.1 Data Set

Our dataset was created by collecting public information available in user profiles in the Google+ network. We inspected the *robots.txt* file and followed the sitemap to retrieve the URLs of Google+ profiles. Since we retrieved the complete list of profiles provided by Google+, we believe our data set covers almost all users with public profiles in Google+ by the time of the data collection. The data collection ran from March 23rd of 2012 until June 1st of 2012. When inspecting the sitemap we found 193,661,503 user IDs. In total we were able to retrieve information from 160,304,954 profiles. Some IDs were deleted or we were not able to parse their information. With the social links of the users, we have constructed a directed graph that has 61,165,224 user nodes and 1,074,088,940 directed friendship edges.

Country identification. To identify a user's country in Google+, we extracted the geographic coordinates of the last location present on the *Places lived* field and identified the corresponding country. We were able to identify the country of 22,578,898 users.

Gender. Google+ provides a self-declared gender field where the user can choose between three categories: *female*, *male* and *other*. As any other profile field in Google+ (except for the name), it is possible to put this information as private, so we do not have this information for all users. Of the 160 millions users, 78.9% provided the gender field publicly, from which 34.4% are female, 63.8% are male and 1.8% selected "other". It is important to notice that Google+ attracted more tech-savvy users [77], and since there's a known gender gap in IT [4], that is probably the reason of Google+ having more men than women overall.

Details of the Google+ platform and a data characterization of an early version

of the dataset are discussed in a previous work [77]. A summary of the number of users for each country can be found in Table 3.2. We only selected countries with at least 5,000 users for each gender.

3.1.1 Online Variables

As doing any type of content analysis for dozens of languages and cultures is extremely challenging, we decided to study how *network* metrics could be indicators for gender gaps. At the country-level, we looked at the following metric which we hypothesized could be an indication of online gender segregation.

- The *assortativity*[1] is the fraction of links to the same gender rather than across genders. A large value can be indicative of either strong same-gender linkage preference, or simply a highly imbalanced gender distribution of the users, which trivially makes cross-gender links less likely.

We also computed the following metrics for each user from the 73 countries in our data set.

- The *in-degree*, also referred to as the number of followers, counts the number of "circles" a user is in. A large in-degree can be seen as an indicator of popularity or status.

- The *out-degree*, also referred to as the number of followees or friends, counts the number of users a user has in their circles.

- The *reciprocity* is the fraction of reciprocal links in relation to the out-degree, i.e. the fraction of times where the act of following is reciprocated by the receiving user.

- The *clustering coefficient* for a particular node is the probability of any two of its neighbors being neighbors themselves. It is calculated by the fraction of the number of triangles that contain the node divided by the maximum number of triangles possible (when all the neighbors are connected), which for a directed graph is equal to $n(n-1)$, where n is the number of neighbors that reciprocate the connection. A large value typically indicates a large degree of "cliqueness" and more tightly connected social groups.

[1] We use "assortativity" rather than "homophily" to emphasize the correlation rather than necessary a causal link.

- The *PageRank* measures the relative importance of a user in the network and, unlike the mere in-degree, is influenced by the "global" social graph structure. A damping factor $d = 0.85$ was used for the iterations of the algorithm. A large PageRank value is often thought of as an indicator of "centrality" or "importance" in the social graph.

- The *differential assortativity* is the "lift" of the fraction of users of the same gender followed by a particular user. It is calculated by dividing the fraction of links to the same gender by the share of that gender for the country of the user. A large value means that users are more likely than by random chance to follow other users of their same gender. The comparison against random chance corrects for the fact that, for example, in an online population with male predominance (e.g.: 80% males, 20% females), men are trivially more likely to follow other males even without any same-gender homophily.

These per-user metrics are then aggregated into a per-country score as described in the next paragraph. Though we group the results by country, connections across countries are included in our analysis. So a reciprocal link between two users in Brazil and Qatar would contribute to the statistics of both countries.

Gender Gap. One of the goals of our study was to devise an "Online Gender Gap" score and to see how this relates to the existing offline Gender Gap scores. We therefore followed the same methodology of computing a "gap" score: First, we group the users by country and gender, and calculate the average of the variable for each country-gender group. After having the aggregated value for each country-gender group, we calculate the gender ratio by dividing the female value by the male value, for each country. Differently from the Global Gender Gap score methodology, we do not truncate the ratio at 1, since we want to analyze the trend even when the value is higher for female users, especially as some of our variables, such as the number of followers, exhibited a counter-intuitive trend. Furthermore, for some of our variables such as the Differential Assortativity, it is also not intuitively obvious if a high or a low gender-specific value is desirable and, correspondingly, it is unclear if high or low values should be truncated.

Note that, in line with the Global Gender Gap report, a large "gap value" is actually *desirable* in the sense that it typically indicates gender equality for the variable considered, whereas a very low gap value is undesirable as it indicates that the variable considered is lower for women than for men.

3.1.2 Offline Variables

The Global Gender Gap Index[2] is a benchmark score that captures the gender disparities in each country. It takes into account social variables from four categories (economy, politics, education and health), such as life expectancy, estimated income, literacy rate and number of seats in political roles. The index is built by (1) calculating the female by male ratio of the variables, (2) truncating the ratios at a certain level (1.0 for most variables), (3) calculating subindexes for each one of the four categories (weighted average in relation to the standard deviation) and (4) calculating the unweighted average of the four subindexes to create the overall index. The scores range from 0 (total inequality) to 1.0 (total equality)[3]. For this study we use the 2013 Global Gender Gap report [55].

We also use additional economic variables and demographic information to see if these are linked to online gender gaps. For population and internet penetration information we use information from the Internet World Stats website[4] on internet usage for 2012. The GDP per capita information was collected from the World Bank website[5] and is for 2011. Information for more recent years was missing for some countries which is why we selected data from 2011. These variables will be used and analyzed in Section 3.4.

3.2 Gender Differences Online

Before we link online variables to offline indicators of gender gaps, we first describe how men and women in 73 countries differ in their usage of Google+. Figure 3.1 shows the gender ratio of the variables for each country. We observe that for some variables there is a female predominance (such as for "Reciprocity" and "Clustering Coefficient"), while for others there's a male predominance (such as "Number of followees"). In most cases, the gender predominance is the same across countries, but for some variables ("Number of followers") there are divergences.

3.3 Online Gender Gaps

To test the significance of the difference between female and male values of the variables we conducted a permutation test that does not make assumptions about the distribu-

[2] `http://www.weforum.org/issues/global-gender-gap`
[3] Since their focus is on *gender equality* the ratios are truncated to have at maximum 1.0
[4] `http://www.internetworldstats.com`
[5] `http://www.worldbank.org`

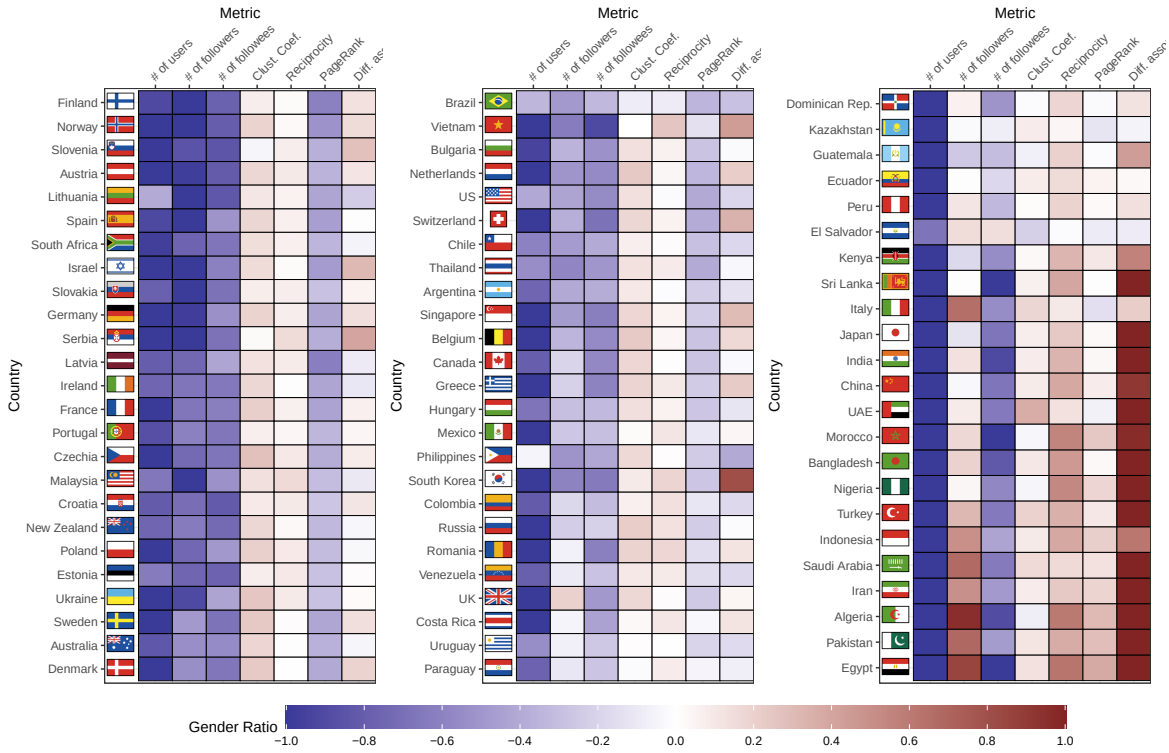

Figure 3.1. A color plot of the logarithm, base 2, of the (female value)/(male value) gender ratio (GR), i.e. $\log_2(\text{GR})$, for the variables in each country. The scale is truncated at -1.0 and 1.0. A value lower than 0 (blue) indicates male predominance, and higher than 0 (red) means female predominance. Countries are sorted according to the average gender ratio considering the 7 metrics, from lowest average (Finland) to highest average (Egypt).

tion of the variables.[6] First, for each country we compute the average of a variable across all female users and compare the value with the one obtained for the male users. Let δ be the observed difference. Then we use the same set of users, but now randomly permute the gender label. The basic idea is to see if the observed difference could have arisen due to random variance or whether it is more systematically linked to the gender of the users. We now calculate the average of the two groups derived from the permutation, and calculate the difference δ_p. We repeat this process 1,000 times to estimate the level of variability of δ_p. Finally, we mark the δ as significant if it was in the bottom/top 0.5% (or 2.5%) of the percentiles of the δ_p. In Table 3.1 we present the significance test result for some variables for a fraction of the countries. In Table 3.3 we present the values for all the countries. For most countries and most variables the difference between female and male is significant.

[6]See Pitman [96] for background information on permutation tests in statistics.

Country	In-degree ♀/♂	Out-degree ♀/♂	Recipr. ♀/♂	Clust. Coeff. ♀/♂	PageRank ♀/♂
United States	34.8/47.1**	20.6/30.3**	0.49/0.50**	0.31/0.28**	2.0e-08/2.6e-08**
Russian Federation	17.7/20.8**	31.0/36.1**	0.45/0.41**	0.38/0.32**	1.5e-08/1.8e-08**
Italy	34.7/22.0	22.7/33.3**	0.51/0.48**	0.33/0.29**	1.8e-08/2.0e-08**
Vietnam	36.9/57.4**	41.7/78.3**	0.41/0.34**	0.29/0.29	1.8e-08/2.0e-08**
Philippines	11.6/16.6**	28.8/38.5**	0.42/0.41	0.40/0.36**	1.4e-08/1.6e-08**
Pakistan	25.4/15.8**	35.3/49.1**	0.40/0.31**	0.32/0.29**	1.6e-08/1.3e-08**
Saudi Arabia	39.3/24.6**	30.2/47.4**	0.37/0.33**	0.29/0.26**	1.7e-08/1.6e-08
Bangladesh	17.4/15.2	30.4/54.1**	0.41/0.30**	0.32/0.30**	1.4e-08/1.3e-08
United Arab Emirates	19.6/18.4	21.4/33.6**	0.46/0.42**	0.28/0.22**	1.7e-08/1.7e-08
Greece	19.0/22.1	26.5/40.3**	0.47/0.44**	0.34/0.30**	1.5e-08/1.8e-08**
Norway	16.8/40.3**	17.6/30.8**	0.57/0.56**	0.35/0.31**	1.7e-08/2.5e-08**
Sri Lanka	20.9/21.1	23.7/50.7**	0.47/0.36**	0.31/0.30*	1.6e-08/1.6e-08
El Salvador	12.8/11.5	31.7/28.7	0.38/0.39	0.21/0.24**	1.4e-08/1.5e-08*
Guatemala	10.1/12.1	21.2/26.2**	0.46/0.40**	0.27/0.29*	1.5e-08/1.5e-08
Slovenia	10.0/18.2**	16.8/30.2**	0.56/0.53**	0.27/0.28	1.6e-08/2.1e-08**

Table 3.1. Significance test results for variables in Google+ for a subset of our 73 countries, ranked in descending order of the number of users. The value on the left is the average female value and the value on the right is the average male value, followed by the significance result ('*' is 95% significant, '**' is 99% significant). The full list of results can be found in Table 3.3.

3.4 Linking Online and Offline Gender Gaps

Whereas the previous section looked exclusively at online gender differences, here we focus on linking online and offline gender gaps across 73 countries.

Figure 3.2 shows the linear regression between online variables and the Global Gender Gap scores. GR stands for Gender Ratio (female divided by male value). We observe that the gap score for the number of users is positively correlated with the gender gap score. Countries with a roughly equal number of male and female users online tend to score better (= higher) for the offline gap scores. Surprisingly we also find that the number of followers and other measures of "status" are negatively correlated for both networks. For example, Pakistan has an offline Gender Gap score of 0.546 (with 1.0 indicating equality) but, at the same time, women who are online in Pakistan have on average (and in median) more followers than their male counterparts. We discuss potential reasons later in the paper.

The two plots in the right column of Figure 3.2 show the linear regression plots of the assortativity variables in Google+. When we analyze the Differential assortativity we observe that most countries, clustered together on the dashed line, have similar values for female and male, meaning that the level of gender assortativity is the same

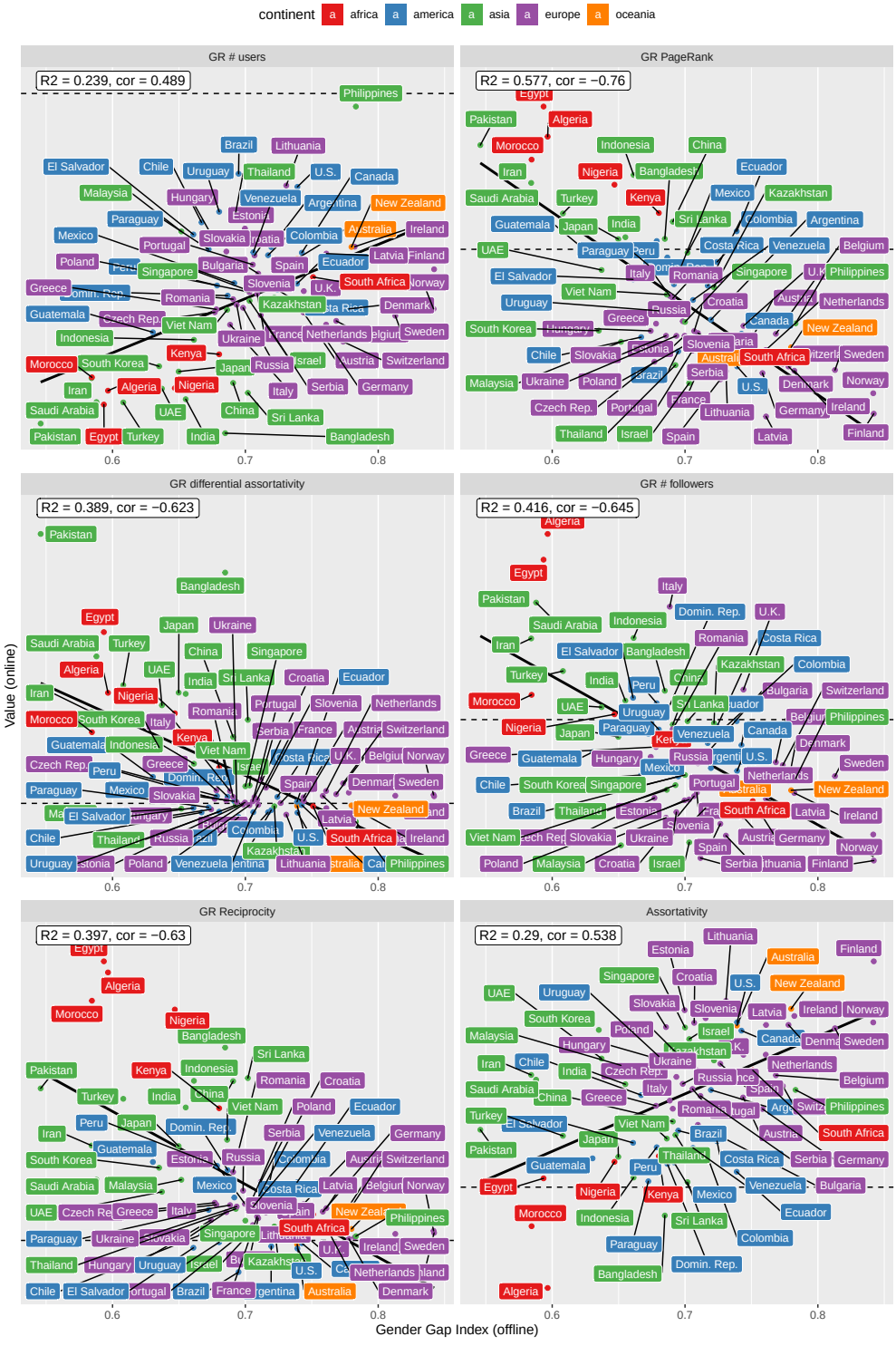

Figure 3.2. Linear regression and correlation between online social network metrics and the Global Gender Gap score. GR stands for Gender Ratio (female by male value). The p-values for the correlation were all lower than 0.01.

for women and men. On the other hand, in countries with a low Gender Gap score there's a female predominance, meaning that women in these countries connect much more among themselves than expected when compared to men. This could be seen as an indication of women "shying away" from cross-gender linkage in such countries. When we analyze not the gap but the actual assortativity of a country we observe a positive correlation with the gap score, meaning that in countries with higher Gender Gap score (= little inequality), there is higher assortativity (= more within-gender linkage). We discuss potential hypotheses explaining this arguably surprising finding in Section 3.6.

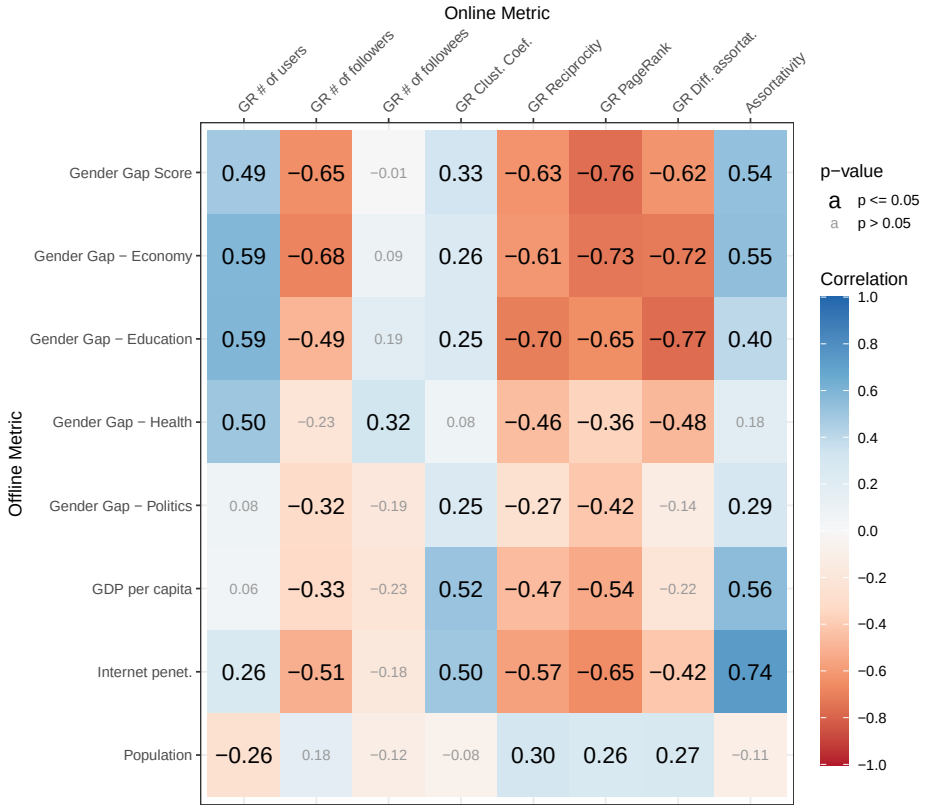

Figure 3.3. Correlation between offline variables and the ratio of online variables of the countries. GR stands for Gender Ratio (female by male value). The not-significant correlations ($p > 0.05$) are labeled in smaller font and light grey color.

Figure 3.3 presents the matrix of correlation between the online and offline variables, essentially summarizing the linear regression fits from Figure 3.2 and adding more variables. As in Figure 3.2, the Gender Gap Score is positively correlated with the gender gap of the number of users in Google+, and, surprisingly, negatively correlated with the gap of the number of followers, reciprocity and PageRank. In terms

of assortativity, there is a negative correlation for differential assortativity, meaning that female users connect more among themselves in countries with a low Gap score, while the actual assortativity of the network is positively correlated, implying more segregation in countries with high Gender Gap score.

3.5 Online Gender Ratio Map

In our final analysis we focus on identifying clusters of countries that have similar online behaviour regarding gender gap metrics.

We use the online gender ratio values calculated with online information from Google+ to create a bi-dimensional scatter plots of the countries. For each country we create a vector of 8 dimensions, each dimension being one of the online metrics we studied here. We combine all the country vectors, resulting in a 73 by 8 matrix.

Having the gender ratio matrix, we need to apply a dimensionality reduction algorithm to obtain two dimensions and create a 2D plot. We experimented using three common techniques: PCA [94], t-SNE [117], and UMAP [82]. The UMAP showed the best results is terms of interpretability of the clusters, and is also very flexible in terms of parameterization. We use $\texttt{number of neighbours} = 10, \texttt{minimum distance} = 1, \texttt{spread} = 5$. The resulting gender ratio country map is presented in Figure 3.4.

We observe that there are basically three clusters of countries. First, in the upper-left part of the map, there are Asian and African countries. It is interesting to observe that a considerable part of these countries has a majority Muslim population [95] (Turkey, Egypt, Algeria, Morocco, Iran, United Arab Emirates, Saudi Arabia, Pakistan, and Bangladesh), but there are also other religions such as Buddhism (Sri Lanka) and even irreligious (Japan and China).

The second cluster is positioned in the top-right part of the map. Most of the American countries are located in this cluster, even though having some European countries such as Italy, United Kingdom and Russia, and also Asian countries (Kazakhstan and Philippines). In terms of religion, we notice that most of the countries has a major Christian population [95] regardless of their continent. There are Latin American countries such as Ecuador, Peru, Colombia and Brazil, Southeast European countries such as Bulgaria and Romania, and even a Southeast Asian country (Philippines).

The third cluster (in the bottom-right part of the map) is the bigger one in terms of the number of countries, and has a majority of European countries. It also contains the two only countries from Oceania (Australia and New Zealand), some Asian countries (Singapore, Malaysia, and Israel), and South Africa as the only Asian country.

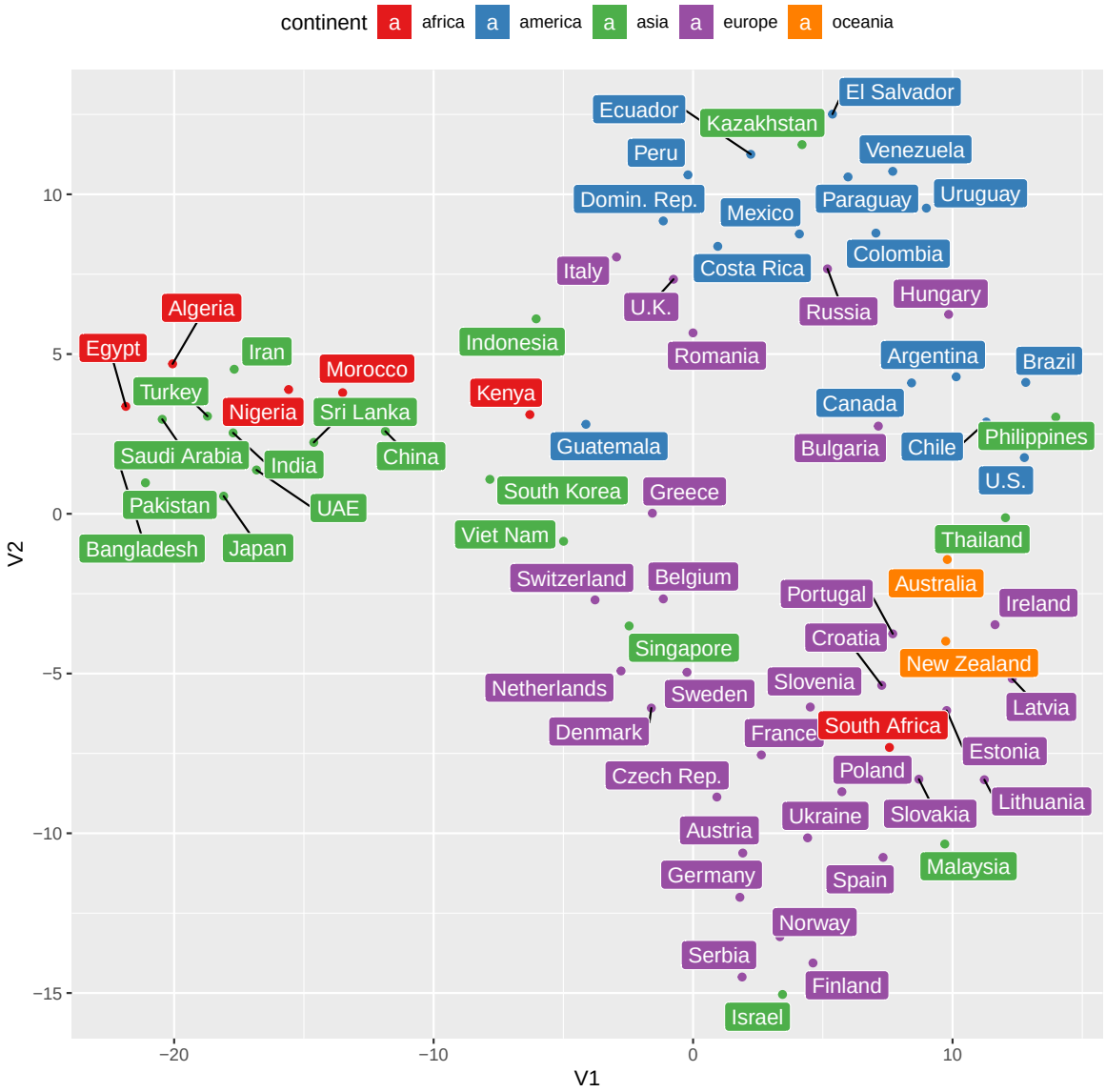

Figure 3.4. Online Gender Ratio country map. Dimensions were generated using the UMAP technique.

Regarding religion, we notice that there is a high diversity, having Israel (majority of Jewish population), Spain (majority of Christian population), and Estonia (majority of irreligious population).

The online Gender Ratio map shows that there are indeed patterns of countries with similar online behaviour regarding gender gap. These patterns can not be easily explained by single characteristics such as geographic location or religion. Further investigation is necessary, to analyze other social-economical and cultural traits that might be influencing these online patterns.

3.6 Discussion

One of our main motivation for this work was to see if online data could be used to derive global indicators of gender inequality and whether these indicators were in some sense "grounded" in that they are linked to existing indicators. Our findings indicate that this indeed the case.

Surprisingly, the directionality of important indicators was *opposite* from what we had expected. Concretely, we found that all indicators of gaps in online social status such as the average number of followers, or the Pagerank on Google+ all had noticeable *negative* correlations (.65 and -.76 correspondingly) with the aggregated offline gender gap score. For example in Pakistan, with a gender gap score of 0.55, indicating a large inequality, we found that women have on average 50% more followers on Google+ than men. Note that the number of followers is typically heavy-tailed [70] and for such distributions it is known that the observed average will increase as the sample size increases[7]. As we have fewer women and men for countries where we observe these effects, the actual effect might hence be even stronger. We also mention that we observed the same effect by looking at medians, rather than averages, indicating a robust result not caused by outliers.

Our current hypothesis is that this unexpected result might be due to the so-called "Jackie Robinson Effect"[8]. Jackie Robinson was a baseball player who became the first African-American to play in Major League Baseball in the modern era. If he had been only good, rather than great, it is unlikely that he would have been given a chance to play rather than a slightly less talented white alternative. Similarly, one might imagine that women that are online in countries where women have more limited online access compared to men must be extraordinary to begin with. In a similar vein it was found that female politicians perform better than their male counter-parts as doing just as well would not suffice to "make it" [6].

The effect above might also be linked to our observation of more within-gender linkage for countries such as Finland or Norway, compared to Egypt or Pakistan. Other potential explanations for this observation could be acts of online "stalking" or "staring" where women attract follow links from men, causing more cross-gender linkage. This latter hypothesis is also consistent with our observation that in countries with more offline gender inequality women have a stronger tendency for within-gender linkage than men, potentially indicative of shying away from cross-gender linkage.

[7] See, e.g., `http://en.wikipedia.org/wiki/Pareto_distribution` which has an infinite mean when $\alpha \leq 1$.

[8] `http://en.wikipedia.org/wiki/Jackie_Robinson`

Of course, our current data set and methodology are not perfect. Clearly, our user set is not representative of the overall population. Generally, we expect people over a higher social status to be overrepresented in our data. But even the fact that for Pakistan we find about 8 times as many male Google+ users as female ones is in itself a signal. Also note that for certain applications the selection bias might be irrelevant. If, for example, the main purpose of using online data is to have a low-cost and real-time alternative to compute the offline gender gap index then as long as it works, despite the selection bias, the selection bias itself becomes irrelevant. As a comparison, if it is possible to accurately predict current levels of flu activity from social media data then there is no reason to question this approach, assuming that the prediction remains valid as the online population continues to change [7, 71, 32].

The example of monitoring flu activity also points to another limitation of our study: the use of only one data source. For flu monitoring using online data, Google Flu Trends [47] is the de-facto standard and baseline to beat. Recently, its use as a figurehead has however been questioned [73]. Still, it seems promising to look at, say, the relative search volume of topics associated with gender roles to see if their search volume could be indicative of gender gaps. Additionally, gender differences on comments on national, political sites could be indicators for political engagement.

Another big limitation is our decision to ignore the content/topics that are discussed. The main reasons for this are (i) technical difficulties when dealing with content analysis for dozens of different languages and character sets, in particular if the results need to be comparable across countries, and (ii) the emphasis of existing offline indices on "hard data" rather than sentiments or more qualitative analysis. Still, it seems valuable to look at the topics discussed by, say, men and women in Mali to get better insights into their lived online experiences. An idea for future work in this topic is to focus on a limited set of countries and languages and study topical differences in depth. Integrating content could also lead to an improvement of the already decent fit between a combination of online indicators and the offline gender gap scores. Finally, it could provide hypotheses for the root causes of the differences we observe.

For future work, it would also be interesting to study the temporal evolution of the gender gaps. The Global Gender Gap Report already does this, being published anually. With these analysis it would be possible to verify rather the online gaps are being closed, and also verify if the countries that are improving offline are the same improving online.

Ultimately, of course, the goal is not just to describe and quantify gender gaps but to close these gaps. Here, a large amount of responsibility undoubtedly lies with politicians and people in positions of power. As good policy making needs to be linked

to quantifying the progress made, and there is a necessity to observe the impact of new policies, measurement efforts are a valid objective in their own right. However, it is well worthwhile thinking about how social media and online social networks could in itself be used as a tool to facilitate the process of closing the gap, rather than as a mere data source. It might for example be possible to automatically strengthen the social capital of underprivileged women or, if nothing else, it could be used as communication channel to support the cause of gender equality. Our contribution to support this cause in this work is to raise awareness of gender gap in online environments, and at the same time to provide a methodology that is capable of measuring gender gaps.

3.7 Conclusion

We presented a large-scale study of gender differences and gender gaps around the world in Google+. Our analysis is based on 17,831,006 users from 73 countries with an identified gender and, to the best of our knowledge, is the first study that links online indicators of gender inequality to existing offline indicators.

Our main contribution is two-fold. First, we describe gender differences along a number of dimensions. Such insights are valuable both as a starting point for in-depth studies on identifying the root causes of these differences, but also when it comes to designing gender-aware systems. Second, we show how applying existing offline methodology for quantifying gender gaps can be applied to online data and that there is a respectable match in form of a 0.8 correlation between online and offline measures, across 73 countries.

Looking at individual variables we also find surprising patterns such as a tendency for women in less developed countries with larger gender differences to have a *higher* social status online as measured in terms of number of followers or PageRank. We hypothesize the existence of an underlying "Jackie Robinson Effect" where women who decided to go online in a country such a Pakistan are likely to be more self-confident and tech-savvy than random male counterparts. Such an effect might also be linked to the fact that we observe a *higher* within-gender link assortativity for countries with *less* offline gender inequality, though alternative explanations include men "stalking" women online.

As more and more economic activity, education, and political engagement happens online we are convinced that, ultimately, quantifying gender inequality has to crucially take into account online activity.

Country		# users			Country		# users		
Code	Name	Female	Male	Total	Code	Name	Female	Male	Total
US	United States	2,186,509	2,910,470	5,096,979	KR	South Korea	16,570	60,696	77,266
IN	India	363,956	1,964,070	2,328,026	SE	Sweden	22,342	54,815	77,157
BR	Brazil	563,173	716,455	1,279,628	BE	Belgium	21,755	55,223	76,978
GB	United Kingdom	210,801	445,343	656,144	AE	United Arab Emirates	12,250	57,399	69,649
ID	Indonesia	136,013	396,028	532,041	DK	Denmark	20,219	47,470	67,689
RU	Russian Federation	140,024	326,464	466,488	CZ	Czech Republic	19,409	46,548	65,957
CA	Canada	147,247	255,750	402,997	SG	Singapore	20,798	43,515	64,313
MX	Mexico	129,566	261,958	391,524	FI	Finland	21,831	41,072	62,903
DE	Germany	98,500	275,813	374,313	GR	Greece	17,578	41,393	58,971
ES	Spain	116,997	221,343	338,340	IE	Ireland	21,277	35,959	57,236
IT	Italy	87,028	226,777	313,805	RS	Serbia	16,458	40,241	56,699
FR	France	98,628	211,602	310,230	CH	Switzerland	14,255	42,085	56,340
JP	Japan	57,234	221,049	278,283	AT	Austria	15,487	37,185	52,672
CN	China	45,551	199,300	244,851	NO	Norway	15,246	35,795	51,041
AU	Australia	87,605	156,493	244,098	IL	Israel	15,101	33,752	48,853
VN	Viet Nam	64,539	152,459	216,998	EC	Ecuador	15,611	31,654	47,265
TH	Thailand	80,655	117,904	198,559	NZ	New Zealand	17,462	29,547	47,009
AR	Argentina	68,877	116,617	185,494	SK	Slovakia	16,061	27,749	43,810
TR	Turkey	25,974	147,023	172,997	LK	Sri Lanka	7,186	35,540	42,726
CO	Colombia	62,590	110,004	172,594	BG	Bulgaria	13,136	25,260	38,396
PH	Philippines	78,760	81,601	160,361	HR	Croatia	13,612	23,944	37,556
MY	Malaysia	60,607	95,842	156,449	MA	Morocco	7,170	29,434	36,604
UA	Ukraine	46,132	105,582	151,714	DO	Dominican Republic	10,750	23,303	34,053
PL	Poland	48,381	102,802	151,183	SV	El Salvador	11,891	19,049	30,940
NL	Netherlands	40,074	104,336	144,410	DZ	Algeria	5,176	24,887	30,063
PK	Pakistan	15,420	128,150	143,570	CR	Costa Rica	9,632	20,186	29,818
IR	Iran	27,153	112,444	139,597	KE	Kenya	6,868	22,522	29,390
CL	Chile	53,286	81,165	134,451	NG	Nigeria	5,050	23,523	28,573
EG	Egypt	19,414	113,495	132,909	GT	Guatemala	7,342	20,189	27,531
ZA	South Africa	34,153	66,871	101,024	UY	Uruguay	9,966	14,552	24,518
SA	Saudi Arabia	15,173	85,416	100,589	LT	Lithuania	10,416	13,801	24,217
PE	Peru	32,296	66,141	98,437	KZ	Kazakhstan	5,727	12,555	18,282
RO	Romania	28,907	63,982	92,889	PY	Paraguay	6,273	10,730	17,003
PT	Portugal	32,218	59,238	91,456	SI	Slovenia	5,644	11,269	16,913
VE	Venezuela	32,623	56,556	89,179	LV	Latvia	5,722	9,979	15,701
BD	Bangladesh	7,029	74,221	81,250	EE	Estonia	5,337	8,337	13,674
HU	Hungary	30,525	48,858	79,383					

Table 3.2. List of countries with their respective 2-letter country codes and the total number of female and male users. We select only countries with at least 5,000 females and males.

Country	In-degree ♀/♂	Out-degree ♀/♂	Recipr. ♀/♂	Clust. Coeff. ♀/♂	PageRank ♀/♂
United States	34.8/47.1**	20.6/30.3**	0.49/0.50**	0.31/0.28**	2.0e-08/2.6e-08**
India	25.5/23.2	20.3/38.2**	0.52/0.41**	0.25/0.23**	2.0e-08/2.0e-08
Brazil	20.4/28.7**	38.0/48.0**	0.37/0.39**	0.16/0.17**	1.7e-08/2.2e-08**
United Kingdom	30.9/26.8	20.5/28.9**	0.47/0.46**	0.33/0.29**	1.8e-08/2.1e-08**
Indonesia	25.0/17.7**	39.5/53.4**	0.43/0.33**	0.36/0.34**	1.9e-08/1.6e-08**
Russian Federation	17.7/20.8**	31.0/36.1**	0.45/0.41**	0.38/0.32**	1.5e-08/1.8e-08**
Canada	33.9/38.9	19.6/29.1**	0.48/0.48	0.31/0.28**	1.8e-08/2.2e-08**
Mexico	10.5/12.6**	22.8/28.0**	0.45/0.41**	0.28/0.27*	1.5e-08/1.6e-08**
Germany	21.5/42.2**	21.9/31.6**	0.49/0.47**	0.35/0.31**	1.6e-08/2.1e-08**
Spain	13.7/29.2**	20.4/29.1**	0.50/0.47**	0.32/0.29**	1.6e-08/2.2e-08**
Italy	34.7/22.0	22.7/33.3**	0.51/0.48**	0.33/0.29**	1.8e-08/2.0e-08**
France	15.6/24.7**	19.8/30.5**	0.49/0.46**	0.33/0.29**	1.6e-08/2.1e-08**
Japan	32.0/35.0	30.8/49.1**	0.44/0.37**	0.34/0.32**	1.9e-08/1.9e-08
China	45.1/46.3	48.0/76.5**	0.41/0.31**	0.27/0.25**	1.9e-08/1.8e-08
Australia	14.8/21.5**	18.5/27.2**	0.48/0.48	0.33/0.29**	1.5e-08/2.0e-08**
Viet Nam	36.9/57.4**	41.7/78.3**	0.41/0.34**	0.29/0.29	1.8e-08/2.0e-08**
Thailand	19.4/29.1**	34.0/48.2**	0.41/0.39**	0.34/0.31**	1.6e-08/2.2e-08**
Argentina	13.4/17.8**	22.7/29.7**	0.43/0.43*	0.29/0.27**	1.6e-08/1.9e-08**
Turkey	18.8/15.1**	29.0/45.7**	0.46/0.36**	0.32/0.28**	1.5e-08/1.4e-08
Colombia	9.6/10.9**	24.8/31.0**	0.44/0.40**	0.28/0.27**	1.4e-08/1.6e-08**
Philippines	11.6/16.6**	28.8/38.5**	0.42/0.41	0.40/0.36**	1.4e-08/1.6e-08**
Malaysia	11.8/32.7**	26.5/38.1**	0.45/0.40**	0.33/0.30**	1.4e-08/1.8e-08**
Ukraine	20.1/37.9**	31.8/43.0**	0.48/0.45**	0.37/0.31**	1.6e-08/1.9e-08**
Poland	8.1/13.6**	17.0/23.9**	0.53/0.50**	0.37/0.32**	1.5e-08/1.8e-08**
Netherlands	15.7/22.3**	18.6/27.5**	0.51/0.50**	0.33/0.28**	1.6e-08/2.1e-08**
Pakistan	25.4/15.8**	35.3/49.1**	0.40/0.31**	0.32/0.29**	1.6e-08/1.3e-08**
Iran	50.2/35.6	34.9/49.0**	0.46/0.39**	0.30/0.29**	1.9e-08/1.7e-08
Chile	9.7/13.5**	17.7/23.4**	0.50/0.50*	0.27/0.26**	1.6e-08/2.0e-08**
Egypt	34.2/18.9**	30.3/62.4**	0.38/0.25**	0.31/0.28**	1.7e-08/1.3e-08**
South Africa	10.5/17.9**	19.4/31.0**	0.45/0.42**	0.29/0.26**	1.4e-08/1.8e-08**
Saudi Arabia	39.3/24.6**	30.2/47.4**	0.37/0.33**	0.29/0.26**	1.7e-08/1.6e-08
Peru	12.2/11.3	27.7/34.9**	0.41/0.36**	0.28/0.28	1.5e-08/1.5e-08
Romania	22.8/24.0	34.4/52.7**	0.43/0.38**	0.35/0.31**	1.5e-08/1.7e-08**
Portugal	13.3/20.4**	22.6/35.9**	0.47/0.46**	0.27/0.26**	1.5e-08/1.9e-08**
Venezuela	13.5/14.4	28.6/34.9**	0.42/0.39**	0.28/0.26**	1.5e-08/1.7e-08**
Bangladesh	17.4/15.2	30.4/54.1**	0.41/0.30**	0.32/0.30**	1.4e-08/1.3e-08
Hungary	10.0/12.4**	17.9/22.5**	0.55/0.53**	0.34/0.31**	1.5e-08/1.8e-08**
South Korea	17.7/26.8**	26.8/42.1**	0.48/0.42**	0.33/0.31**	1.6e-08/2.0e-08**
Sweden	16.8/23.6**	17.6/28.2**	0.58/0.57*	0.37/0.31**	1.7e-08/2.3e-08**
Belgium	13.8/17.6*	17.9/26.4**	0.50/0.49**	0.34/0.29**	1.6e-08/1.9e-08**
United Arab Emirates	19.6/18.4	21.4/33.6**	0.46/0.42**	0.28/0.22**	1.7e-08/1.7e-08
Denmark	12.7/18.4**	14.8/23.5**	0.57/0.57	0.34/0.29**	1.7e-08/2.2e-08**
Czech Republic	12.2/20.2**	17.0/27.1**	0.56/0.52**	0.38/0.31**	1.6e-08/2.1e-08**
Singapore	14.8/20.6**	19.5/30.0**	0.51/0.49**	0.27/0.24**	1.7e-08/2.1e-08**
Finland	13.4/47.0**	13.7/23.5**	0.60/0.59*	0.37/0.35**	1.6e-08/2.5e-08**
Greece	19.0/22.1	26.5/40.3**	0.47/0.44**	0.34/0.30**	1.5e-08/1.8e-08**
Ireland	13.9/22.2**	17.3/27.4**	0.49/0.48	0.35/0.31**	1.6e-08/2.1e-08**
Serbia	13.9/46.9*	19.8/31.8**	0.53/0.47**	0.31/0.30	1.5e-08/2.0e-08**
Switzerland	22.4/29.2	20.6/33.3**	0.50/0.48**	0.31/0.28**	1.7e-08/2.2e-08**
Austria	14.2/27.9**	17.9/31.4**	0.52/0.49**	0.37/0.33**	1.5e-08/1.9e-08**
Norway	16.8/40.3**	17.6/30.8**	0.57/0.56**	0.35/0.31**	1.7e-08/2.5e-08**
Israel	23.2/61.5	24.5/37.4**	0.50/0.49	0.26/0.23**	1.8e-08/2.5e-08**
Ecuador	8.5/8.5	27.6/31.4**	0.40/0.36**	0.32/0.31**	1.4e-08/1.3e-08
New Zealand	14.3/22.4**	16.7/27.8**	0.51/0.50**	0.33/0.29**	1.6e-08/2.0e-08**
Slovakia	6.4/12.8**	13.1/21.1**	0.61/0.58**	0.32/0.30**	1.6e-08/2.0e-08**
Sri Lanka	20.9/21.1	23.7/50.7**	0.47/0.36**	0.31/0.30*	1.6e-08/1.6e-08
Bulgaria	14.9/19.1**	25.2/36.2**	0.48/0.46**	0.34/0.31**	1.5e-08/1.8e-08**
Croatia	8.9/14.5**	15.0/26.4**	0.54/0.50**	0.32/0.30**	1.4e-08/1.7e-08**
Morocco	20.7/18.3	27.1/57.9**	0.44/0.30**	0.25/0.26	1.7e-08/1.4e-08**
Dominican Republic	16.7/16.0	27.5/39.3**	0.43/0.38**	0.27/0.27	1.6e-08/1.7e-08
El Salvador	12.8/11.5	31.7/28.7	0.38/0.39	0.21/0.24**	1.4e-08/1.5e-08*
Algeria	20.7/10.6**	27.6/51.4**	0.34/0.22**	0.25/0.27	1.3e-08/1.0e-08**
Costa Rica	14.6/15.1	20.3/27.6**	0.50/0.46**	0.27/0.27	1.7e-08/1.8e-08
Kenya	13.1/14.8	28.6/42.0**	0.42/0.34**	0.27/0.26	1.6e-08/1.5e-08
Nigeria	8.7/8.4	31.9/47.7**	0.31/0.21**	0.26/0.27	1.2e-08/1.1e-08*
Guatemala	10.1/12.1	21.2/26.2**	0.46/0.40**	0.27/0.29*	1.5e-08/1.5e-08
Uruguay	13.2/13.9	23.9/28.6*	0.46/0.46	0.27/0.27	1.5e-08/1.7e-08**
Lithuania	7.9/19.3**	19.3/34.5**	0.51/0.49**	0.30/0.28**	1.5e-08/2.0e-08**
Kazakhstan	16.5/16.8	33.6/35.6	0.38/0.37	0.33/0.32	1.4e-08/1.5e-08
Paraguay	16.8/18.2	28.1/34.0**	0.45/0.42**	0.23/0.23	1.8e-08/1.8e-08
Slovenia	10.0/18.2**	16.8/30.2**	0.56/0.53**	0.27/0.28	1.6e-08/2.1e-08**
Latvia	11.8/19.7**	26.2/35.3*	0.51/0.48**	0.34/0.31**	1.5e-08/2.3e-08**
Estonia	8.9/15.0**	15.0/25.7**	0.54/0.51**	0.26/0.25	1.6e-08/1.9e-08**

Table 3.3. Significance test results for variables in Google+ for our 73 countries, ranked in descending order of the number of users. The value on the left is the

Chapter 4

International Online Values with Word Embeddings

Human values are one of the key characteristics that influence the culture of social groups. They are beliefs used by a person to make decisions related to life and make actions, influencing the mode of conduct and way of thinking of individuals. The importance of God in life, whether abortion is justifiable, or if it is important to be rich, are examples of questions that people will have different visions, being influenced by the cultures they have contact with. When formulating a conception for human values, Rokeachz states that "[...] human values will be manifested in virtually all phenomena that social scientists might consider worth investigating and understanding". Differently from traditional methodologies that use surveys to measure values, we propose a technique that explores the phenomena of writing texts online to measure values on a global scale.

Twitter originated as a simple microblogging service, where people could post small texts of 140 characters and follow other users to receive their posts. But since its release in 2006, Twitter has not only increased its number of users (126 million daily active users in 2019 [107]) and released new features (retweet, reply, embedded images and video, URL preview, polls, etc), but also diversified its utilization. If people used to simply share texts and news, Twitter is now a multi-purpose online environment, where entities like companies, politicians, celebrities, and robots publish content and interact with each other, having their own values and goals in the platform. Additionally, Twitter is utilized by people and entities from all over the globe, having its interface and personalized content (e.g. trending topics) available in more than 47 languages[1]. These characteristics make Twitter an interesting place for studying online worldwide

[1] Acessing `https://twitter.com/settings/language` on 06-Oct-2019

social phenomena.

In the field of natural language processing, word embedding algorithms emerged as better alternatives for creating mathematical representations of textual datasets [84, 74, 49]. Compared to classical methods (e.g. one-hot encoding) they create models with a smaller number of dimensions while capturing the semantics of the language. Since the word embeddings are trained with texts written by humans, they are prone to capture and propagate social biases, such as gender stereotypes [20]. There is also criticism arguing that analogies might not be the most adequate tool to measure and identify bias [89].

Inspired by the psychological test IAT [51], the WEAT [24] is a technique that measures implicit associations between words in the model, allowing one to identify potentially harmful biases. We conjecture that word embeddings can reflect not only biases and stereotypes but also *human values*.

In this work, we develop, describe and test a methodology to measure human values manifested in written texts for different countries, applied to an international online community. We use a dataset of 1.7 billion tweets of the year 2014, identify the location of the tweets, and train a word embedding model for 59 countries. The intrinsic semantics of these textual models are used to calculate several *online values* for the countries, which are compared to their respective *offline value* from the World Values Survey. We show that some online values are indeed correlated with their corresponding offline value. We also present the online cultural map, which is a bi-dimensional scatter-plot of the countries created by a factor analysis of the online values.

The rest of the chapter is organized as follows. Section 4.1 presents definitions and descriptions for the terms and platforms we approach in our work. Next, in Section 4.2 we describe our methodology for collecting and training our datasets. Then, in Section 4.3 we show our results for the inquiries, the correlations with offline data, as well as the online cultural map. Finally, Section 4.4 concludes our work with some interesting discussion and future work.

4.1 Background

Before describing our methodology, it is important to describe and define some key aspects of our research. In this section we define culture and values, describe the Twitter platform with details and also the World Values Survey.

4.1.1 Conceptualizing Culture and Values

The task of defining "culture" is very difficult, many attempts, propositions, and interpretations of the term were made by several authors. By doing a literature review starting from 19th century, Avruch and of Peace divide the utilization of the term into three categories [10]: (1) culture as a special intellectual characteristic, which only a portion of a social group has, (2) culture as a characteristic that everyone has, but that can be classified in an evolutionary spectrum (from 'savagery' to 'civilization') and (3) culture as unique characteristics of different and varied peoples or societies, rejecting the judgment present in the other views. Our understanding of culture is aligned with the third view, and we employ a comparative approach rather than a judgmental one. As shown by Spencer-Oatey, many definitions for "culture" have been proposed [111], and we present here one of them, written by the same author:

> Culture is a fuzzy set of basic assumptions and values, orientations to life, beliefs, policies, procedures and behavioural conventions that are shared by a group of people, and that influence (but do not determine) each member's behavior and his/her interpretations of the 'meaning' of other people's behavior. [110]

In this definition, we notice that *values* are one of the aspects that compose the culture of a social group. More specifically, by using the framework of characteristics of culture proposed by Spencer-Oatey, "culture is manifested at different layers of depth" [111], being *espoused* values the second layer. In this depth, the focus is on what people *report* when questioned about their behavior. In our study, this manifestation will happen with written text on Twitter. We present here the definition of values written by Macionis:

> Values [are] culturally defined standards that people use to decide what is desirable, good, and beautiful and that serve as broad guidelines for social living. People who share a culture use values to make choices about how to live. [76]

There are four authors known for their relevant studies regarding cultural human values: Milton Rokeach (social psychologist), Shalom Schwartz (social psychologist), Geert Hofstede (social psychologist) and Ronald Inglehart (political scientist). Rokeach presents not only a conceptualization for human values but also a classification system (Rokeach Value Survey) for measuring them, consisting of a rank-order methodology of 36 values, organized in two groups of equal size [103]. Following, Schwartz develops the

Theory of Basic Human Values [106], directly inspired by Rokeach's work, consisting of 10 universal values and measured by applying the Schwartz Value Survey. Starting from 1967 and being developed through the years, the Hofstede's cultural dimensions theory [58] has a methodology for measuring values consisting of 6 cultural dimensions, applied on employees of IBM worldwide. Finally, we highlight the work of Inglehart, who developed the World Values Survey (WVS) [61], a questionnaire methodology to measure several attitude items, followed by a factor analysis that identifies two dimensions of values [62]. The WVS is explained in more details in Section 4.1.3.

It is not our goal to compare these human values theories and the corresponding methodologies and systems of measurement. There are already studies [85, 39] that discuss in details the similarities, differences, and advantages of each technique. We choose the World Values Surveys as a source of comparison and inspiration for designing our method due to its abundance and availability of data, and the coverage of a considerable number of countries. It is important to notice that our goal is not to emulate the WVS but to use it for comparison.

4.1.2 Twitter Platform

Twitter is an online social service where people communicate by publishing 140 character messages[2] known as *tweets*. The platform describes itself as[3]:

> what's happening in the world and what people are talking about right now.

The development of Twitter started in March 2006[4], and it was publicly released in July 2006, earlier named *Twttr* [8]. As of 2018 Twitter has 326 million active users [52], having personalized content for 239 countries and interface available in 47 languages[5]. It has a diverse user base, being accessed by celebrities, authorities, politicians, and even head of states.

Twitter is commonly described as a *microblogging* service. This definition is limited, considering the diverse content present in the tweets and the myriad ways and purposes for using and interacting with the platform. For instance, a report [3] from 2009 analyzed 2,000 tweets and classified them into six categories, and observed that 40% were "pointless babble" and 38% "conversational", followed by "pass-along

[2]From September 2017, Twitter increased the limit to 280 characters

[3]`https://about.twitter.com/en_us.html`

[4]First published tweet: `https://twitter.com/jack/status/20`

[5]The number of countries and languages was obtained by acessing `https://twitter.com/settings/account` in March 2019

balue", "self-promotion", "spam", and "news". It is also important to note that Twitter allows the publication of several media, such as images, gif, and videos, making it a complex multimedia online environment. It is known that Twitter has several purposes of utilization, ranging from committing crimes [16] and terrorism [28] to predicting disease outbreaks [80] and reporting eartquakes [104].

Users can *follow* other users and build a list of *followees* (friends). Concomitantly, users can be followed by others, building a list of *followers*. The *timeline* is the list of tweets published by the user himself, and the *feed* of a user is the list of tweets published by its followees, being the principal way of consuming tweets. Traditionally, the feed was presented in reverse chronological order (i.e. most recent tweet in the top). Since February 2016, Twitter added a feature to show "the best tweets first" [87]. After this, Twitter is gradually adding more features and complexity to the feed, showing even tweets from people you do not follow (e.g. a tweet liked by a friend).

Additionally to *tweeting* and reading tweets, there are other actions and features available. One can *like* a tweet (represented by a pink heart symbol), which will not only increase the like count of that tweet, but also add it to the user's "Liked tweets" list[6]. Another action is the *retweet*, which will republish a tweet in the user's timeline[7]. Finally, there is the action called *reply*, which will respond a certain tweet with another tweet. The reply allows the users to easily keep track of the messages, making Twitter a good environment for conversations. More recently, Twitter improved the reply feature to encourage the creation of *threads*, which are ordered lists of tweets of the same author.

Other than the text of the tweet (including emojis), there are two important textual features present in Twitter: the "mention" and the "hashtag". The *mention* is the act of explicitly mentioning someone[8] in the tweet, by writing `@username`, being "username" the screen name (i.e. login name, handle) of the particular account. It is possible to mention more than one account, as long as it fits the tweet length. The effect of a mention is generating a notification for all the accounts being mentioned, and generating a hyperlink in the text of the tweet linking to the profile of the corresponding account.

The *hashtag* is a marked word inside the text of the tweet, by writing `#word`,

[6]In earlier versions of the platform, the action was called *favoriting*, represented by a yellow star symbol

[7]In earlier versions of the platform, when the retweet feature was not available, users explicitly retweeted by appending "RT" followed by the user name of the original author of tweet before the original text of the tweet.

[8]It is important to notice that an account in twitter can represent not only a person, but any entity, like a company, an event or a website. For simplicity, "account" and "person" will be used interchangeably to refer to any Twitter account

being "word" the intended word or phrase to be highlighted[9]. Similarly to mentions, a single tweet can contain multiple hashtags. The direct effect of a hashtag is generating a hyperlink in the text of the tweet, linking to a page with the search results for that particular hashtag (.i.e other tweets using the same hashtag). The hashtag concept emerged in Twitter, but during the following years was adopted by major online social platforms such as Facebook and Instagram.

4.1.3 World Values Survey

The *World Values Survey* (WVS) is a project that has the goal of researching values and beliefs from people all over the world [118]. It started in 1981, and since then regular national surveys are conducted in more than 100 countries. The answers for the survey questions are analyzed and compared across time, and can give support for studies about several social, political and economic topics, such as globalization, tolerance for ethnic minorities, and gender equality.

The surveys are not conducted in the same year for all the countries. They are organized in *waves*, consisting of all surveys in a certain period of time. For instance, Wave 1 represents 1981-1984, and Wave 6 (the most recent available) represents 2010-2014. For our work we will use Wave 6 (W6) [63].

Its methodology consists of interviewing a representative sample of individuals in each country. A master questionnaire is developed in English, then translated into the appropriate national languages where the survey is applied. Following each wave, the researchers deliberate about the questions, either to remove or add new questions.

The content of the questionnaire is diverse, and the questions account for several aspects of the individual values and beliefs. Examples of questions are "How important is God in your life?", "How proud are you to be [nationality] (e.g. American, Brazilian, etc)?", and "I see myself as someone who is reserved". The answer for the questions may vary, but they are commonly designed as a ten-level Likert scale [75], where the respondent can either strongly disagree (1) or strongly agree (10) with the statement of the question.

The data for WVS is freely available in its official website[10]. Besides the questionnaire, codebook and results of the surveys by country, it provides the actual survey replies in a tabular format (columns are questions and rows are individuals). In our work we will handle the WVS in an aggregated form rather than individual. For each

[9] Even though a hashtag does not allow space between words, there are a couple of techniques used to use a phrase or term as hashtag. For example, `#ThisIsAPhrase` or `#this_is_a_phrase` could be used to represent "This is a phrase".

[10] `http://www.worldvaluessurvey.org`

of the selected questions we calculate the mean value of replies for each country. This value will give us the average value in terms of the Likert scale for a particular question in a particular country. We call this average the *WVS Score*.

4.2 Methodology

In this section we describe our methodology, from data collection to calculating the online values. We describe the several methodologies of Twitter collection, present our Twitter dataset, explain the country identification algorithm, the word embedding techique, the WEAT test, and in the end we present our methodology to measure values online by using word embeddings.

4.2.1 Twitter Collection

4.2.1.1 Collection Methodologies

There are different ways to collect Twitter data. Earlier in Twitter lifetime, user ids were numerical and small. This allowed one to sequentially probe the numerical IDs and collect virtually all Twitter profiles, as well as tweets and list of followers and followees [26]. Since the Twitter ID system was modified to use higher randomly generated numbers, this method became invalid.

Another common technique is the *snowball* crawling method. It works by executing an exploratory Breadth-first search in the social graph of Twitter. The collection starts by selecting an initial user *seed* (or a list of seeds), extracting the user connections, updating the user list, and repeating the process. The algorithm continues until a certain threshold is reached or no new users are found [48, 102].

The third collection methodology in Twitter is done by *querying*, either through *search* or *streaming*[11]. The collection depends on a list of query terms, that will be used to search Twitter and collect the tweets having that terms. The query term can be a word, a phrase or a hashtag. After having the query terms defined, the collector will retrieve all the tweets returned by Twitter containing the corresponding term. It is important to note that Twitter provides an official API, having functions designed to make searches. There are limitations regarding the number of requests allowed in a certain period of time (e.g 200 requests per hour), and also limitations about the period coverage (e.g return only tweets from 1 weeks ago or newer)[12].

[11]

[12]From the official Twitter API documentation [`https://developer.twitter.com/en/docs/tweets/search/api-reference/get-search-tweets.html`, accessed on 05-Aug-2019]: "Keep in

The fourth method to collect information in Twitter is by acquiring a randomly selected sample of tweets. There are different ways to have access to these samples, commonly purchasing access from GNIP[13]. These samples are generally organized as daily snapshots, existing different percentages of coverage (1%, 10%, 50% or 100%). For instance, a snapshot of 1% for a particular date, consists of a 1% random sample of all tweets published on that day. These snapshots are commonly refered as *gardenhose*, *decahose* or *firehose*.

Finally, the most recent and fifth method of collection is by using the Premium API. In 2014, Twitter announced that it would index all tweets, from the past or the future, improving the coverage and quality of search results [129]. Then, in 2017 Twitter released the Premium API, which is a paid service that extends the API functions, removing the period coverage limitation present in the free regular search API [116]. This allows one to retrieve all the indexed tweets for determined search parameters (query term, period, etc).

All the aforementioned collection methods have their own limitations and biases, and also have different use cases, being adequate for different goals and scenarios. The researcher has to take into account its resources (equipment, budget, time), the topic of research, and the limitations to evaluate which methodology will suffice its needs.

4.2.1.2 Internet Archive

For this work, the author uses the method of randomly selected sample of tweets (fourth method). Since there is no specific topic or event being covered in the study, and we want general published tweets from all over the world, any collection method consisting of selecting list of words to be queried is inadequate. Also, methods consisting of graph search (snowball) are known to have biases [48, 102], which could affect the country coverage of this study. Besides that, the cost of collecting user profiles is not worthy, since we are not focusing on the user and the corresponding social graph. Our goal is to have a huge collection of *tweets*, from several countries, in different languages.

The Internet Archive is a non-profit organization with the goal of building a digital library, providing free public access to several artifacts in digital form. It allows people to download images, videos, books, software, websites and other media. One of the digital artifacts that Internet Archive started indexing and publishing is tweets. The "Twitter Stream Grab"[14] project archives a collection of tweets in the JSON format,

mind that the search index has a 7-day limit. In other words, no tweets will be found for a date older than one week."

[13] `http://support.gnip.com/sources/twitter/`

[14] `https://archive.org/details/twitterstream`

including metadata such as profile information of the author. It consists of a sample of 1% of the public twitter stream. Internet Archive publishes monthly collections, internally consisting of samples for all the days in the corresponding month. The publications are not regular, and might be delayed. For instance, as of this date (January 2019), the last available collection is from October 2018.

In this sense, we downloaded the full collection for the year 2014, which is the last year of the last wave of the World Value Survey. The collection consists of 12 tar compressed files, one for each month[15]. In total, we have **1,709,071,452** tweets (representing 1% of all tweets in 2014).

4.2.2 Location Identification

We want to capture social values across several countries, so the first step is to identify the country of origin of the tweet or user. There are basically two methods to extract location in Twitter. The first one is to use the GPS coordinates metadata contained in geo-tagged tweets. When posting a tweet, the user can enable an option to mark the exact location from where the tweet is being posted. The second method is by exploring the field "location" in the profile of the tweet author (poster). This is a free-text field where the user can write anything she wants as a location (e.g. "Los Angeles", "India", "100 Fictional Street, London, UK").

The second method is more adequate for the purpose of our work and will be used. The geo-tagged method, even though being more precise, is rarely utilized [109] (less than 1% of the tweets), which would limit the amount of tweets covered and result in a small dataset for the countries. Besides that, the geo-tag feature does not necessarily reflect the location of origin or residence of the user, it rather represents the current location. For instance, a tourist visiting some foreign country could post tweets during her trip. On the other side, the "location" field is an explicit information for the location of origin or residence.

Having access to the "location" field is not sufficient. We have to extract the country from the text the user wrote. Before we detail our method of country identification, it is important to notice some limitations. Being a free-text field, the user can write literally anything. For instance, "`nowhere`", "`in your heart`" and "`I don't know`" are all valid inputs. There is also the possibility of the user lying when filling the field. Another potential problem is for ambiguous place name locations (e.g. two cities that have the same name but are located in different countries). We acknowledge

[15]The collection of January 2014 is empty, and have no tweets.

these potential problems can lead to errors in the identification of the country, which can create noise in the dataset.

The core of our country identification algorithm uses the *Nominatim* API[16], provided by *OpenStreetMaps*. The OpenStreetMaps is an open collaborative mapping project with the goal of providing free editable maps of the world. One of its services is Nominatim, which is a geocoding and reverse geocoding tool, allowing one to search for names of places, addresses or specific points in the map. Given a generic string (in any language), the API will return several location information, including latitude, longitude, city, county, state and country of the identified place. For instance, "copacabana beach" will be identified as being from the city "Rio de Janeiro", the state "RJ" and country "Brazil". The API will actually return a list of places, ordered by relevance according to the queried string. Our algorithm gets the first place of the list, and extracts the `country_code` field (e.g: `CA` for Canada).

Running the function of country identification for each of our tweets would be unfeasible, since we have around 1.7 billion tweets. This process would not only take a lot of time to finish, but could also overload the Nominatim service. With that in mind, we make an strategy of creating a *text location dictionary*, where the key is the original written text in the location field (e.g. "New York"), and the value would be the country code (e.g. "US").

In order to create our text location dictionary, we first extract all the *unique* strings (lower cased) of the user location fields from all of our tweets. In total, we extracted **27,604,098** unique location strings. The algorithm of this process is described in Figure 4.1.

The next step is filtering the less common strings, aiming to reduce the number of requests necessary and remove very specific strings, which intuitively are more prone to be errors (typos, jokes, etc). We select only strings with more than 100 occurrences, and then apply the country identification function (Nominatim). This process resulted in **336,680** location strings with a valid country identified, and its algorithm is described in Figure 4.2.

With the text location dictionary built, we can create a tweet dataset for each country. We select a list of 59 countries [17], which are the ones contained in the Wave 6 of the World Value Survey. In order to create the 59 tweet datasets we load the dictionary into memory, go through our complete tweet dataset, and verify their user location field. If the location is empty or it is not in the dictionary, we ignore the tweet.

[16] `https://nominatim.openstreetmap.org`

[17] The World Values Survey also includes Hong Kong as a separated territory, but our location identification methodology is not able to identify Hong Kong separately from China.

Algorithm 1: Extract Location Strings

Input: A set $T = \{t_1, t_2, \ldots, t_n\}$ of tweets
Output: A table $L = \{(loc_1, count_1), (loc_2, c_2), \ldots, loc_m, count_m\}$ with the frequency of unique location strings present in T

$L \leftarrow \emptyset$;
for $tweet \in T$ **do**
  $location \leftarrow$ `LowerCase`$(tweet.user.location)$;
  if $location \notin L$ **then**
    $L_{location} \leftarrow 0$;
  end
  $L_{location} \leftarrow L_{location} + 1$;
end

Figure 4.1. Algorithm for extracting unique location strings from the set of tweets. `LowerCase` converts a string to lower case.

Algorithm 2: Resolve Locations

Input: A table $L = \{(loc_1, count_1), (loc_2, c_2), \ldots, loc_n, count_n\}$ with the frequency of unique location strings
Output: A table $C = \{(loc_1, country_1), (loc_2, country_2), \ldots, loc_m, country_m\}$ with the corresponding countries identified for the locations strings in L

for $(location, count) \in L$ **do**
  if $count > 100$ **then**
    $response \leftarrow$ `NominatimQuery`$(location)$;
    $country \leftarrow country\_code$ of the $address$ of the first location in $response$;
    $C_{location} \leftarrow country$;
  end
end

Figure 4.2. Algorithm for identifying the country of location strings. `NominatimQuery` makes a request to Nominatim geocoding API.

If the location is the dictionary, we save the tweet in the corresponding country dataset that was identified by the dictionary. We present the algorithm of the algorithm of this final step in Figure 4.3. Table 4.1 presents the number of tweets in each country dataset.

Algorithm 3: Create Countries Tweet Datasets

Input: A set $T = \{t_1, t_2, \ldots, t_n\}$ of tweets, a table $C = \{(loc_1, country_1), (loc_2, country_2), \ldots, loc_m, country_m\}$ with string locations and corresponding countries, and a set S of selected countries

Output: A list of set of tweets $TC = \{T_{c1}, T_{c2}, \ldots, T_{cs}\}$ from the countries in S

```
for country ∈ S do
    TC_country ← ∅;
end
for tweet ∈ T do
    if tweet has location information AND tweet is not a retweet then
        location ← LowerCase(tweet.user.location);
        country ← C_location;
        if country ∈ S then
            text ← CleanTweet(tweet.text);
            lang ← tweet.lang;
            TC_country ← TC_country ∪ (text, lang);
        end
    end
end
```

Figure 4.3. Algorithm for creating the datasets of tweets for the countries. `LowerCase` converts a string to lower case. `CleanTweet` removes hashtags, mentions and URLs from the tweet text.

4.2.3 Language Model

After creating the tweet datasets for the countries, we need to create language models that captures the intrinsic association between words, ideally reflecting the cultural values of the countries. In order to achieve this goal, we will use *word embeddings*, a natural language processing (NLP) modeling technique.

Word embeddings are a set of NLP algorithms with the purpose of mapping words or phrases to mathematical vectors, placed in a multi-dimensional space. Several methods could be used to train the model and create the mapping, including probabilistic modeling [49], dimensionality reduction [74] and neural networks [83].

One of the most popular word embedding techniques is *word2vec* [84]. By using a neural network of two layers, it process a considerable large corpus of text and creates a vector space. Each word can then be mapped to this space. The vectors are positioned in such a way that words with similar contexts are close to each other. Differently from a one-hot encode model, it has lower dimensionality (in the order of hundreds) and the

			Number of tweets		
Code	Country	Main language	Total	In main language	In English
US	US	English	55,088,053	49,362,360	49,362,360
JP	Japan	Japanese	29,836,540	37,184,080	832,319
BR	Brazil	Portuguese	15,305,954	12,605,659	1,625,299
AR	Argentina	Spanish	10,059,133	8,611,059	855,574
ES	Spain	Spanish	7,717,853	5,906,840	1,010,264
MX	Mexico	Spanish	6,275,939	5,178,347	903,008
TR	Turkey	Turkish	6,255,539	6,132,328	366,487
PH	Philippines	English	6,226,766	3,188,789	3,188,789
RU	Russia	Russian	6,186,556	5,138,169	551,170
CN	China	Japanese	3,844,676	3,446,182	640,134
CO	Colombia	Spanish	3,711,036	2,709,804	810,856
DE	Germany	English	3,428,472	1,950,988	1,950,988
MY	Malaysia	Indonesian	3,254,736	1,667,041	1,218,538
IN	India	English	2,805,073	2,199,776	2,199,776
AU	Australia	English	2,651,664	2,503,352	2,503,352
KR	South Korea	Korean	1,970,919	1,258,854	367,851
NL	Netherlands	Dutch	1,877,977	990,339	622,617
CL	Chile	Spanish	1,787,414	1,332,835	269,264
TH	Thailand	Thai	1,730,430	1,317,134	333,684
EG	Egypt	Arabic	1,411,587	1,385,425	248,931
ZA	South Africa	English	1,386,417	1,191,821	1,191,821
UA	Ukraine	Russian	1,283,010	807,524	154,103
KW	Kuwait	Arabic	1,222,679	1,592,162	109,233
NG	Nigeria	English	1,074,153	857,511	857,511
PL	Poland	Polish	1,055,165	543,700	462,717
UY	Uruguay	Spanish	1,002,006	824,160	79,612
TW	Taiwan	Japanese	980,163	908,628	115,414
PK	Pakistan	English	818,626	616,261	616,261
EC	Ecuador	Spanish	817,726	653,804	115,795
SG	Singapore	English	781,147	593,399	593,399
SE	Sweden	Swedish	759,889	315,032	338,241
PE	Peru	Spanish	712,215	509,828	122,372
NZ	New Zealand	English	531,020	450,150	450,150
IQ	Iraq	Arabic	360,256	250,062	93,570
BY	Belarus	Russian	351,247	276,071	68,273
MA	Morocco	English	339,685	169,710	169,710
QA	Qatar	Arabic	311,220	309,139	66,826
RO	Romania	English	285,551	134,237	134,237
TN	Tunisia	Arabic	236,378	172,414	58,285
JO	Jordan	Arabic	232,592	217,747	45,883
KZ	Kazakhstan	Japanese	226,151	107,194	58,620
BH	Bahrain	Arabic	223,600	147,223	43,112
PS	Palestine	Arabic	222,538	168,857	37,945
LB	Lebanon	Arabic	192,467	109,546	83,344
DZ	Algeria	Arabic	188,395	97,868	51,504
GH	Ghana	English	183,888	160,639	160,639
SI	Slovenia	English	174,368	106,951	106,951
AZ	Azerbaijan	English	169,738	109,091	109,091
YE	Yemen	Arabic	159,713	177,078	20,958
EE	Estonia	English	134,719	71,258	71,258
LY	Libya	Arabic	111,735	120,143	14,595
AM	Armenia	English	110,071	31,726	31,726
TT	Trinidad & Tobago	English	108,688	84,122	84,122
GE	Georgia	English	103,464	51,532	51,532
CY	Cyprus	English	95,146	43,955	43,955
KG	Kyrgyzstan	English	81,792	65,328	65,328
ZW	Zimbabwe	English	62,408	47,956	47,956
UZ	Uzbekistan	Russian	35,414	14,672	7,917
RW	Rwanda	English	16,196	12,007	12,007

Table 4.1. List of countries with their respective 2-letter country codes, the most popular language, the total number of tweets in the dataset, the number of tweets written in the main language, and the number of tweets written in English.

vector space holds semantic and morphological patterns of the language, which could be extracted by applying vector arithmetic operations. For instance, the relationship "woman is to man as queen is to king" could be obtained with the following vector operation $V_{king} \approx V_{queen} - V_{woman} + V_{man}$, where V_{word} is the vector for the word '*word*' in the space created by the model.

The word2vec model quality depends on large corpus of text. We could directly use our tweet dataset to create the country models, but this might not be ideal for two reasons. First, even though we have a considerable high number of tweets, this is not true for all the countries. Second, it is possible that the tweet corpus doesn't contain certain words of interest related to the values that will be tested (i.e. low vocabulary coverage). To mitigate these problems, we will pre-train our country model with a neutral and embracing textual dataset: *wikipedia*[18].

The Wikimedia Foundation provides regular and updated dumps of the articles of all language versions of Wikipedia[19]. It is important to notice that wikipedia is *language* centered, rather than country centered. For instance, there is not a Brazilian Wikipedia, but there is a Portuguese Wikipedia. We download the dumps for 16 potential languages that will be used to pre-train our country models, then train a word2vec model for each of these languages.

Our approach to create the country language models is to load the wikipedia language models as a base, then retrain it with the proper tweets of the particular country. Since the tweets of the country datasets are filtered regarding only location, we have to control for the language. We use the `lang` field of the JSON tweet metadata to identify the language of each tweet. For each country, we choose the most popular language contained in its tweets dataset, as shown in Table 4.1, and use it as the base for training its model. We also want to evaluate the impact of using different languages, so we also train models for the country considering only the tweets written in English in their datasets.

To create the final language model of a country we load the corresponding Wikipedia language model, and retrain it with the filtered tweets of the country in that language. Besides training with Wikipedia, we also train models using only the

[18] `https://www.wikipedia.org`
[19] `https://dumps.wikimedia.org`

tweets. In the end, we will have four models for each *country*:

$$\begin{aligned} MT_{country} &= word2vec\,(T_{country}[\text{Main Lang.}]) \\ MTE_{country} &= word2vec\,(T_{country}[\text{English}]) \\ MW_{country} &= word2vec\,(W_{lang} + T_{country}[\text{Main Lang.}]) \\ MWE_{country} &= word2vec\,(W_{\text{English}} + T_{country}[\text{English}]) \end{aligned} \tag{4.1}$$

We define W_{lang} as the wikipedia dump for language $lang$, and $T_{country}[lang]$ as the tweet dataset for the country, filtered for language $lang$. The plus sign (+) is used as an append operator for the texts in the datasets. In the end of the process, each country of our dataset will have four word2vec language models trained with the aforementioned methodology: MT (tweets in corresponding language), MTE (tweets in English), MW (Wikipedia and tweets in corresponding language), and MWE (Wikipedia and tweets in English).

We fixed the word2vec parameters for all our models. The *size* (number of dimensions) is 600, the *window* (number of words of the context in the document) is 10, the *min_ count* (minimum frequency for a word) is 10, and the *sample* (threshold of the higher-frequency words to be downsampled) is 0.00001. For the retraining of the MW and MWE models, we set the *epoch* [20] as 100, with the goal to increase the influence of the tweets in the previously trained Wikipedia model. These values of the parameters were empirically chosen.

4.2.4 Word-embedding Implicit Biases

Being an artificial intelligence technology trained with human generated data, word embeddings are susceptible to containing biases. Bolukbasi et al. [20] exposed the risks of using word embeddings by showing that models trained with news articles contained gender stereotypes, and then present an algorithm to measure these biases. A complementary work by Caliskan et al. [24], goes on the direction of identifying and measuring human stereotypes in word embeddings models. They propose the WEAT (Word-Embedding Association Test), which is based on IAT (Implicit Association Test), a psychological test for measuring human biases based on reaction times. Instead of using the reaction time, WEAT will explore the distance between words in the dimensional space created by the word embedding model. A second test called WEFAT (Word-Embedding Factual Association Test) is also proposed, which, accord-

[20]The 'epoch' represents the number of iterations over the corpus. In the case of our finetuning methodology, it would define how many times the Twitter data would be iterated and incorporated in the base Wikipedia model.

ing to the authors, is adequate for comparing values of concepts in the word embedding space and factual properties of the world.

Our methodology is based on WEFAT, and relies on the belief that it is possible to capture not only stereotypes, but also *social values* of different cultures and nations. Given a target word w, and two sets of attribute words A and B, we can define the static s associated with each word

$$s(w, A, B) = \frac{\texttt{mean}\left(\texttt{cos}\left(\vec{w}, \vec{a}\right), \forall a \in A\right) - \texttt{mean}\left(\texttt{cos}\left(\vec{w}, \vec{b}\right), \forall b \in B\right)}{\texttt{stddev}\left(\texttt{cos}\left(\vec{w}, \vec{x}\right), \forall x \in A \cup B\right)} \tag{4.2}$$

which is basically a normalized association score comparing the average distance between w and the words in A and the average distance between w and the words in B. Further, we will introduce our concept of *inquiry*, which will use the association score to measure and represent questions from the World Values Survey.

4.2.5 Online Values Inquiry

Our prime goal is to emulate questions from the World Values Survey by using the Word Embedding model trained for the countries. In order to do that, we define the *Online Values Inquiry* (OVI), which is basically a set of words that replicate an specific question of the World Values Survey. An Online Values Inquiry is represented as

$$OVI_{m,w,A,B} = s(w, A, B) \tag{4.3}$$

where m is the word embedding model to be used to measure the word distances, w is the target word (which generally represents the main topic of the question), and A and B are two sets of opposite attribute words (commonlly holding "positive" and "negative" words respectively).

The OVI will measure an association score (as previously defined) with the given words. A positive value means that the target word w is closer to the set of words from A (generally "positive" words), while a negative value will indicate a proximity with the words from set B (generally "negative" words). A value of zero implies that there is no difference between the distances.

Since we will be measuring OVI's for different languages, it is important to have a methodology to generalize the measurement for different countries. First we choose the target question from WVS that we will be capturing. Secondly, we define the set of *English* words that we think will capture that question. After, we translate the set of words for each of our covered languages. Finally, for each of the four models of each

country, we measure the corresponding OVI according to the proper language of the model.

4.2.6 WVS Score

We are analyzing values in an *aggregated* manner for each country rather than in the individual level like the World Values Survey. In order to compare our Online Values Inquiry with the answers in WVS, we need to summarise the replies from the questionnaire. We apply a methodology of calculating a *normalized average* answer.

The WVS questionnaire has a considerable diversity of questions and answers. There are binary questions (e.g. '1 - Yes' and '2 - No'), Likert scale based questions (agreement scale from '1 - Strongly disagree' to '4 - Strongly agree'), and even questions with a scale of 10 options. Also, it is important to notice that for some questions, the *lowest* value in the reply scale is the strongest regarding agreement, like Question V148 ("Do you believe in God? 1 - Yes; 2 - No"), and for other questions the *highest* value in the reply scale will represent the strongest agreement, like Question V192 ("Science and technology are making our lives healthier, easier, and more comfortable; 1 - Completely disagree; ... 10 - Completely agree."). To standardize the reply and scale of all questions, we normalize the reply values so that it will always be between -1.0 and 1.0, being the lowest value the strongest disagreement, and the highest value the strongest agreement. Given a question q from the WVS questionnaire (Q), we calculate

$$Min_q = \texttt{min}\left(D_q \cdot r, \forall r \in Q[q]\right) \tag{4.4}$$

$$Max_q = \texttt{max}\left(D_q \cdot r, \forall r \in Q[q]\right) \tag{4.5}$$

where r is and individual reply value in the original scale of the corresponding question q, and D_q is the *direction* of the scale of the question q, being 1 if the question has the highest value in the reply meaning agreement, and -1 if the highest value in the reply means disagreement. Taking as example the two questions mentioned earlier, we would have $Min_{V148} = -2; Max_{V148} = -1$, and $Min_{V192} = 1; Max_{V192} = 10;$. Having the minimum and maximum values for each question, we can calculate the normalized reply (nr). Given the original reply r of question q we define

$$nr = 2 \cdot \frac{D_q \cdot r - Min_q}{Max_q - Min_q} - 1. \tag{4.6}$$

Basically, what we are doing is *rescaling* the reply from the original scale to

the $(-1, 1)$ scale, taking the agreement direction into consideration. In the end, all the reply values will be in the same standard: -1.0 strongest disagreement, and 1.0 strongest agreement. Finally, we define the *WVS Score* WVS, defined for a question q and a country *country*, calculated as

$$WVS_{q,country} = \texttt{mean}\,(nr, \forall r \in Q_{country}[q]) \tag{4.7}$$

where $Q_{country}[q]$ is the set of replies from the WVS questionnaire for a particular country, filtered for a specific question q, and nr is an individual normalized reply value for that question (as defined in Equation 4.6). Intuitively, the WVS Score will measure the average reply of a question in a country, in terms of an agreeement scale. This value will be useful for comparison and the calculation of the correlation between online and offline values.

It is important to notice that some questions has explicit "Not available" options. For instance, Question V8 ("How important is work in your life?"), has reply values -1 ('Dont't know'), -2 ('No answer') and -3 ('Not applicable'), and Question V211 ("How proud are you to be [nationality]?") has option 5 ('I am not [nationality]'). To calculate the normalized reply value and the WVS Score, we remove the 'Not available' replies, so that they will not be considered in the score.

4.2.7 Limitations

It is important to acknowledge the biases and limitations of our work. We report them so that our results could be properly interpreted and not be overstated.

Since we are collecting and measuring data from the Web, we are limited to analyzing behaviour only from *Internet users*. Even though being fastly growing since its creation, the Internet (as of June 2019) accounts for only 58.8% of the world population [53]. There is also a geographical (and consequently, cultural) gap on its utilization, ranging from a 39.6% penetration rate in Africa to a 89.4% rate in North America [53]. More in particular, we are dealing with an even more restricted group, which are *Twitter users*. As of the first quarter of 2019[21], Twitter reports to have 321 million monthly active users [107], which accounts for nearly 4.1% of the world population. In this sense, our analysis is not fully representative of the whole world, and will probably miss cultural traits from places with no Internet access (e.g. rural areas).

One of the crucial steps of our methodology is the country identification of the users publishing the tweets. A tweet being wrongly identified as being from a certain

[21]The estimation number of monthly active users in Twitter in the fourth quarter of 2014 (the year of the dataset we use in our work) is 288 million [29]

country will influence the language model of the wrong country. We have two limitations in this aspect: (1) self-report data and (2) the accuracy of the reverse geocoding API. The first concern is that we rely on what people writes and report on their profiles, so there is a probability of the person lying about where she/he lives. The second issue is that the Nominatim API has its own errors and limitations. These limitations are mitigated by the fact of the frequency of the location strings being heavily-tailed, i.e. a few more popular strings covers most of the tweets. It is worth mentioning another restriction related to location identification. As previously mentioned, we filter out very rare location strings, in order to reduce the number of request and to remove very specific strings that are potential jokes and misspellings. Unfortunately, this approach will remove real existing places that are simply rare. Since our focus is on the country-level, this is a minor issue. It is possible, though, to mitigate this problem by applying a manual inspection step, such as creating a crowdsourcing task to validate location texts.

Regarding the language models we also identify some limitations. First, it is important to notice that people in a country will speak many languages. Particularly, there are countries with more than one official language. We highlight the case of South Africa, which have 11 official languages [126], and India, which have 2 official languages (Hindi and English), but also 22 regional languages (including Hindi) [119]. Even though being possible to create a word-embedding model with mixed languages, our methodology is very language-centered, since it requires list of words to measure the values. For this reason, we choose the most popular language in the dataset of a particular country to be its main language. Secondly, for the english language models (MTE and MWE), there's an intrinsic bias related to non-english speaking countries: only a portion of the population will be able to write english tweets. These models will be biased to include people in higher social, economical and educational status than the average population. Finally, there is the possibility of the language itself influence the values revealed by a person. For instance, the topics discussed by a peruvian in spanish might be different from the ones she/he writes in English.

As previously described, we use Wikipedia as a neutral and representative source dataset, which is then used to train a base language model of the languages. This is a very common procedure in the word-embedding literature. Nonetheless, it is important to notice that Wikipedia has its own intrinsic biases. Being an online encyclopedia that anyone can edit, it will potentially capture the visions and values of its editors. Another topic that is relevant to discuss is that the *functions of language*[22] of text in Twitter

[22]According to Jakobson [65], there are six functions of language: referential, emotive, conative, phatic, metalingual, and poetic

and Wikipedia are essentially different. Being encyclopedic texts, Wikipedia articles will in most cases have a *referential* function. Tweets, in the other hand, can be used in a multitude of ways, so it can have diverse functions. For instance, an study of Italian political tweets showed that the referential function is present in tweets as well, but other functions such as *emotive* is very representative [34]. In this sense, mixing tweets and Wikipedia articles in the same language model, might be misleading. That being said, we advocate for the use of Wikipedia not as and *end*, but as *base* of the language model, which will then be modified by the tweets to bring its own representations. Additionally, we analyze language models using only tweets, so that we can isolate the influence of Wikipedia.

Being a global and standardized study that can be used to measure culture in different countries, the World Values Survey was a clear inspiration for our own work. As stated before, our goal is not to replicate the WVS, but to use it as a source of cultural information in the offline. Furthermore, there's a fundamental difference between our study and the World Values Survey, regarding how the cultural behaviour is gathered. The WVS is a survey, so it have answers of specific persons on the *individual* level. Our methodology, on the other hand, combines a group of individual manifestations (tweets) then creates a representation for the whole group (country), being, in this sense, an *aggregated* approach. It could be possible to create a methodology to capture cultural values of individuals in Twitter (since tweets are created individually), but it is not the goal of our methodology. It is also important to notice that the World Values Survey has its own biases and limitations that should be taken into consideration.

Another intrinsic difference between our methodology and the World Values Survey is related to the fact that the latter is a questionnaire. A survey has a collection of questions carefully created and compiled to measure specific things (e.g. values), answered by specific persons in a *private* environment. Otherwise, our methodology relies on tweets written by several accounts (people, institutions, companies, bots, etc) on a multitude of topics and situations, published in a *public* environment. The WVS is a *direct instigation* of specific topics, while the tweets are *natural manifestation* of diverse topics. Nevertheless, both approaches are affected by a common problem: trusting on what the person is revealing. People can intentionally or unconsciously lie in a survey, trying to "reveal their best self". Also, in the Online environment, people can create fake profiles, lie, or simply create an "online personna" that expresses only the "good" parts of themselves. It is important to differentiate between what one really thinks and what one publicly expresses or reveal for other people.

We enumerate and summarise all the limitations of the online values work in Table 4.2, categorized among four types of limitation: Selection bias, Accuracy, dataset

intrinsic bias, and Methodology limitation.

Type	Limitation
Selection bias	• Only Internet users are covered • Only Twitter users are covered • Users that tweet more are more likely to be covered • Rare locations might not be covered • Only users of certain languages are covered (multilingual countries) • Tweets written in English have socioeconomic and education bias
Accuracy	• The location identification algorithm might wrongly identify the country in some case
Dataset intrinsic bias	• Wikipedia has its own pre-existing biases • Twitter and Wikipedia have different functions of language
Methodology limitation	• The OVI methodology uses an aggregated aprroach, instead of individual like in the WVS • Tweets are indirect manifestations of the values, instead of a direct instigation like in the WVS • Tweets might have lies and fake personas • Public tweets will have only public expressions of the people

Table 4.2. List of limitations of the online values work, categorized by type.

Some of these limitations can be mitigated in further studies. The selection biases could be reduced by using a higher sample of the Twitter (10% or more instead of 1%). The accuracy of the location identification could be improved by using other APIs or datasets jointly with Nominatim. Applying the OVI methodology in other datasets besides Wikipedia and Twitter can enlighten the discussion on the use of different functions of language. Finally, regarding the limitations of the methodology, we argue that it might be possible to create similar methods (also using word embeddings) to measure online values in the individual level, as well as using other sources of text that reflect more direct and private responses, as long as it uses written text.

Even with these limitations and biases, we believe that our work is relevant and methodologically robust to provide not only insights and findings regarding values and culture on the Internet, but also a framework that allows people to measure online values.

4.3 Results

In this section we present and discuss the results for the measurement of values in the online environment using word embeddings. First, we present the list of OVIs and how they were created in Section 4.3.1. Next, we show in Section 4.3.2 the calculated

online value scores for all the countries, including analysis of correlation between the four language models and a cultural map that aggregate countries with similar values. Finally, in Section 4.3.3, we compare the online values obtained by our methodology with other external variables, such as the WVS question and socioeconomic indicators.

4.3.1 Creating Online Values Inquiries

As mentioned before, our main inspiration for capturing online values is the World Values Survey. In that manner, we create a list of 22 OVIs, or *inquiries* [23], to measure specific values related to one or more questions from the WVS. The complete list of inquiries is presented in Table 4.3, including the inquiry name (for easier referencing), the list of words defining the OVI, the code of reference of the corresponding WVS question, and the original text of the question presented in English version of the survey.

All the inquiries were manually designed with the goal of reflecting the same value captured by the original WVS question. It is important to notice that different set of words can be used to evaluate the same value, which can generate different results. The list we present is a *proof of concept* and can be improved. In this sense, the process of creating an OVI in our study is *exploratory* rather than confirmatory. It might be possible to create a methodology to automatically create the set of words of an inquiry given a WVS question, but this is not the goal of our work. Furthermore, we envision that the possibility of manually designing OVIs is appealing for sociologists, demographers, and other specialists that might want to use our methodology, allowing them to use their own knowledge and expertise when crafting the set of words in the inquiries.

Now we will show some examples of WVS questions to illustrate the creation of the inquiries. For instance, take WVS Question V152. The text of the question is "How important is God in your life?", and the answer is a scale from 1 to 10, being 1 "Not at all important" and 10 "Very important". It is clear that the question has "god" as the main topic of the question, so our target word of the OVI will be "god". Next, we need to define the list of "positive" and "negative" [24] words. Since the original question measures a "level of importance" we make the inquiry have a positive score for considering god "more important", and a negative score the opposite. In that case,

[23] From this point forward we use the short term *inquiry* interchangeably with its complete name, Online Values Inquiry, or its acronym OVI.

[24] The positive and negative here is not necessarily related to the set of words being "good" or "bad", but a mere indication of the final score being positive or negative considering the target word being closer to one of the set of words.

we build the positive set of words to be $(good, great, important)$, and the negative set $(bad, useless, optional)$.

Some questions in WVS are, in a sense, impossible to be captured with our methodology, either because they are very person-centric (demographic) or are too complex (have multiple topics). Take for instance WVS question V230, that asks rather the person works for the government, a private business or a private non-profit organization. Another good example is question V242, that asks the respondent's age. Now, to give an example of a complex question, take question V208, that asks rather it is justifiable "For a man to beat his wife". It is not clear which one is the main topic (word) of the question ('man', 'beat' or 'wife'), so that it could be used as the target word of the OVI. As stated before, our goal is not to reproduce the whole WVS survey, but we highlight here one of the limitations of our methodology.

There are some questions in the survey that are grouped together due to having the same common prefix and different suffixes (topics). For instance, questions V198-V210 all ask rather the person thinks something can be never justifiable or always justifiable, in a scale from 1 to 10. The difference between them, are the topic. Question V203 will ask rather homosexuality is justifiable, and question V204 will ask rather abortion is justifiable. In the first iteration of creating the inquiries we wondered rather a generic list of positive and negative words could be used to capture all these questions, changing simply the target word (e.g: "homosexuality" and "abortion"). We evaluated using the original set of pleasant and unpleasant words [25] from the original WEFAT paper [24]. We noticed that, even though the question is the same for two different topics, the set of words used to capture that value should also be related to the main topic of discussion. For instance, for the Abortion inquiry we use the words $(good, right, life, health)$ as positive and the words $(bad, wrong, death, fetus)$ as negative. These words are related to the pro and anti-abortion discussion.

Another insight regarding the creation of inquiries that is important to take notice is about the number of words in the OVI. We observed that using fewer words is generally better because it is more stable (better correlation) and have higher coverage (more countries with valid scores). Since we are dealing with multiple countries and languages, using a very specific and rare word in English, might cause the inquiry in another language not having that specific word, due to the fact of not having a direct single-word translation. Besides that, even if the word has a translation, it might be missing in the word embedding model, specially in the models utilizing only Twitter data.

[25] Positive: $(joy, love, peace, wonderful, pleasure, friend, laughter, happy)$. Negative: $(agony, terrible, horrible, nasty, evil, war, awful, failure)$.

Inquiry				World Values Survey Question		
Name	**Target Word**	**Positive Words**	**Negative Words**	**Var. Code**	**Question**	**Answer Options**
God	god	good, great, important	bad, useless, optional	V148	Believe in: God	1=Yes, 2=No
				V152	How important is God in your life	Scale: 1="Not at all important", 10="Very important"
Science	science	good, great, love	bad, wrong, hate	V192	Science and technology are making our lives healthier, easier, and more comfortable.	Scale: 1="Completely disagree", 10="Completely agree"
				V193	Because of science and technology, there will be more opportunities for the next generation.	Scale: 1="Completely disagree", 10="Completely agree"
Nationality Pride	<country name>	good, love, pride	bad, hate, shame	V211	How proud of nationality	Scale: 1="Very Proud", 4="Not at all proud"
Prostitution	prostitution	sex, work, law	bad, shame, ugly	V203A	Justifiable: Prostitution	Scale: 1="Never justifiable", 10="Always justifiable"
Homosexuality	homosexual	respect, pride, beautiful	hate, shame, ugly	V203	Justifiable: Homosexuality	Scale: 1="Never justifiable", 10="Always justifiable"
Abortion	abortion	good, right, life, health	bad, wrong, death, fetus	V204	Justifiable: Abortion	Scale: 1="Never justifiable", 10="Always justifiable"
Divorce	divorce	good, normal, allowed	bad, forbidden, sin	V205	Justifiable: Divorce	Scale: 1="Never justifiable", 10="Always justifiable"
Euthanasia	euthanasia	rest, peace, relief	sin, kill, evil	V207A	Justifiable: Euthanasia	Scale: 1="Never justifiable", 10="Always justifiable"
Violence	violence	protection, necessary, legit	unacceptable, repugnant, evil	V210	Justifiable: Violence against other people	Scale: 1="Never justifiable", 10="Always justifiable"
Stealing	steal	necessary, legit, forgivable	unacceptable, wrong, dishonest	V200	Justifiable: Stealing property	Scale: 1="Never justifiable", 10="Always justifiable"
Suicide	suicide	relief, peace, understand	sin, wrong, tragedy	V207	Justifiable: Suicide	Scale: 1="Never justifiable", 10="Always justifiable"
Religion	religion	good, great, important	bad, useless, optional	V9	Important in life: Religion	Scale: 1="Very Important", 4="Not at all important"
Work	work	good, happy, enjoy	bad, sad, tired	V8	Important in life: Work	Scale: 1="Very Important", 4="Not at all important"
Politics	politics	good, debate, elections	bad, sad, corruption	V7	Important in life: Politics	Scale: 1="Very Important", 4="Not at all important"
Friends	friends	good, love, happy	bad, hate, sad	V5	Important in life: Friends	Scale: 1="Very Important", 4="Not at all important"
Family	family	good, love, happy	bad, hate, sad	V4	Important in life: Family	Scale: 1="Very Important", 4="Not at all important"
See Myself Reserved	me	reserved, shy, introvert	social, communicative, extrovert	V160A	I see myself as someone who: is reserved	Scale: 1="Disagree strongly", 5="Agree Strongly"
See Myself Lazy	me	lazy, slow	busy, fast	V160C	I see myself as someone who: tends to be lazy	Scale: 1="Disagree strongly", 5="Agree Strongly"
See Myself Nervous	me	nervous, angry	calm, relaxed	V160I	I see myself as someone who: gets nervous easily	Scale: 1="Disagree strongly", 5="Agree Strongly"
See Myself Happy	me	happy, glad	unhappy, sad	V10	Feeling of happiness	Scale: 1="Very happy", 4="Not at all happy"
Child Rising	child	obedience, religion, faith	independence, determination, creativity	Y003	Autonomy Index	-
Life Priority	important	security, economy	freedom, rights	Y002	Post-materialist index (4-item)	-

Table 4.3. List of OVIs (inquiries) with their respective name for reference, target word, positive attribute words, negative attribute words, and information of the corresponding WVS question, containing its variable code of reference, the question text in the survey, and the options avaiable to respond. The "God" and "Science" inquiries are linked to two WVS questions. The "Child Raising" and "Life priority" inquiries, instead of a proper WVS question, are inspired by an WVS Index derived from a couple of questions.

In the next section we will show the online values scores for all the inquiries presented in Table 4.3.

4.3.2 Online Values

Once we have defined the inquiries and their respective target and attribute words, we will calculate the corresponding association scores for each one of the four models of each country.

4.3.2.1 Country Values

We start by analyzing the actual value of the association scores of the inquiries among the countries. We plot color matrices tables, one for each type of model we have, presented in Figure 4.4. Each facet in the plot corresponds to a particular model type: MT, MTE, MW, and MWE (top to bottom). Each row is an inquiry from Table 4.3, and each column is a country. For easier referencing, we add in the top of each column the image of the flag of the country and the respective language utilized. The color of

each cell (tile) represents the actual value of the association score, ranging from dark purple (most negative value), to dark green (most positive value).

The cells without a tile are caused by the fact of not being possible to calculate the score in that case. This will happen when a certain word of the OVI is not present in the respective word embedding model, making it impossible to calculate the distances. It is noticeable how the models utilizing tweets (the two facets in the top of Figure 4.4) have a considerable amount of missing scores, while the models utilizing Wikipedia + tweets (the two facets in the bottom of Figure 4.4) have most of the scores complete. This is expected, since the Twitter corpora is more limited compared to Wikipedia, which is an encyclopedic corpora. Also, the tweets-only models are a subset of the Wikipedia models, being by definition more restricted.

Now, looking at the color patterns, it is interesting to notice how different inquiries have different patterns. We have inquiries like "Euthanasia" and "Suicide" that have predominant negative values (all countries with a purple color). On the other side, we have inquiries such as "Friends" and "Family", with all the countries having a positive value (green color). Intuitively these inquiries were able to capture some expected behaviour of common sense, such as people, in general, liking families and friends and disliking suicide and euthanasia. Further in Section 4.3.3 we will make more detailed comparisons between the online scores and offline values.

There are also inquiries with a high diversity of scores, having countries with both positive and negative values, which is the case of the inquiry "See Myself Reserved". It is important to note that, even for inquiries having a consistent and predominant *signal*, the *power* of the association score might be different among the countries, resulting in stronger or weaker relationships. This phenomena, as can be observed in Figure 4.4, validates our methodology in the sense of being able to measure differences between the countries. The ability of ranking countries according to a certain inquiry will be explored to evaluate the similarities between the online and the offline.

4.3.2.2 Ranking of Countries

We will highlight now some specific cases that are worthy discussing in more detail. We will analyze three inquiries: "God", "Homosexuality" and "Abortion". For better visualization, we will plot a world map with the distribution of the scores on the map, and a ranked list of the corresponding countries from the higher (top) to lower (bottom) score. Each one of these maps corresponds to a certain specific row from Figure 4.4. So, for each inquiry we will have a specific version of the model selected, properly indicated in the title of the corresponding figure.

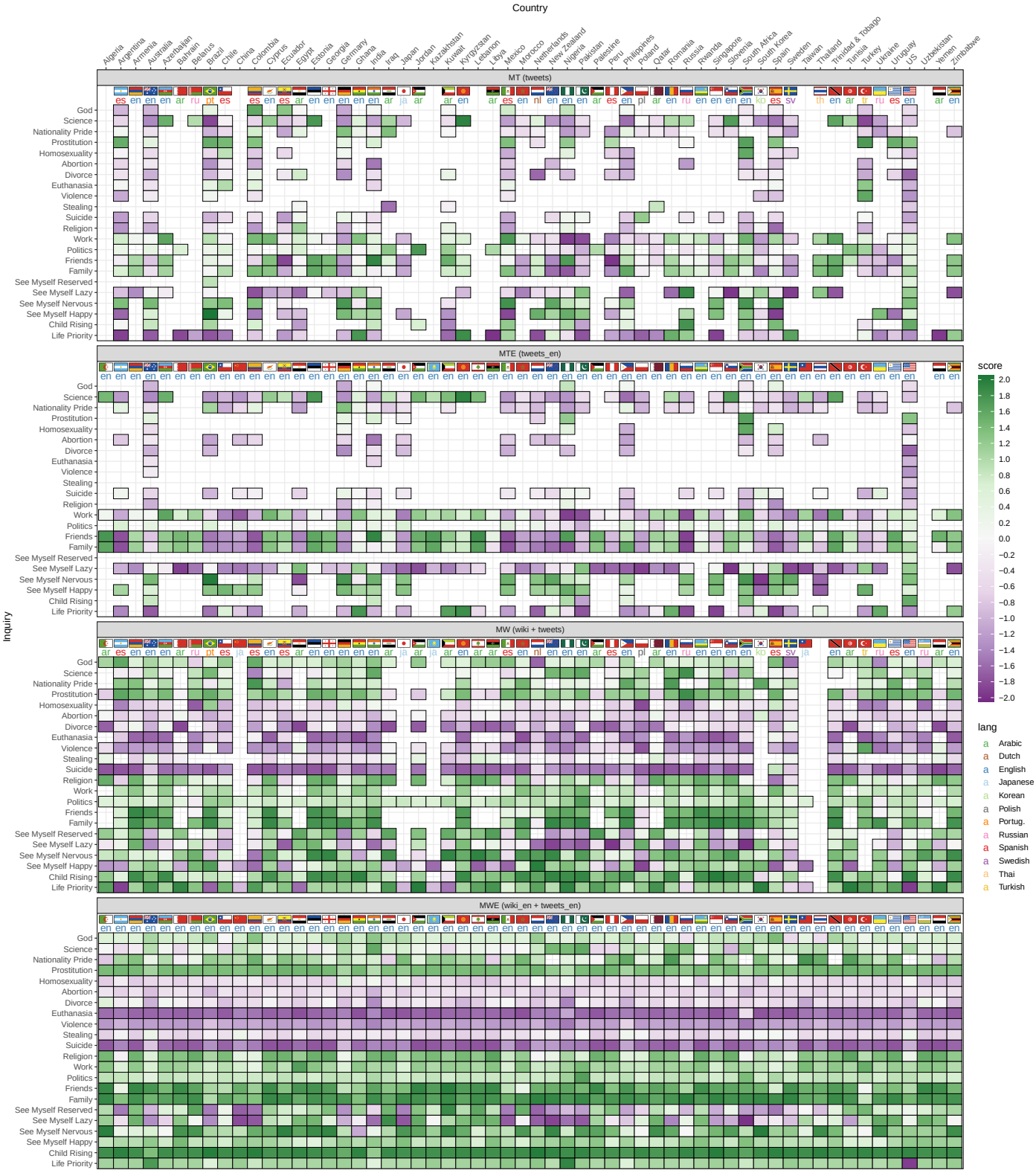

Figure 4.4. Inquiries matrix plot for the four types of model.

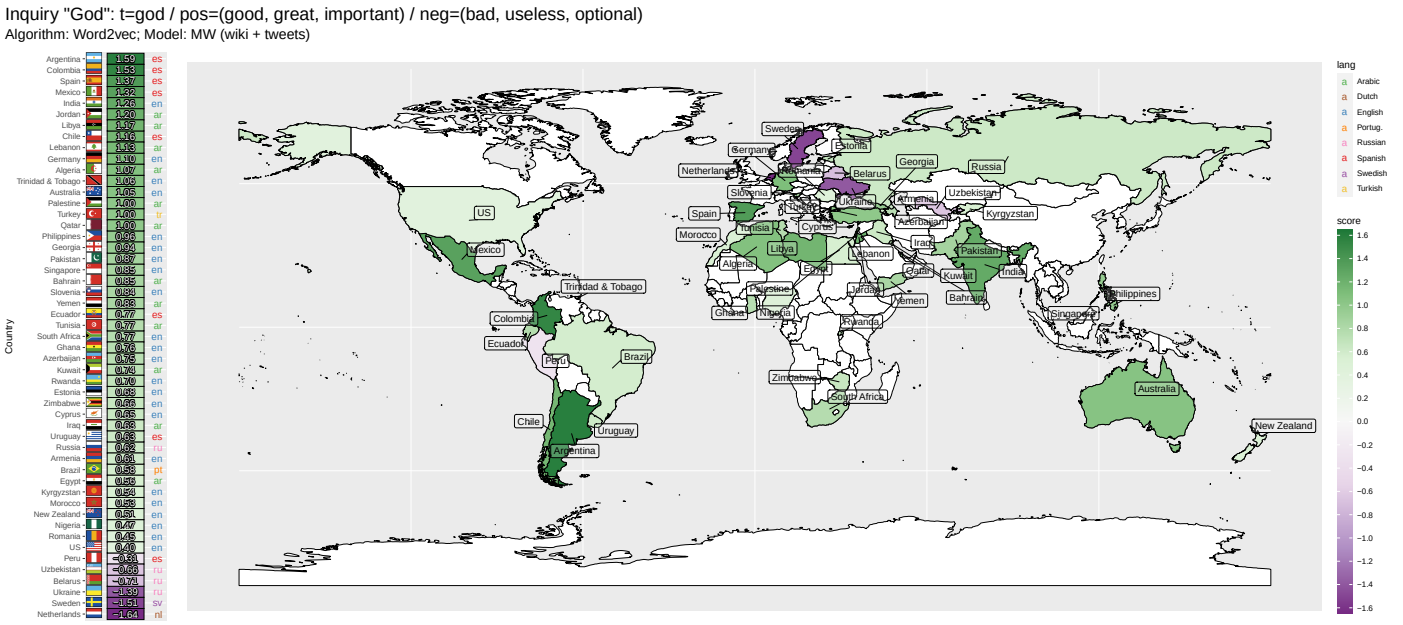

Figure 4.5. Ranking of countries and world map for the God inquiry.

Religion First, we present the "God" inquiry, plotted in Figure 4.5. We notice that there is a major *positive* pattern for the global score of the "God" inquiry, meaning that for most of the countries the word 'god' is closely associated with the words 'good', 'great' and 'important'. According to a report by the WIN-Gallup International Association from 2015 [127], with the exception of Spain, all of the other four countries in the top of our ranking have most of their population as identifying themselves as a religious person: Argentina (72%), Colombia (82%), Mexico (68%), India (76%). Following the top of our ranking, we also see examples of arabic-speaking countries with major religious population, such as Lebanon, Algeria, Palestine, and Turkey (respectively 80%, 90%, 75%, 79% of religious people [127]).

Now analyzing the bottom of our ranking in Figure 4.5, we see Netherlands and Sweden, which are one of the *least religious countries* according to the same report [127]. Sweden is behind only from Japan and China, being the third least religious country, having 59% of its population self-declaring as not religious and 17% as a "convinced atheist". Netherlands occupies the 5th position, with 51% of not religious and 15% of atheists.

We highlight also two cases of divergence between our online ranking and the WIN-Gallup report. Germany is the 10th *most* religious country in our ranking (Figure 4.5), but it is actually not a very religious country, having only 34% of its population self-declaring as religious [127]. On the other hand we have Peru, being the sixth *least*

religious country according to our online ranking, but being placed as the 20th most religious country, with 82% of its citizens being religious.

Overall, the "God" inquiry was able to capture religiosity very well. Even though having some exceptions, the top and bottom of the rank seems to be very consistent with external measurements. Further, on Section 4.3.3, we will measure the actual correlation between the online inquiry and the world values survey score.

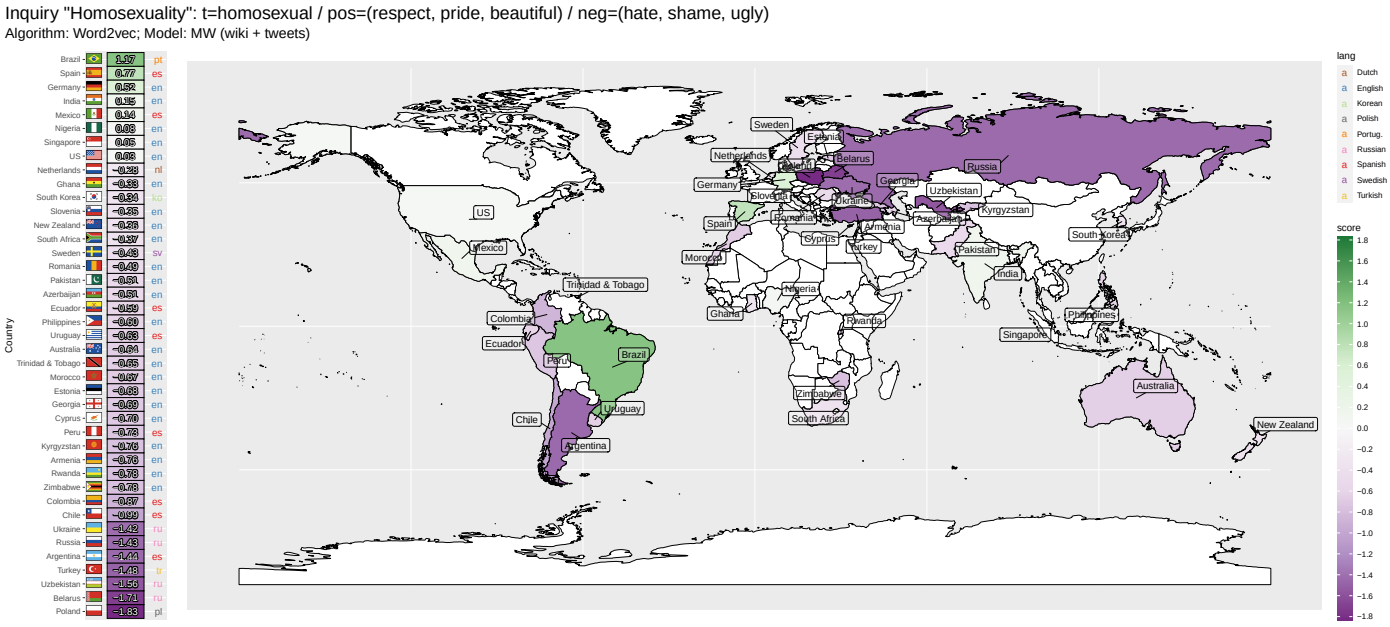

Figure 4.6. Ranking of countries and world map for the Homosexuality inquiry.

Homosexuality Next, the "Homosexuality" inquiry is presented in Figure 4.6. We observe that there is a major negative pattern towards the word 'homosexual'. With the exception of 8 countries in the top of the ranking, all the others have the target word closely related to the negative words 'hate', 'shame' and 'ugly'. Taking a closer look at the top 3 of our ranking we see Brazil, Spain and Germany, all having stronger relationship of the word 'homosexual' with the positive words 'respect', 'pride' and 'beautiful'. Interestingly, all of these countries are known to have some of the biggest LGBT pride events of the world [125]. In Brazil, the LGBT Pride Parade of 2011 in São Paulo had 4 million people [67], being the second biggest in the world [125]. The third biggest LGBT event of all times was in Spain, the 2017 World Pride festival in Madrid, holding 3.5 million people [99], which also held the Europride in 2007 (2.3 million people), and other local and national parade events with over 1.2 million people

(2012, 2016, 2019) [125]. In Germany, the Cologne LGBT parades are known to bring over a million people for many years (2002, 2013, 2018, 2019) [125].

Moving now to the bottom of our ranking (Figure 4.6), we see Poland, Belarus and Uzbekistan as the countries with strongest association of the word 'homosexual' to negative words. Lesbian, gay, bisexual, and transgender people faces legal challenges not experienced by non-LGBT residents in all of these three countries [123, 121, 124]. Poland is actually ranked as the worst European Union country for LGBT rights, according to a 2020 report [60] by the ILGA-Europe [26]. Even though same-sex sexual activity being legal in Belarus since from 1994 [121], there are still reports of aggression and violation of freedom of expression regarding Gay pride [22]. In Uzbekistan, sex between two men is illegal, with the possibility of being punished up to 3 years in prison [124].

We notice also some discrepancy in our ranking of the "Homosexuality" inquiry. Argentina for instance is placed as the 5th country with strongest *negative* relationship, but it is actually among the most *advanced* countries regarding LGBT rights [120], being the first country in Latin America and tenth in the world to legalize same-sex marriage in 2010 [14], and having "one of the world's most comprehensive transgender rights laws" [72]. In the opposite direction is Nigeria, which is placed in our ranking as the 6th country with the strongest *positive* association, but actually does not allow or recognize LGBT rights [122], and even criminalizes same-sex marriage [15].

The "Homosexuality" inquiry captured some interesting and consistent behaviour, placing countries with popular LGBT events in the top, and countries with limited (or nonexistent) LGBT rights in the bottom, even though having same inconsistencies. We acknowledge that this inquiry is limited in the sense of not capturing all the spectrum of sexuality and gender identity, missing for example the bisexual and the transgender communities. The utilization of the word "homosexual" was motivated to reflect the actual World Values Survey question, that asks rather *homosexuality* is justifiable (Table 4.3).

Abortion Finally, we observe the "Abortion" inquiry in Figure 4.7. There's a dominant *negative* pattern, meaning that the word 'abortion' is closely related to the negative words 'bad', 'wrong', 'death', and 'fetus', for all of the countries. However, there are considerable differences among the countries, where the negative association score will vary from -0.02 to -1.51.

[26] ILGA-Europe: European region of the International Lesbian, Gay, Bisexual, Trans and Intersex Association

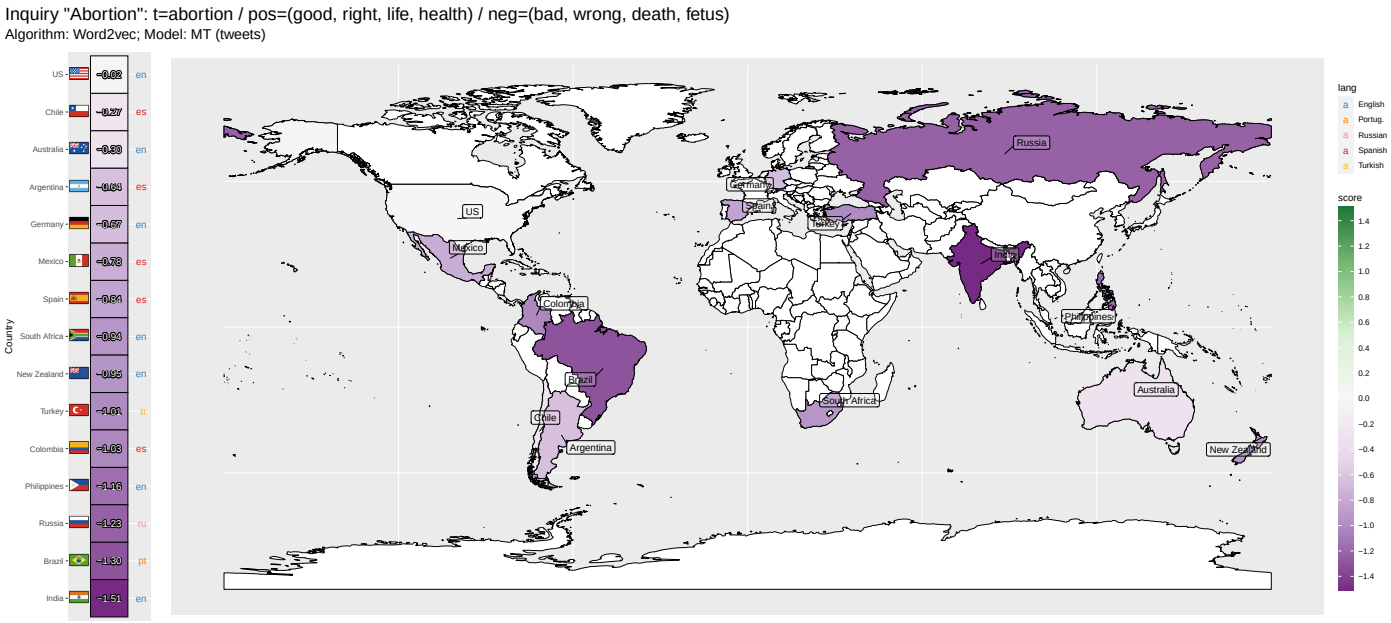

Figure 4.7. Ranking of countries and world map for the Abortion inquiry.

To discuss and analyze the resulting ranking, we compare it with a 2013 report published by the United Nations called "World Abortion Policies 2013" [88]. This report summarises the national laws regarding abortion in the 193 UN member states. It presents a table registering in which legal grounds the abortion is permitted, divided in seven categories: (1) "To save a woman's life", (2) "To preserve a woman's physical health", (3) "To preserve a woman's mental health", (4) "In case of rape or incest", (5) "Because of fetal impairment", (6) "For economic or social reasons", and (7) "On request".

Looking at the top 5 positions of the inquiry ranking (Figure 4.7), with the exception of two South American countries (Chile and Argentina), we have three countries where the abortion is allowed for all the 7 situations presented by the UN report: United States, Australia, and Germany. In the bottom of the ranking, with the exception of Russia (where abortion is allowed for the 7 situations), we have countries that do not permit abortion on request (India, Brazil, Philippines, and Colombia), do not permit abortion for economic or social reasons (Brazil, Philippines, and Colombia). In particular, Philippines allows abortion only when the woman's life is in risk.

The "Abortion" inquiry seems to be capturing the social phenomena for some cases, but there are also a considerable amount of exceptions. It is important to notice that we analyzed here the *legal* aspect of abortion, which might be different from the population values. Further, in Section 4.3.3 we will compare and calculate the

correlation between the inquiries and the World Values Survey scores of the countries.

4.3.2.3 Intra-model Correlation

Now we will check the correlation between our four types of models. We want to verify how differently the inquiries are in two aspects: (1) using native language versus English, and (2) using only tweets versus using Wikipedia plus tweets. This comparison is important to evaluate the compromise of using one strategy instead of another. For instance, it might be infeasible to create inquiries for different inquiries, so adopting a common language strategy might be more appropriate. To achieve that we make the following comparisons:

- MT / MTE: compare native language vs. English in the tweet models
- MW / MT: compare Wikipedia vs tweet models in the native language model
- MW / MWE: compare native language vs. English in the Wikipedia models
- MWE / MTE: compare Wikipedia vs tweet models in the English model

For each inquiry, we separately calculate the Pearson correlation for each one of the four model combinations. When comparing two models we use only the matching countries with valid association scores in both models. We only calculate the correlation if there are at least 3 countries with valid association score. Figure 4.8 presents the matrix plot of correlation between the models for all the inquiries. Correlations with a p-value higher than 0.10 have small font and gray color, and correlations between 0.05 and 0.10 have a medium font.

We observe that there's a major strong positive correlation for MT / MTE (first column), meaning that, in general, the ranking of the countries in relation to the association score of the inquiries of the Tweet-only native language model (MT) is very similar to the corresponding inquiry utilizing the Tweet-only English model (MTE). The same phenomena can be observed for the MW / MWE (third column), which also has a predominant positive correlation for most of the inquiries.

Comparing now the tweets-only model with the Wikipedia + tweets equivalent, we observe that for most of the inquiries there is a weak positive correlation, which is not significant ($p - value > 0.05$) in some cases. This is true both for the native language models (MW / MT) and the English model (MWE / MTE). Even not having strong correlations in general, there are still some exceptions. For instance, the "Violence" inquiry has a correlation of 0.84, and the "See Myself Happy" inquiry a correlation of 0.90, both for MW / MT. Curiously, the "Euthanasia" inquiry presents a strong

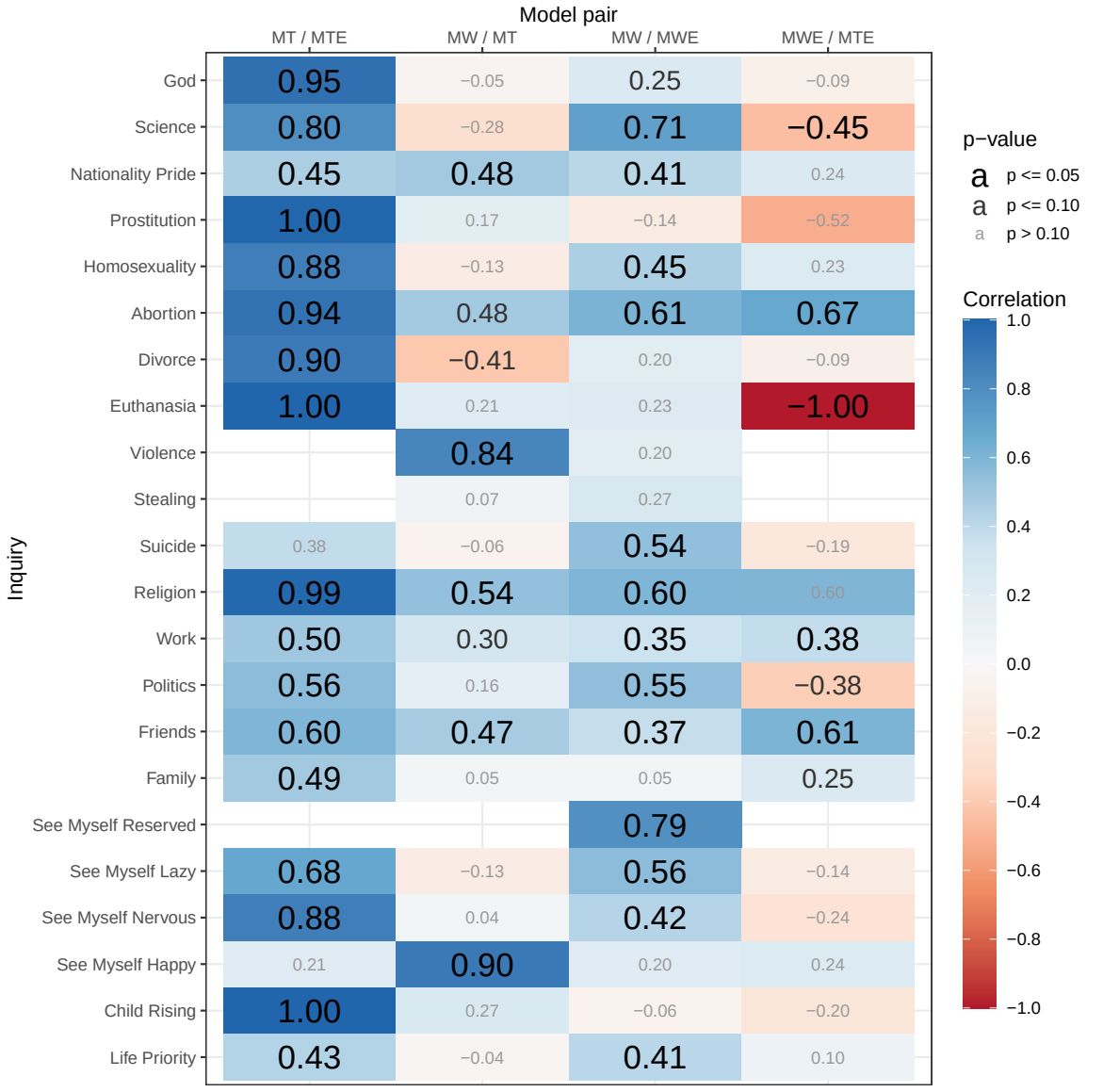

Figure 4.8. Correlation matrix of the inquiries comparing the 4 versions of the word-embedding models.

negative correlation in the MWE / MTE, meaning that the Wikipedia + tweets model has the opposite ranking of countries for its tweets-only counterpart.

In conclusion, we noticed that inquiries with the same type of corpus, but with different languages are, in general, correlated. This indicates that using the same language for all the countries might be a good compromise in scenarios where creating multi-language inquiries are infeasible. As previously discussed (Section 4.2.7), it is important to remember though that using a language that is not native for a country will have an inherent bias (e.g. only people with foreign language education will be

able to communicate in that language). Looking now at the aspect of the type of corpus we noticed that tweet-only models, in general, are not so much correlated with its equivalent Wikipedia + tweets model. This implies that the type of corpus has a major influence in the association scores of the inquiries. Considering that Wikipedia has an encyclopedic text, and tweets are more personal texts, they will have different functions of language (as discussed in Section 4.2.7).

4.3.2.4 Online Cultural Map

Next, we will identify clusters of similar countries regarding their online values measure by our methodology. We use the association score values of the country to measure and plot these similarities. Figure 4.9 shows the original Inglehart-Welzel Cultural Map. Ronald Inglehart and Christian Welzel states that the cultural variation of countries in the world can be depicted by two dimensions [9]:

- **Traditional vs. Secular-rational**: traditional values are related to the importance of religion and traditional family values, while secular-rational values has weaker beliefs on these values, and sees themes like divorce and abortion as more tolerable.
- **Survival vs. Self-expression**: survival values are related to strong beliefs in physical security, economic development, lower trust and lower tolerance, while self-expression values are related to prioritizing participation in the economic and political life, acceptance of foreigners, defending gender equality, and acceptance of LGBT community.

Our goal in this analysis is to create a cultural map of online values. We will evaluate two approaches: dimensionality reduction and factor analysis. The first has the advantage of being generic, in the sense of being able to use any set of inquiries, and the former has the advantage of being specially crafted for capturing specific value dimensions, which is exactly the technique used to build the Inglehart-Welzel cultural map.

There are a number of algorithms to do the task of dimensionality reduction, that can be utilized for different goals such as reducing noise, selecting variables, and simplification of data. In our case, we want to create a 2D visual representation of the online values, so our goal is visualization. In that case there are three techniques that are commonly used: PCA [94], t-SNE [117], and UMAP [82].

For all the dimensionality reduction strategies we use the same dataset: a matrix of 58 countries by 22 variables (inquiries from Table 4.3). We replace missing values

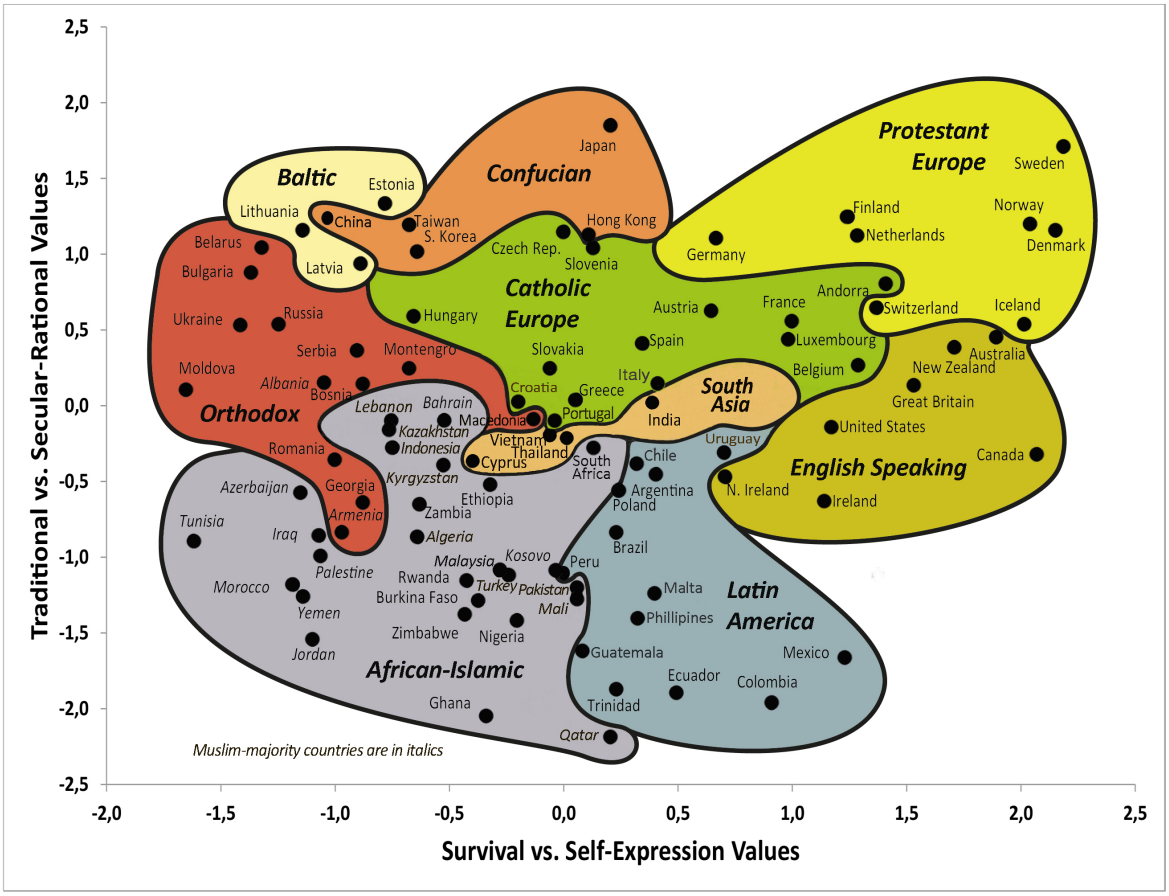

Figure 4.9. Inglehart–Welzel Cultural Map, using data from the Wave 6 of World Values Survey (2010-2014). Original image from WVS Findings webpage [9].

with zero, which is a central-neutral value for the association score. Since the MT and MTE models has scarce data, we only run the analysis for the MW and MWE models.

We experiment with the three algorithms, but will present and discuss only the t-SNE results. Even though presenting different arrangements of the countries, similar findings can be derived from them. Also, it is important to notice that each algorithm has their own parameters, that can also influence the final result of the cultural map.

In Figure 4.10 we show the results for t-SNE ($perplexity = 10$, $theta = 0.0$), for the MW models (left) and MWE models (right). The colors of the countries represents their major language, utilized by the MW models. Even though all the countries of MWE (right) are utilizing English, we choose to color them with the same colors from

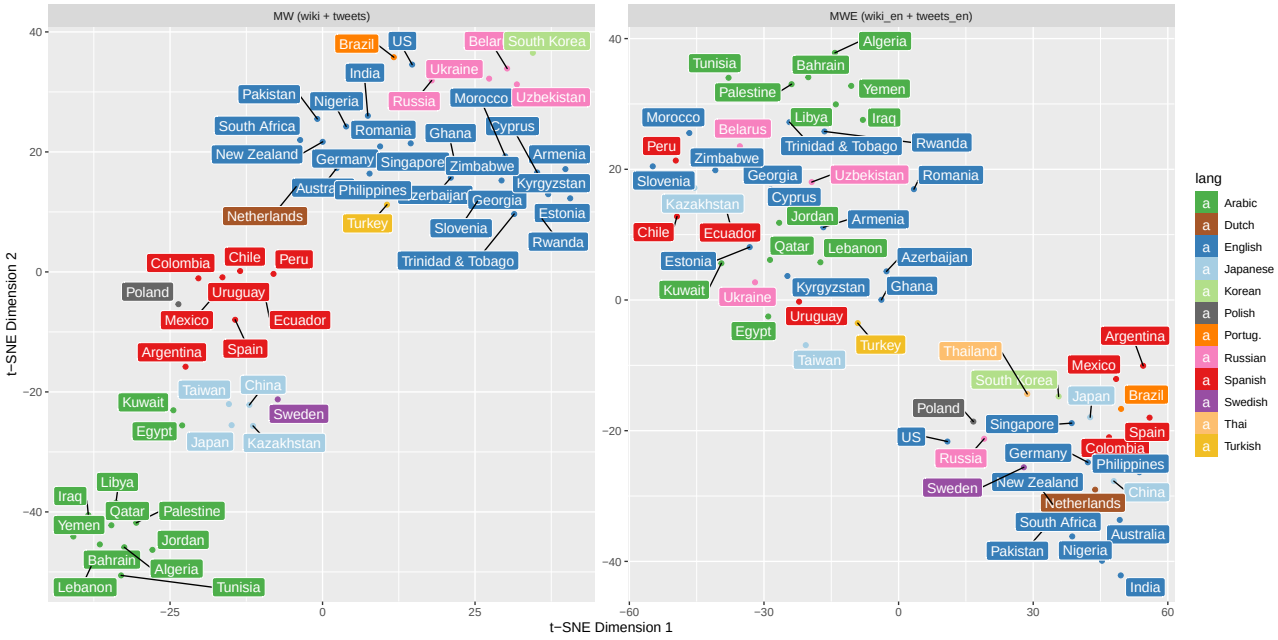

Figure 4.10. Cultural map of the countries considering online values, utilizing t-SNE, for MW and MWE models.

MW, to better compare the effect of the language on the cultural map. It is important to note that, unlike the Inglehart-Welzel cultural map, the x and y axis does not have a semantic.

We observe that the language of the model seems to be the main factor of the clusterization of the countries, as we can see on the plot of the left. When using English as common language for all the countries (plot on the right), the effect of the language is diminished, and two groups of countries are formed, with no apparent pattern. In the end, the cultural map produced by t-SNE is not very meaningful. We believed one of the reasons for the poor performance might be the fact of not all inquiries being a good fit to capture cultural traits. A possible improvement could be to make a "inquiry selection", similarly to a feature selection, so that only good quality inquiries would be utilized.

Instead of improving the dimensionality reduction techniques, we focus on experimenting with another strategy for creating the cultural map, which not only has explicit inquiry selection, but also produce a cultural map where the axes has semantics: Confirmatory Factor Analysis (CFA). This is actually the technique utilized by Inglehart and Welzel to create their cultural map. Table 4.4 shows the two factors and the corresponding questions from the World Values Survey utilized by them. These factors are created with the goal of capturing the cultural dimensions explained earlier:

Traditional vs. Secular-rational, and Survival vs. Self-expression.

Dimension and Item	WVS Question	Inquiry
Traditional vs. Secular-Rational Values		
TRADITIONAL VALUES EMPHASIZE THE FOLLOWING:		
God is very important in respondent's life.	V152	God
It is more important for a child to learn obedience and religious faith than independence and determination.	Autonomy Index	Child Rising
Abortion is never justifiable.	V204	Abortion*
Respondent has strong sense of national pride.	V211	Nationality Pride
Respondent favors more respect for authorizty.	V69	-
(SECULAR-RATIONAL VALUES EMPHASIZE THE OPPOSITE)		
Survival vs. Self-Expression Values		
SURVIVAL VALUES EMPHASIZE THE FOLLOWING:		
Respondent gives priority to economic and physical security over self-expression and quality-of-life.	Post-materialist index	Life Priority
Respondent describes self as not very happy.	V10	See Myself Happy*
Respondent has not signed and would not sign a petition.	V85	-
Homosexuality is never justifiable.	V203	Homosexuality*
You have to be very careful about trusting people.	V24	-
(SELF-EXPRESSION VALUES EMPHASIZE THE OPPOSITE)		

Table 4.4. Items of the dimensions of the Inglehart-Welzel cultural map, with the corresponding WVS variable and online Inquiries. Inquiries marked (*) have the score inverted (multiplied by -1) to reflect the same direction of the item.

The table also shows the corresponding WVS question variable code (or index) associated with the item, and the corresponding OVI from Table 4.3 that will be utilized for the online factor analysis. To reflect the direction of the item, we multiply by -1 the association score of three inquiries ("Abortion", "See Myself Happy", "Homosexuality"). For instance, the abortion inquiry is a score of *agreement* of abortion, but the item utilized in the dimension is the opposite (i.e. a measurement of abortion *never* being justifiable). We run a Confirmatory Factor Analysis (CFA), utilizing the online values captured by the inquiries. The formula utilized by our online CFA is the following:

$$\begin{aligned} TraditionalVsSecularRational &= God + ChildRising + Abortion + \\ &\quad NationalityPride \\ SurvivalVsSelfExpression &= LifePriority + SeeMyselfHappy + \\ &\quad Homosexuality \end{aligned} \tag{4.8}$$

We run a CFA for each model (MW and MWE) and present the resulting factors of the countries in Figure 4.11. The X-axis is the Survival vs. Self-Expression dimension, and the Y-axis is the Traditional vs. Secular-Rational dimension, similarly like the original cultural map in Figure 4.9. Similarly like the dimensionality reduction, we use colors to indicate the language of the MW models, and keep the same colors for the

MWE model, even though the former is utilizing the English language.

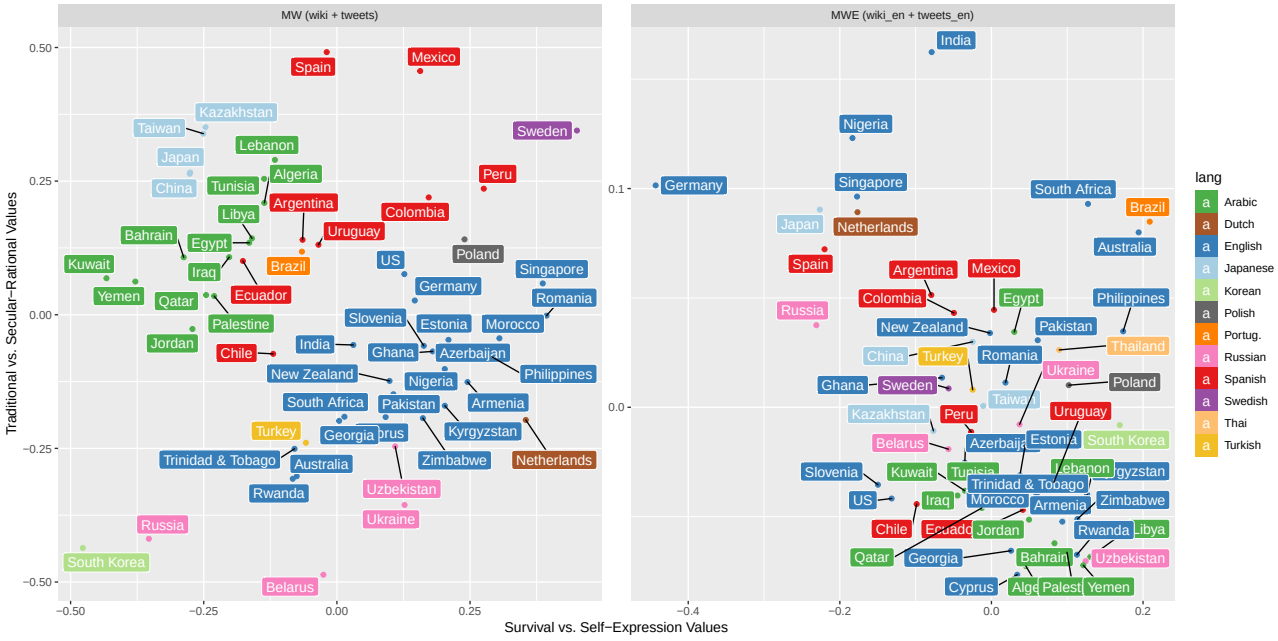

Figure 4.11. Cultural map of the countries considering online values, utilizing Confirmatory Factor Analysis, for MW and MWE models.

We observe that the cultural map utilizing native language (left facet of Figure 4.11) has some similarities withe the Inglehart-Welzel cultural map from Figure 4.9. Sweden is placed in the top-right corner both in the offline and the online cultural map, placing itself as one of the countries with higher Secular-Rational and Self-Expression values. India is in a central position of both maps, being a country with mixed values, in the sense of being placed in the middle of both dimension values. Taiwan, Japan and China are close to each other in the online map, while also being placed in the same cluster of the offline cultural map (Confucian). Some countries of the offline Latin America cluster, are also placed together in the online cultural map, like Argentina, Uruguay, Brazil and Ecuador.

There are also some differences between the offline map from Figure 4.9 and the online map from Figure 4.11 (left). Orthodox countries like Russia, Ukraine, and Belarus are closer to traditional values in the online map, while being closer to the Secular-Rational value in the offline. Even though being clustered together in the online map, Latin America countries have higher Secular-Rational values online than offline. South Korea is placing itself as a very Traditional and Survival country online, even though being more Secular-Rational offline.

Finally, looking at the online cultural map of the English model (MWE, left facet of Figure 4.11), we also observe some patterns. Interestingly, some countries with similar language and culture are close to each other even when not using their native language. This is the case of the Latin America countries like Colombia, Argentina and Mexico, and African-Islamic countries like Iraq, Tunisia and Qatar. In general, the English online cultural map is more disperse, and the cultural clusters proposed by Inglehart and Welzel are not very eminent. Since the model is trained using only tweets written by people capable of communicating in English, it will have an intrinsic bias related to presenting values of people with higher education in countries where English is not a native language.

4.3.3 Offline Values

In the following analyzes we will compare the online values of the inquiries with other "offline" metrics and indicators. We will take the same approach as the previous analyzes and make a country-based approach.

4.3.3.1 WVS Scores

We start by analyzing the values from World Values Survey itself. We select all the 24 WVS Questions we are studying (Table 4.3) and calculate the WVS Score (Section 4.2.6) for all the countries. Some questions are not available in the WVS questionnaire of some countries, so the WVS Score is not available in these cases.

We present a color matrix plot in Figure 4.12 with all the WVS Scores, each row is a question, and each column is a country, where the colors represent the WVS Score from the lowest value (purple) to the highest value (green). This matrix plot can be seen as the analogous offline version of inquiries matrices plots from Figure 4.4.

We observe that, similarly to the online values, the offline values have differences between countries, and also between questions. There are questions with a majority positive score, like "Important in life: Family", and questions with a majority negative score, like "Justifiable: Stealing property".

As a first analysis between online and offline values, we will verify rather there is an agreement between the inquiries and the WVS Questions regarding the *signal* of the scores, regardless of the ranking of the countries. Does our inquiry methodology captures the same positive or negative trend of a particular value? To answer that we calculate the percentage of countries that has the same signal in the online score (inquiry) as in the offline score (WVS). We do that for each association of inquiry and question presented in Table 4.3. In some cases, either an inquiry association score or

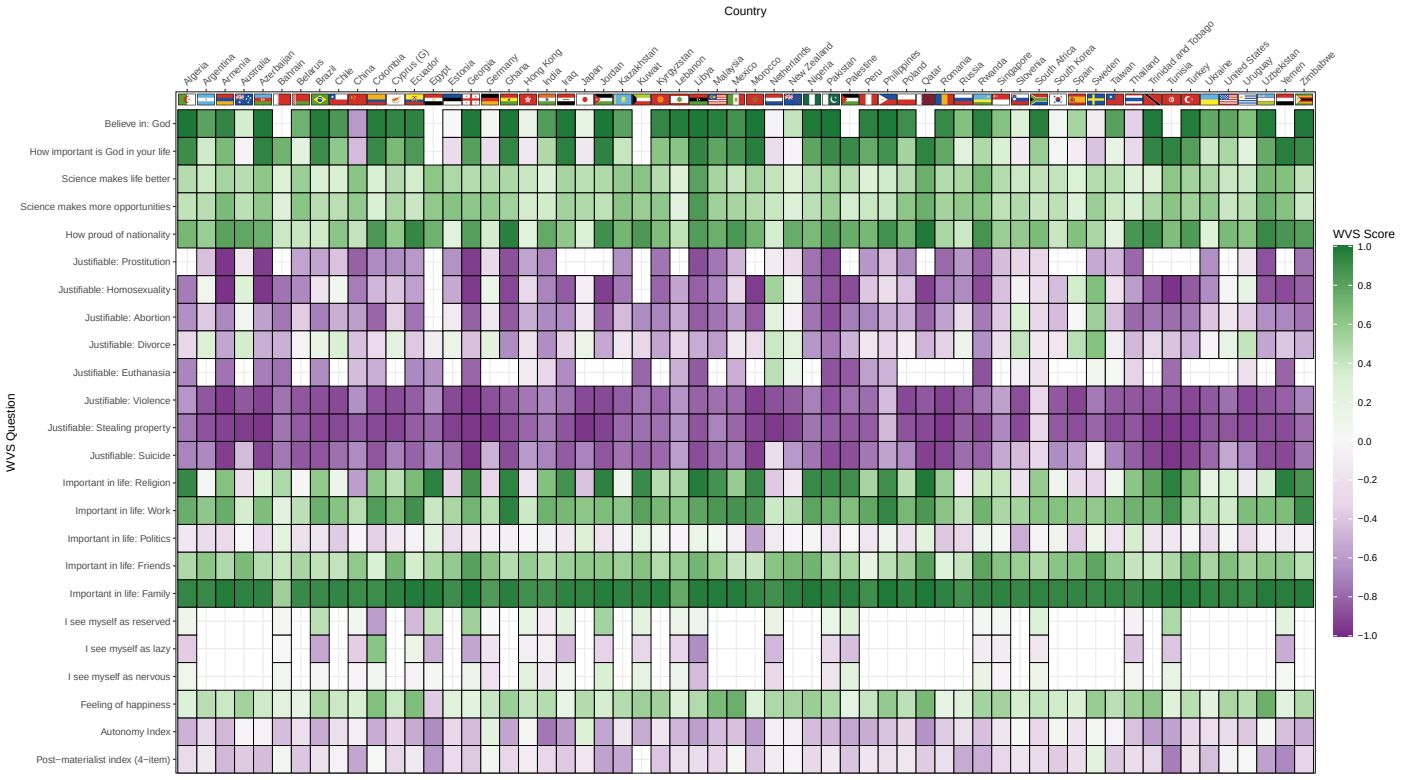

Figure 4.12. Matrix plot of the WVS Score for the selected questions. Each row is a WVS question, and each column is a country. The color of each represents the corresponding WVS Score.

an WVS Score is not available, so we consider only the pairs with valid scores both for online and offline.

In Figure 4.13 we present a plot of the percentage agreement for all the WVS/Inquiry pairs (rows), for each one of the four types of models (columns). We observe that, in general, the agreement is very high, in some cases having a 100% agreement score. Even though having also high percentages, the tweets-only models (first two rows), in general, have a lower agreement than the Wikipedia + tweets models (last two rows). As discussed before, the tweets-only models has a scarcity of data, so a single country with a disagreement will have high impact on the final percentage.

We highlight the "Prostitution" and the "Autonomy Index" as two questions with low agreement (below 25%). This might be an indication that the words being utilized in the OVI are not the ideal choices to capture the equivalent WVS questions. On the other hand we have many questions, like "Believe in god", "Abortion", "Friends", "Family", and others, with very high agreement scores (above 90%).

We showed that the online values methodology was able to capture, at least, the

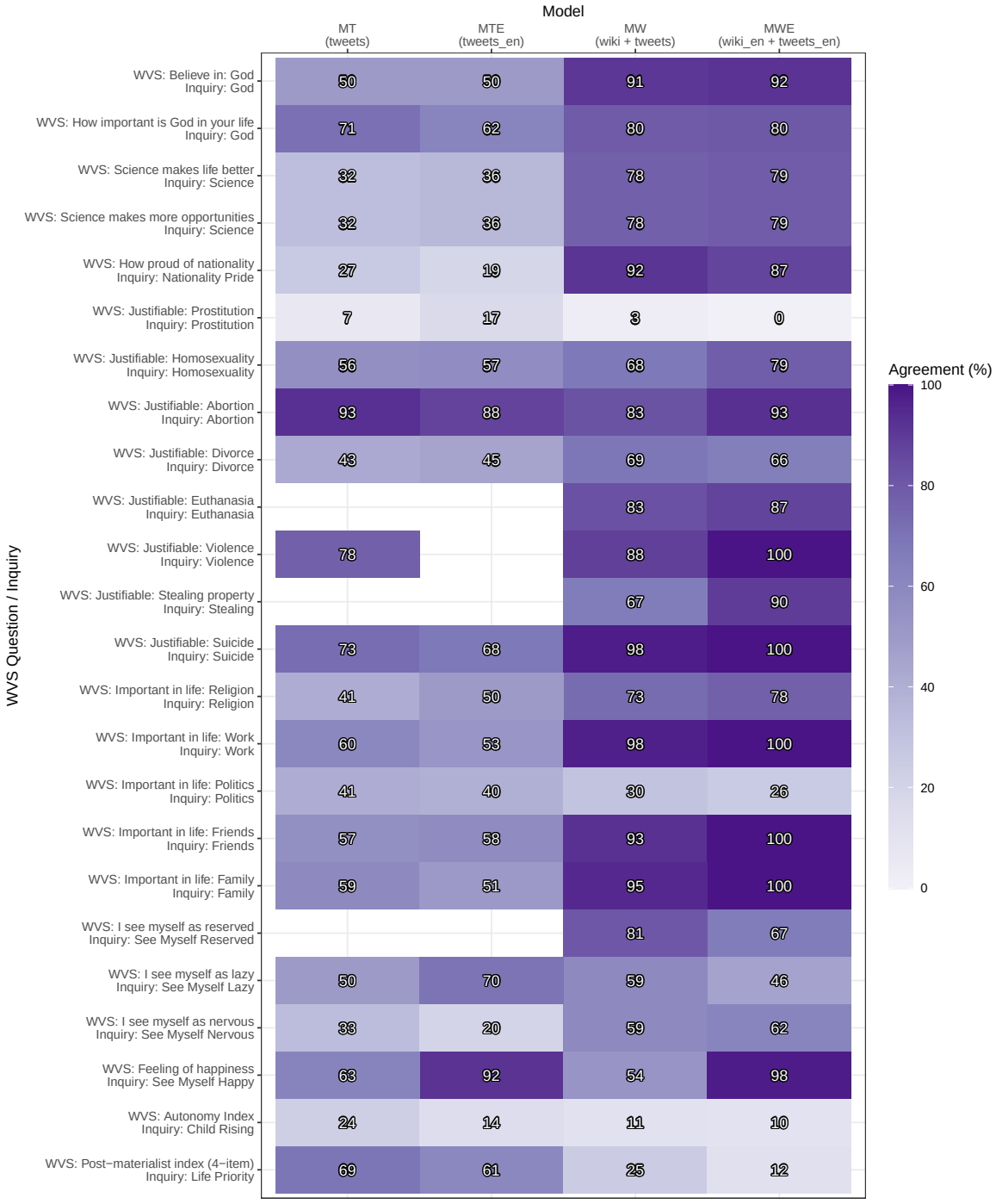

Figure 4.13. Agreement percentage between the association scores of the inquiries and the WVS Score, for all the four types of models.

signal of the corresponding offline value. In the following section we will take into consideration the ranking of the countries.

4.3.3.2 Online-Offline Correlation

We want to analyze now rather the online values methodology is able to capture the same *strength* as the offline value. We want to know if countries with lower WVS Score for a question will also have a lower inquiry association score, and vice-versa. Will the ranking of countries of the online values be the same as the ranking for the offline values?

In order to investigate that we will measure the Pearson Correlation coefficient (and the corresponding p-value) for all the pairs of WVS Question and Inquiry from Table 4.3, for all the four types of model. A table plot with all the correlation values is presented in Figure 4.14. Each row is a WVS Question and Inquiry combination, and each column is a model type. The cell label shows the actual correlation value, being the color a visual representation of the same (red is negative, and blue is positive), and the size of the font a representation of three categories of p-value ($p \leq 0.05$, $p \leq 0.10$, and $p > 0.10$).

We observe that the three religion-related questions have the highest and more consistent correlations: "Believe in God", "How important is God in your life?", and "Important in life: Religion". Curiously, despite the scarcity of the tweets-only models, they presented stronger relationship for the religion questions, having correlations as high as 0.69.

The two science questions presented an overall positive correlation, but with no significance ($p - value > 0.10$), with the exception of the question "Science makes life better", which had a positive significant correlation of 0.44 for the MT model. A similar pattern, but with negative correlation is observed for the "Justifiable: Prostitution" question, which have a negative correlation for most of the models, having a significant negative correlation of -0.32 for the MWE model. The negative correlation is probably relate to it also having a low agreement (as seen in the previous section).

We notice that both "Homosexuality" and "Suicide" had no significant correlation for the tweets-only models, while having significant positive correlations in the Wikipedia + tweets model. In these cases, the scarcity of data of the tweets models might have influence on the power of the correlation, since they have less data points to be used in the calculation of the correlation.

Interestingly, the "Abortion" question had a discrepancy: positive mildly significant ($p - value \leq 0.10$) correlation for the tweets model, and a low negative mildly significant correlation for the Wikipedia + tweets models. This is an indication that for some themes, there might be stronger differences between the encyclopedic text and the public opinion discourse. For instance, the texts about abortion in Wikipedia

Figure 4.14. Correlation matrix between the association scores of the inquiries and the WVS Scores, for all the four types of models.

might bring a neutral trend that is not present in Twitter, where there are probably a strong polarized discourse defending or attacking abortion.

Overall, the MW model had the best performance, having 5 questions with sig-

nificant ($p - value \leq 0.05$) positive correlation, followed by MT and MWE, both with 4 questions with significant positive correlation. Even though having less positive results, the tweets-only models, when having a significant result, have stronger correlation than the Wikipedia + tweets, taking for example the "Important in life: Religion" question, that have a correlation of 0.54 for the MT model, and correlations of 0.40 and 0.39 for the MW and MWE models. The advantage of the Wikipedia + tweets models is increasing the vocabulary and, consequently, increasing the number of data points, bringing strength for the correlation calculation, with the disadvantage of having influence of a neutral encyclopedic text, which might influence on having a weaker correlation. Differently, the tweets-only models have the advantage of having influence only from what people expresses, which will probably capture a stronger relationship and correlation, with the disadvantage of having a more limited vocabulary, which might make impossible to calculate the inquiry association scores in some cases.

We now take one of the values and analyze it in more detail: Religion. We present in Figure 4.15 a scatter plot of the countries for the four models of word embeddings for the WVS question "Important in life: Religion" with its corresponding "Religion" inquiry. First, as previously mentioned and discussed in this work, there is a data scarcity for the tweets-only model (facets in the top of the image), compared to the Wikipedia-based models (facets in the bottom). In the case of this value, all of the models presented a positive significant correlation ($p - value < 0.10$). It is interesting to see some consistencies between the models (like previously analyzed in Figure 4.8). For instance, Spain is consistently in the left-bottom of the scatter plot, presenting itself as one of the least religious countries in our analyzes. Regarding the language, we notice that, even though having some clusters (indicated by a group of country-points of the same color close to each other), we observe that there are countries with the same base-language model, with very different online values, such as Spain and Ecuador (in both facets of the left), indicating that language is not the only factor that explains the online value.

In the end, we show that the online values calculated by our methodology have, indeed, strong correlation with the corresponding offline values in some cases, particularly for religion-related values. Next, we will use other offline measurements besides the WVS Score to analyze rather they can explain the online values.

4.3.3.3 Offline Indicators

We observed that the association scores of the inquiries varies among the countries. In our final analysis we want to understand the factors that influence the online values

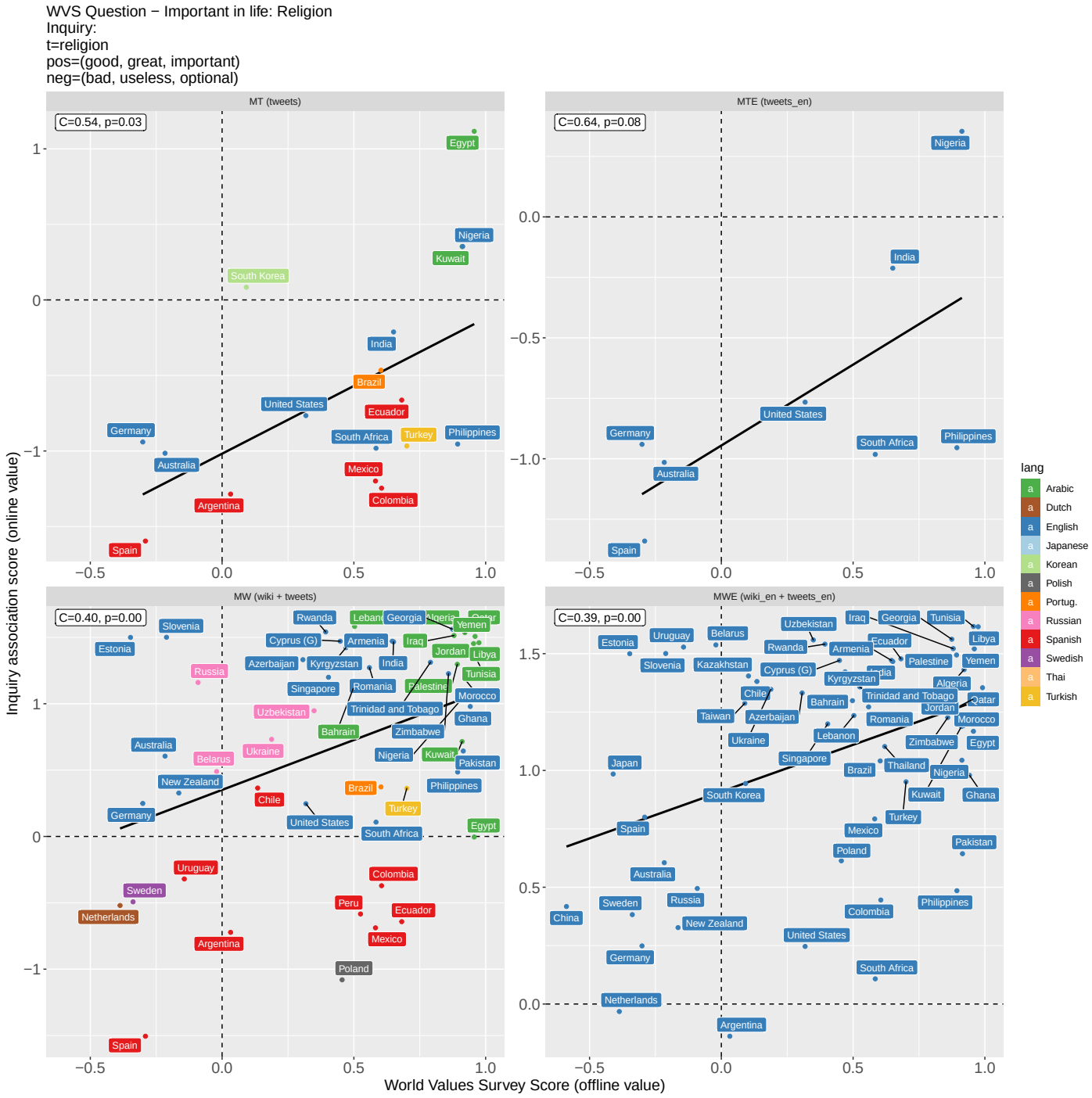

Figure 4.15. Scatter plot of countries, regarding the WVS Score for the question "Important in life: Religion" and the online inquiry association score, for the four models of word embedding. C is the spearson correlation of the points, and p is the corresponding p-value. The color scale is to differentiate the language. The horizontal and vertical dashed lines are guides for the online and offline scores, both centered at zero.

around the world. We investigate the influence of other external variables on the online values measured by our Inquiry methodology.

The list of variables we use is inspired by Ballatore et al. [12]. We use socioeconomic indicators related to the population of the country, the economy (e.g. GDP), and Internet infrastructure (e.g. Internet penetration) retrieved from World Bank [27]. We also include variables related to scientific publications, retrieved from a dataset of the SciMago research group [28], such as number of citable documents and number of citations of the countries. We also include a categorical variable indicating the region of the country in the World, retrieved from United Nations [29], which categorizes countries in 17 regions [30]. Due to consistency to the World Values Survey data and our Twitter dataset, we use data from 2014. All the variables are listed in Table 4.5.

Variable	**Data Source**	**Year**
WVS Score	World Values Survey	-
Population, total	World Bank	2014
GDP (current US$)	World Bank	2014
International tourism, receipts (current US$)	World Bank	2014
Individuals using the Internet (% of population)	World Bank	2014
Secure Internet servers	World Bank	2014
Scimago - Citable documents	SciMago	2014
Scimago - Citations	SciMago	2014
Scimago - Citations per document	SciMago	2014
Scimago - Documents	SciMago	2014
Scimago - H index	SciMago	2014
Scimago - Self-citations	SciMago	2014
Sub-region Name	United Nations	-

Table 4.5. List of variables utilized by the regression models.

We build several linear regression models (LM) predicting the OVI (association score representing a certain online value) using the socioeconomic indicators as variables. The linear model is defined for a certain embedding model m_c of a country c and an specific inquiry, defined by the target word w, positive attribute words A, and negative attribute words B, associated to a World Values Survey question q. The

[27] `https://data.worldbank.org`
[28] `http://www.scimagojr.com`
[29] `https://unstats.un.org/unsd/methodology/m49/overview`
[30] Considering the countries of our dataset, only 14 regions are present.

model is represented as following:

$$\begin{aligned} OVI_{m_c,w,A,B} = \alpha + & \beta_1 \cdot WVS_{q,c} + \beta_2 \cdot \text{Population}_c + \beta_3 \cdot \text{GDP}_c + \\ & \beta_4 \cdot \text{Int.TourismReceipt}_c + \beta_5 \cdot \text{InternetPenetration}_c + \\ & \beta_6 \cdot \text{SecureInternetServers}_c + \beta_7 \cdot \text{CitableDocuments}_c + \\ & \beta_8 \cdot \text{Citations}_c + \beta_9 \cdot \text{CitationsPerDocument}_c + \beta_{10} \cdot \text{Documents}_c + \\ & \beta_{11} \cdot \text{HIndex}_c + \beta_{12} \cdot \text{SelfCitations}_c + \beta_{13} \cdot \text{SubRegion}_c \end{aligned} \tag{4.9}$$

Since there is a scarcity of data in the tweets-only models (MT and MTE), resulting in a low number of country data points for the linear models, we choose not to use them for these analysis. We build in total 48 linear models, one for each inquiry and question from Table 4.3, both for the MW and MWE models. We show the results of the linear models in a grid plot in Figures 4.16 (MW) and 4.17 (MWE). Each line contains the result for one linear model, and each column is the estimate for the intercept (α) and the β values of each one of the variables. The shape and color of the point represents rather it is a positive or negative value, and we show the points only for the estimates with $p - value < 0.05$. The red number label in the left part of the grid is the Adjusted R^2 of the linear model, and the asterisk is marked in the statistically significant models ($p - value < 0.05$).

First, we observe that there are some LMs with high quality fit, particularly for the MWE models (Figure 4.17). From the 24 LMs of MWE, 17 are statiscally significant, being six of them with an Adjusted R^2 higher than 70%. For instance, the linear model for the online value of Euthanasia can have 96% of its variation explained by the variables, and the online value of "See myself reserved" can have 94% of its variation explained.

Analyzing now the importance of the variables, it is interesting to observe that the WVS score is rarely significantly correlated with the association score. This indicates that that are other external factors besides the offline value from the World Values Survey that have influence on the online value. Particularly, the "Secure Internet Servers" variable is frequently correlated with the online value (8 LMs out of 24 in MWE). This result suggests that the digital infrastructure of a country can have influence on the measures of online values. This can also be related to the fact of the online measurement being biased towards Internet and Twitter users.

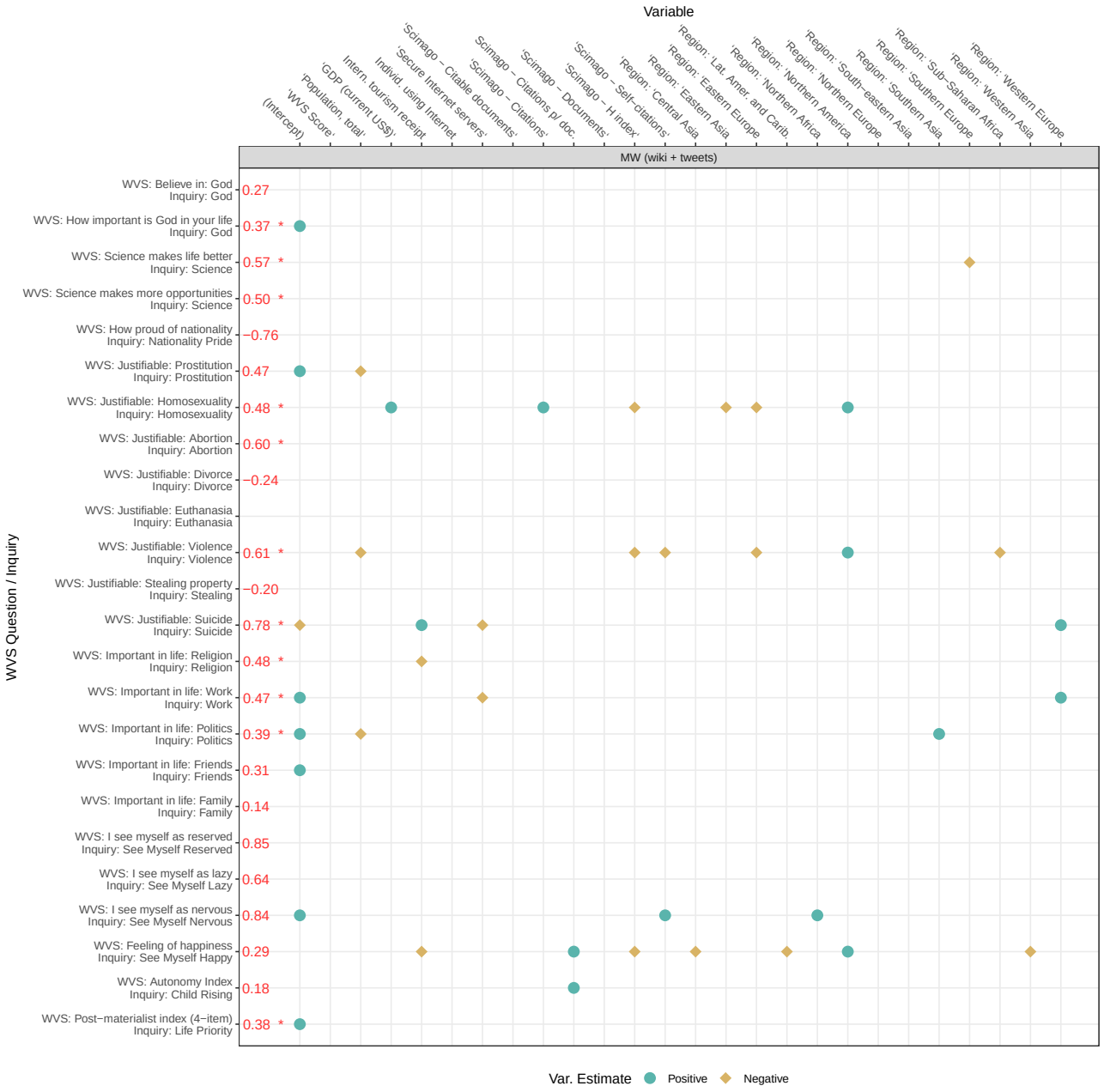

Figure 4.16. Variable estimates for the linear regression models of the online inquiries in relation to offline indicator, using the MW models. Red labels in the left indicate the Adjusted R^2, and the asterisks indicates statistical significance ($p - value < 0.05$).

4.4 Conclusion

We proposed here a methodology to measure human values using word embedding models. Our analysis focused on comparing cultural differences between countries,

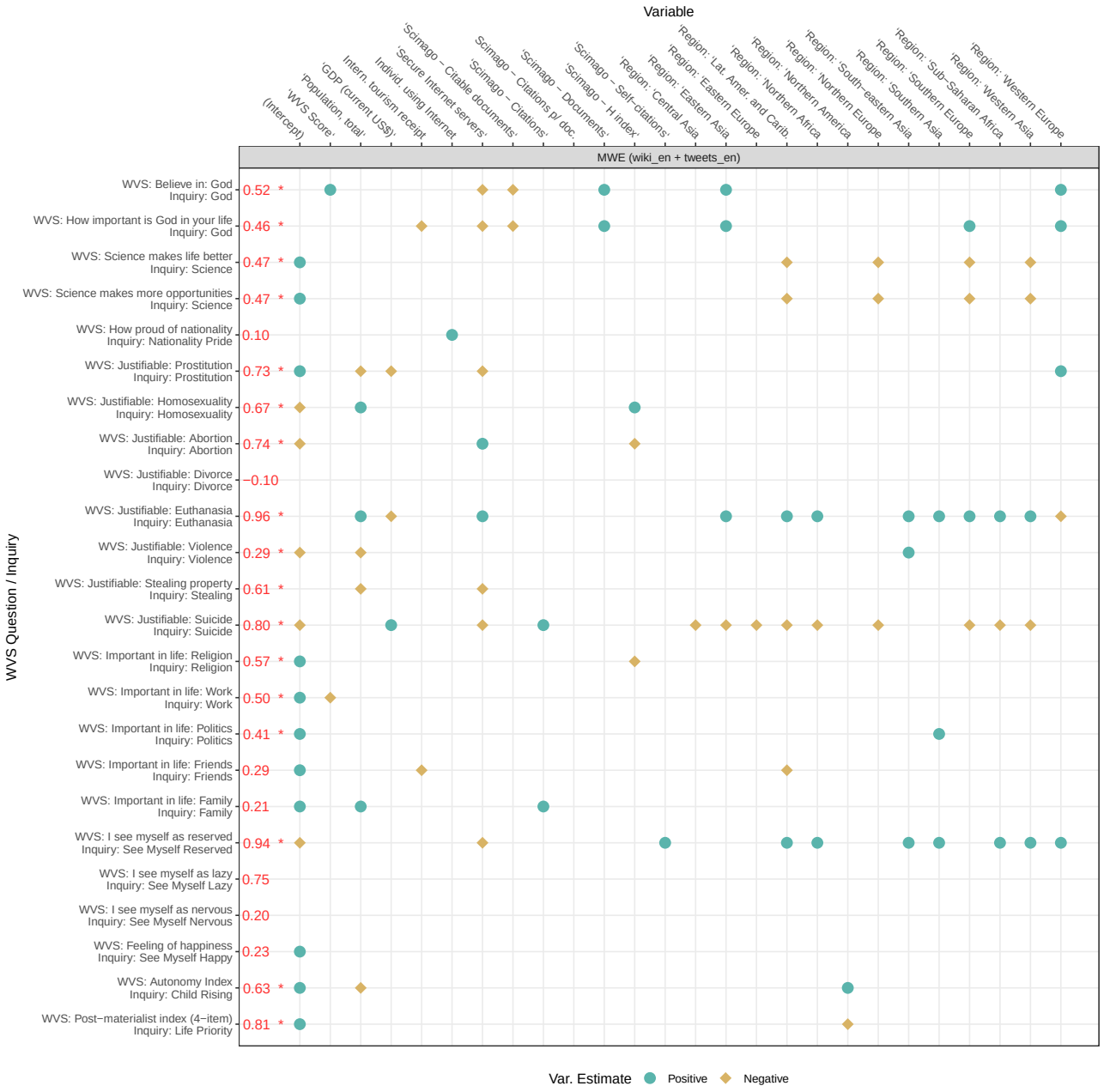

Figure 4.17. Variable estimates for the linear regression models of the online inquiries in relation to offline indicator, using the MWE models. Red labels in the left indicate the Adjusted R^2, and the asterisks indicates statistical significance ($p - value < 0.05$).

using written text from online communities. The methodology allows one to create an Online Values Inquiry (OVI), which is a set of words used to calculate distances in the word embedding model, designed to capture specific human values. We evaluate our methodology by creating models using Wikipedia and Twitter data for more than 50

countries, and designing 24 OVIs inspired by the World Values Survey.

Our results showed that the inquiries are capable of capturing differences between countries. Some online values are very diverse, having some countries with high positive agreement scores and others with low negative score. There are also online values with a more homogeneous trend, having almost all the countries with a positive or negative score, while still having intrinsic differences on the power of the agreement.

By clustering countries with similar online values and creating a bidimensional cultural map, we show that using a generic dimensionality reduction technique is possible, specially when having good quality Inquiries. However, if the goal is to capture other offline cultural maps with specific semantic dimensions, a factor analysis presented better results considering the similarity with the original offline map.

When comparing the online values (measured using OVI) with the offline values (measured with the World Values Survey), we show that our methodology was able to capture the signal of the values, meaning that an offline overall positive agreement for a certain value will also have an overall positive score online. Next, comparing the actual power of the values and the corresponding ranking of the countries, we show that there is a strong positive correlation between the online and the offline for some human values, specially for the inquiries related to religion. Finally, we create regression models with other offline socioeconomic indicators to evaluate their correlation with the OVI, and show that geographical location and the digital infrastructure of the countries are among one of the strongest characteristics related with the online value score.

We presented a robust and flexible framework that allows people to measure values online, and we believe that it can be explored and improved in several ways. First, the list of Online Values Inquiries could be extended, allowing people to measure other online human values. It is also possible to include more countries in our study, that would not only increase the international coverage, but also include more data points for the correlation and regression analysis. Another possibility of future work is to use other embedding algorithms besides word2vec, like GloVe, FastText or BERT. These algorithms could be compared to evaluate rather they differ on the online values measurement, and also on their performances when being utilized to measure online values. Finally, it would be interesting to make a temporal analysis of the evolution of the online values.

Chapter 5

Concluding Remarks

In this dissertation we studied whether and how it is possible to measure social characteristics from several countries using different online sources of information for different phenomena, by using many computational techniques. Our contributions are related in the sense of utilizing international online data to calculate a digital indexes, but they are independent in the sense of analyzing two online social networks (Google+ and Twitter) and two social attributes (gender gap and values).

First, in Chapter 1 we introduced our research topic, discussing the influence of Internet on culture, and how a online social imaginary is being constructed. Next, in Chapter 3 we presented our online gender gap work, which is based on the paper "International Gender Differences and Gaps in Online Social Networks" [78], previously published in the Proceedings of the 6th International Conference on Social Informatics (SocInfo 2014).

Finally, Chapter 4 we presented our work in progress about online values. We did a literature review of publications comparing online and offline data and also papers that study values using other approaches. After that, we conceptualized some important terms for our research, described the online platform being studied (Twitter), and the World Values Survey. Then, we described our whole methodology, including data collection, pre-processing, location classification, word embedding model training, and our proposed online values measurement technique.

The hypotheses related to online gender gaps were both confirmed. By calculating the online Gender Ratio for all the metrics and countries (Figure 3.1) we show that different countries have different female/male ratios, confirming **H1**. Next, comparing the online gender ratios with the offline metrics of the countries (Figure 3.3), we observe that there is indeed significative (p-value < 0.05) correlation, either positive (e.g. GR of the number of users) or negative (e.g. GR of the Recriprocity) depending on the

metric, confirming **H2**.

Following, the three hypotheses related to the online human values were also confirmed. By calculating the association scores for the 24 OVIs, considering the four types of model and all the countries, we observe that there is indeed a difference between the human values (i.e. some values are mostly positive, others are mostly negative, and there are also those that are heterogeneous), confirming **H3** (Figure 4.4), and at the same time there is a difference between countries (for the same value countries have higher or lower scores), confirming **H4** (Figures 4.4, 4.5, 4.6, 4.7). Finally, when comparing the online values with the scores from World Values Survey, we observe that there is a strong agreement in relation to the signal (Figure 4.13) and in relation to the power ranking for *some* of the values (specially religion-related ones), partially confirming **H5**.

Our results show that both of the methodologies are capable of measuring cultural traits from online environments and comparing these characteristics between different social groups. In our case we adopt an international approach, applying the analysis in the country level, but we believe that our methodologies are generic and adjustable to allow one to compare any compilation of social groups, either in the geographical level (e.g. cities, neighborhoods) or other socioeconomic factors (e.g. income and wealth).

On the other hand, the techniques presented here are inadequate for tracking social traits in the individual level. This is a limitation, but we advocate that this is also *beneficial* for the society. With the increasing concerns of the hazards and harmful usage of social tracking and surveillance, empowered by the scope and the amount of data being shared on the Internet, we see that it is worthwhile to provide methods that allows researchers to study social behaviour while preserving anonymity and privacy of the Internet user.

Since we are using online data from specific sources (Google+ and Twitter) we acknowledge that our findings will be dependant on the context of these online social networks. In the same way that the Internet is actually *part* of the "offline world", each platform and website is also part of the Internet, having its own market niche and intrinsicalities. We speculate that more general findings, such as countries with lower Gender Gap offline having lower gender gap regarding online number of users, or the fact of religion-related values being easier to capture online, will be more consistent and present similar results in other websites. On the other hand, findings related to popular political debate, such as the results regarding abortion and homosexuality, will be highly dependent on the political views of the users of the platform, and also on the time period of the data collection.

Our findings corroborate with the idea that the online and the offline are highly

interconnected. For instance, we observe that countries with higher disparity on the number of men and women online are also countries with a higher gender gap. When looking at online human values, we show that online religiousness is highly correlated with the equivalent offline manifestation of religion.

At the same time, we also observe some divergences between the online and the offline measurement. In our online gender gap study, we show that women in countries with higher offline gender gap have actually more followers than men online. Regarding online human values, we have cases like the "Homosexuality" value in Argentina, which is known to be one of the most advanced countries towards LGBT rights, but had a high disagreement score online. These cases are important to highlight that, even though being connected with the offline, the online behaviour might have its own peculiarities and underlying phenomena that should be taken into account when being studied.

The results of the correlation between online and offline values suggests that some online human values are worthy of being scrutinized. We believe that studies towards religion could be extended, not only related to values, but also to analyze hate speech discourse against religious minorities in different countries of the world. Other values like "Homosexuality" could also be enhanced, particularly to understand the reasons for the disparities between online and offline, and also to analyze the online discrimination discourse. More generally, values that are more closely related to other political discourses, like "Abortion", seems to be better capture by the OVI methodology.

Differently, values linked to "well being" and feelings (e.g. "Feeling of happiness" and "I see myself lazy") had lower agreement scores and correlation, indicating that they are probably harder to be measured. This might be related to one of the limitations of the OVI methodology: it relies on aggregated information. These human values would probably be better captured with a technique the is capable of operating on the individual level.

Considering the discussion and findings of our work, we emphasize our conviction that the online and the offline should be considered as two spaces from the same world, having at the same time closely related experiences, and their own idiosyncrasies and cultural manifestations. We should look the Internet as an extra environment of human interactions, that will not only have influence from the offline spaces, but will also influence these same offline spaces. The online and the offline "worlds" have a symbiotic relationship, worthy of being studied.

We believe that the study presented in this dissertation shows the achievability of studying the *online social imaginary*. We develop and present methodological frameworks for capturing social and cultural traits from the online environment, by collecting

and processing data from different online sources. The international characteristic of the Internet is valuable, presenting itself as a dynamic and diverse environment that can be studied by researchers from different fields, so that it will give us insights and better comprehension of the online culture from people of distinct regions of the world.

Bibliography

[1] Abitbol, J. L., Karsai, M., Magué, J.-P., Chevrot, J.-P., and Fleury, E. (2018). Socioeconomic dependencies of linguistic patterns in twitter: A multivariate analysis. In *Proceedings of the 2018 World Wide Web Conference*, WWW '18, pages 1125--1134, Republic and Canton of Geneva, Switzerland. International World Wide Web Conferences Steering Committee.

[2] Althoff, T., Sosič, R., Hicks, J., C King, A., Delp, S., and Leskovec, J. (2017). Large-scale physical activity data reveal worldwide activity inequality. *Nature*, 547.

[3] Analytics, P. (2009). Twitter Study - August 2009. *Fast Company*.

[4] Andrews, J., Hinton, L., and Ash, S. (2017). Women in tech: Time to close the gender gap. Technical report, PwC UK Research.

[5] Antenucci, D., Cafarella, M., Levenstein, M., Ré, C., and Shapiro, M. D. (2014). Using social media to measure labor market flows. Technical report 20010, National Bureau of Economic Research.

[6] Anzia, S. F. and Berry, C. R. (2011). The jackie (and jill) robinson effect: Why do congresswomen outperform congressmen? *American Journal of Political Science*, 55:478--493.

[7] Aramaki, E., Maskawa, S., and Morita, M. (2011). Twitter catches the flu: Detecting influenza epidemics using twitter. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing*, EMNLP '11, pages 1568--1576, Stroudsburg, PA, USA. Association for Computational Linguistics.

[8] Arrington, M. (2013). Odeo Releases Twttr. TechCrunch. `https://techcrunch.com/2006/07/15/is-twttr-interesting/`. [Online; accessed 11-March-2019].

[9] Association, W. V. S. (2013). World Values Survey - Findings and Insights. `http://www.worldvaluessurvey.org/WVSContents.jsp?CMSID=Findings`. [Online; accessed 13-Sep-2020].

[10] Avruch, K. and of Peace, U. S. I. (1998). *Culture & Conflict Resolution*. Cross-Cultural Negotiation Books. United States Institute of Peace Press. ISBN 9781878379825.

[11] Baghal, T. A., Sloan, L., Jessop, C., Williams, M. L., and Burnap, P. (2019). Linking twitter and survey data: The impact of survey mode and demographics on consent rates across three uk studies. *Social Science Computer Review*, 0(0):0894439319828011.

[12] Ballatore, A., Graham, M., and Sen, S. (2017). Digital hegemonies: The localness of search engine results. *Annals of the American Association of Geographers*, 107(5):1194–1215.

[13] Bastos, M., Mercea, D., and Baronchelli, A. (2018). The geographic embedding of online echo chambers: Evidence from the brexit campaign. *PLOS ONE*, 13(11):1–16.

[14] BBC (2010). Argentine Senate backs bill legalising gay marriage. BBC. `https://www.bbc.com/news/10630683`. [Online; accessed 23-Aug-2020].

[15] BBC (2013). Nigerian parliament bans same-sex marriage. BBC. `https://www.bbc.com/news/world-africa-22722789`. [Online; accessed 23-Aug-2020].

[16] Bello, M. and DiBlasio, N. (2013). Twitter: The new face of crime. USA Today. `https://www.usatoday.com/story/news/nation/2013/09/29/twitter-crime-dark-side/2875745/`. [Online; accessed 12-March-2019].

[17] Bimber (2000). Measuring the Gender Gap on the Internet. *Social Science Quarterly*, 81(3).

[18] Bollen, J., Mao, H., and Pepe, A. (2011a). Modeling public mood and emotion: Twitter sentiment and socio-economic phenomena. In *ICWSM*.

[19] Bollen, J., Mao, H., and Zeng, X.-J. (2011b). Twitter mood predicts the stock market. *J. Comput. Science*, 2(1):1–8.

[20] Bolukbasi, T., Chang, K.-W., Zou, J., Saligrama, V., and Kalai, A. (2016). Man is to computer programmer as woman is to homemaker? debiasing word embeddings. In *Proceedings of the 30th International Conference on Neural Information Processing Systems*, NIPS'16, pages 4356--4364, USA. Curran Associates Inc.

[21] Bond, B. J. (2009). He posted, she posted: Gender differences in self-disclosure on social network sites. *Rocky Mountain Communication Review*, 6(2):29--37. ISSN 15426394.

[22] Bortnik, V. and Sementsov, S. (2008). Are all equal before the law? https://web.archive.org/web/20090227045600/http://www.pride.by/en/show.php?id=33. [Online, archived; accessed 23-Aug-2020].

[23] Breen, A. (2015). I, Fashion: How technology is changing the way we dress. The Globe and Mail. https://www.theglobeandmail.com/life/fashion-and-beauty/fashion/i-fashion-how-technology-is-changing-the-way-we-dress/article26028412/. [Online; accessed 24-Aug-2019].

[24] Caliskan, A., Bryson, J., and Narayanan, A. (2017). Semantics derived automatically from language corpora contain human-like biases. *Science*, 356(6334):183--186. ISSN 0036-8075.

[25] Center, W. M. (2014). The status of women in the u.s. media 2014. http://www.womensmediacenter.com/page/-/statusreport/WMC-2014-status-women-with-research.pdf.

[26] Cha, M., Haddadi, H., Benevenuto, F., and Gummadi, K. P. (2010). Measuring user influence in twitter: The million follower fallacy. In *in ICWSM '10: Proceedings of international AAAI Conference on Weblogs and Social*.

[27] Chen, J., Hsieh, G., Mahmud, J. U., and Nichols, J. (2014). Understanding individuals' personal values from social media word use. In *Proceedings of the 17th ACM Conference on Computer Supported Cooperative Work & Social Computing*, CSCW '14, pages 405--414, New York, NY, USA. ACM.

[28] Cheong, M. and Lee, V. C. S. (2011). A microblogging-based approach to terrorism informatics: Exploration and chronicling civilian sentiment and response to terrorism events via twitter. *Information Systems Frontiers*, 13(1):45--59. ISSN 1572-9419.

[29] Clement, J. (2019). Twitter: number of monthly active users 2010-2019. Statista. https://www.statista.com/statistics/282087/number-of-monthly-active-twitter-users/. [Online; accessed 02-Feb-2020].

[30] Collier, B. and Bear, J. (2012). Conflict, criticism, or confidence: an empirical examination of the gender gap in wikipedia contributions. In *CSCW*, pages 383–392.

[31] Costa, P. T., Terracciano, A., and McCrae, R. R. (2001). Gender differences in personality traits across cultures: robust and surprising findings. *Journal of personality and social psychology*, 81(2):322--331. ISSN 0022-3514.

[32] Culotta, A. (2013). Lightweight methods to estimate influenza rates and alcohol sales volume from twitter messages. *Language Resources and Evaluation*, 47(1):217–238.

[33] Cunha, E., Magno, G., Gonçalves, M. A., Cambraia, C., and Almeida, V. (2014). How you post is who you are: Characterizing google+ status updates across social groups. In *Proceedings of the 25th ACM Conference on Hypertext and Social Media*, HT '14, pages 212--217. ACM.

[34] Di Fraia, G. and Missaglia, M. C. (2014). *The Use of Twitter In 2013 Italian Political Election*, pages 63--77. Springer International Publishing, Cham.

[35] Dutton, W. H. and Reisdorf, B. C. (2019). Cultural divides and digital inequalities: attitudes shaping internet and social media divides. *Information, Communication & Society*, 22(1):18–38.

[36] Fatehkia, M., Kashyap, R., and Weber, I. (2018). Using facebook ad data to track the global digital gender gap. *World Development*, 107:189 – 209. ISSN 0305-750X.

[37] Feingold, A. (1994). Gender differences in personality: a meta-analysis. *Psychological bulletin*, 116(3):429--456.

[38] Fiorio, L., Abel, G., Cai, J., Zagheni, E., Weber, I., and Vinué, G. (2017). Using twitter data to estimate the relationship between short-term mobility and long-term migration. In *Proceedings of the 2017 ACM on Web Science Conference*, WebSci '17, pages 103--110, New York, NY, USA. ACM.

[39] Fischer, R. and Schwartz, S. (2011). Whence differences in value priorities?: Individual, cultural, or artifactual sources. *Journal of Cross-Cultural Psychology*, 42(7):1127–1144.

[40] Floridi, L. (2014). *The Onlife Manifesto: Being Human in a Hyperconnected Era*. Springer International Publishing. ISBN 9783319040929.

[41] Formichelli, J. (2016). Urban planning tools synthesize and collect data to improve the quality of city life. Phys.org. `https://phys.org/news/2016-06-urban-tools-quality-city-life.html`. [Online; accessed 24-Aug-2019].

[42] Garcia, D., Mitike Kassa, Y., Cuevas, A., Cebrian, M., Moro, E., Rahwan, I., and Cuevas, R. (2018). Analyzing gender inequality through large-scale facebook advertising data. *Proceedings of the National Academy of Sciences*, 115(27):6958--6963. ISSN 0027-8424.

[43] Garcia, D., Weber, I., and Garimella, V. R. K. (2014). Gender asymmetries in reality and fiction: The bechdel test of social media. In *ICWSM*.

[44] García-Gavilanes, R., Mejova, Y., and Quercia, D. (2014). Twitter ain't without frontiers: Economic, social, and cultural boundaries in international communication. In *Proceedings of the 17th ACM Conference on Computer Supported Cooperative Work and Social Computing*, CSCW '14, page 1511–1522, New York, NY, USA. Association for Computing Machinery.

[45] Garcia-Gavilanes, R., Quercia, D., and Jaimes, A. (2013). Cultural dimensions in twitter: Time, individualism and power. In *International AAAI Conference on Web and Social Media*.

[46] Garcia Gavilanes, R. O. (2013). On the quest of discovering cultural trails in social media. In *Proceedings of the Sixth ACM International Conference on Web Search and Data Mining*, WSDM '13, page 747–752, New York, NY, USA. Association for Computing Machinery.

[47] Ginsberg, J., Mohebbi, M., Patel, R., Brammer, L., Smolinski, M., and Brilliant, L. (2009). Detecting influenza epidemics using search engine query data. *Nature*, 457:1012–1014. doi:10.1038/nature07634.

[48] Gjoka, M., Kurant, M., Butts, C. T., and Markopoulou, A. (2009). A walk in facebook: Uniform sampling of users in online social networks. *CoRR*, abs/0906.0060.

[49] Globerson, A., Chechik, G., Pereira, F., and Tishby, N. (2007). Euclidean embedding of co-occurrence data. *J. Mach. Learn. Res.*, 8:2265--2295. ISSN 1532-4435.

[50] Graham, M. (2012). Geography/internet: Ethereal alternate dimensions of cyberspace or grounded augmented realities? *The Geographical Journal*, 179.

[51] Greenwald, A. G., McGhee, D. E., and Schwartz, J. L. K. (1998). Measuring individual differences in implicit cognition: the implicit association test. *Journal of Personality and Social Psychology*, 74(6):1464–80.

[52] Grothaus, M. (2018). Twitter's Q3 earnings by the numbers. Fast Company. `https://www.fastcompany.com/90256723/twitters-q3-earnings-by-the-numbers`. [Online; accessed 01-March-2019].

[53] Group, M. M. (2019). World Internet Users and 2019 Population Stats. Internet World Stats. `https://www.internetworldstats.com/stats.htm`. [Online; accessed 02-Feb-2020].

[54] Guo, L., Zhang, D., Wu, H., Cui, B., and Tan, K. (2017). From raw footprints to personal interests: Bridging the semantic gap via trip intention aggregation. In *2017 IEEE 33rd International Conference on Data Engineering (ICDE)*, pages 123–126. ISSN 2375-026X.

[55] Hausmann, R., Tyson, L. D., Zahidi, S., and Editors (2013). The global gender gap report 2013. `http://www3.weforum.org/docs/WEF_GenderGap_Report_2013.pdf`.

[56] Hawelka, B., Sitko, I., Beinat, E., Sobolevsky, S., Kazakopoulos, P., and Ratti, C. (2014). Geo-located twitter as proxy for global mobility patterns. *Cartography and Geographic Information Science*, 41:260–271.

[57] Heil, B. and Piskorski, M. (2009). New twitter research: Men follow men and nobody tweets. `http://blogs.hbr.org/2009/06/new-twitter-research-men-follo/`.

[58] Hofstede, G., Hofstede, G., and Minkov, M. (2010). *Cultures and Organizations: Software of the Mind, Third Edition*. McGraw-Hill Education. ISBN 9780071770156.

[59] Hyde, J. S. (2005). The Gender Similarities Hypothesis. *American Psychologist*, 60(6):581--592.

[60] ILGA-Europe – the European Region of the International Lesbian, Gay, Bisexual, Trans and Intersex Association (2020). Country ranking - rainbow europe. `https://rainbow-europe.org/country-ranking#eu`. [Online; accessed 23-Aug-2020].

[61] Inglehart, R. (1997). *Modernization and Postmodernization: Cultural, Economic, and Political Change in 43 Societies*. Political science/sociology. Princeton University Press. ISBN 9780691011806.

[62] Inglehart, R. and Baker, W. E. (2000). Modernization, cultural change, and the persistence of traditional values. *American Sociological Review*, 65(1):19--51. ISSN 00031224.

[63] Inglehart, R., Haerpfer, C., Moreno, A., Welzel, C., Kizilova, K., Diez-Medrano, J., Lagos, M., Norris, P., Ponarin, E., Puranen, B., et al. (2014). World values survey: Round six - country-pooled datafile 2010-2014. Madrid: JD Systems Institute.

[64] Iosub, D., Laniado, D., Castillo, C., Fuster Morell, M., and Kaltenbrunner, A. (2014). Emotions under discussion: Gender, status and communication in online collaboration. *PLoS ONE*, 9(8):e104880.

[65] Jakobson, R. and Ruwet, N. (1969). *Essais de linguistique générale*. Arguments (Ed. de Minuit, Collection). Editions de Minuit.

[66] Joinson, A. N. (2008). Looking at, Looking Up or Keeping Up with People?: Motives and Use of Facebook. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '08, pages 1027--1036. ACM.

[67] Jr., G. B., de Castro, C. M., Izumi, C., and Sassaki, R. (2011). Parada Gay leva 4 milhões para a Paulista. Folha de São Paulo. `https://m.folha.uol.com.br/cotidiano/2011/06/935237-parada-gay-leva-4-milhoes-para-a-paulista.shtml`. [Online; accessed 23-Aug-2020].

[68] Kalimeri, K., Beiró, M. G., Delfino, M., Raleigh, R., and Cattuto, C. (2019). Predicting demographics, moral foundations, and human values from digital behaviours. *Computers in Human Behavior*, 92:428 – 445. ISSN 0747-5632.

[69] Klasen, S. (1999). Does gender inequality reduce growth and development? evidence from cross-country regressions. *Policy research report on gender and development. The World Bank*.

[70] Kwak, H., Lee, C., Park, H., and Moon, S. B. (2010). What is twitter, a social network or a news media? In *WWW*, pages 591–600.

[71] Lampos, V. and Cristianini, N. (2012). Nowcasting events from the social web with statistical learning. *ACM Trans. Intell. Syst. Technol.*, 3(4):72:1--72:22.

[72] Lavers, M. K. (2016). Argentina joins global LGBT rights initiative. Washington Blade. `https://www.washingtonblade.com/2016/03/24/argentina-joins-global-lgbt-rights-initiative/`. [Online; accessed 23-Aug-2020].

[73] Lazer, D., Kennedy, R., King, G., and Vespignani, A. (2014). The parable of google flu: Traps in big data analysis. *Science*, 343:1203–1205.

[74] Lebret, R. and Collobert, R. (2014). Word embeddings through hellinger pca. In Bouma, G. and Parmentier, Y., editors, *EACL*, pages 482–490. The Association for Computer Linguistics.

[75] Likert, R. (1932). *A Technique for the Measurement of Attitudes*. Number N° 136-165 in A Technique for the Measurement of Attitudes. publisher not identified.

[76] Macionis, J. (2016). *Sociology*. Pearson; 16 edition. ISBN 9780132372640.

[77] Magno, G., Comarela, G., Saez-Trumper, D., Cha, M., and Almeida, V. (2012). New kid on the block: exploring the google+ social graph. In *Proceedings of the 2012 ACM conference on Internet measurement conference*, IMC '12, pages 159--170, New York, NY, USA. ACM.

[78] Magno, G. and Weber, I. (2014). *International Gender Differences and Gaps in Online Social Networks*, pages 121--138. Springer International Publishing, Cham.

[79] Manjoo, F. (2017). How Netflix Is Deepening Our Cultural Echo Chambers. The New York Times. `https://www.nytimes.com/2017/01/11/technology/how-netflix-is-deepening-our-cultural-echo-chambers.html`. [Online; accessed 10-May-2019].

[80] Marques-Toledo, C. d. A., Degener, C. M., Vinhal, L., Coelho, G., Meira, W., Codeço, C. T., and Teixeira, M. M. (2017). Dengue prediction by the web: Tweets are a useful tool for estimating and forecasting dengue at country and city level. *PLOS Neglected Tropical Diseases*, 11(7):1–20.

[81] Matheson, R. (2018). Measuring the economy with location data. MIT News Office. `https://news.mit.edu/2018/startup-thasos-group-measuring-economy-smartphone-location-data-0328`. [Online; accessed 10-May-2019].

[82] McInnes, L., Healy, J., and Melville, J. (2018). Umap: Uniform manifold approximation and projection for dimension reduction.

[83] Mikolov, T., Chen, K., Corrado, G., and Dean, J. (2013a). Efficient estimation of word representations in vector space. *CoRR*, abs/1301.3781.

[84] Mikolov, T., Sutskever, I., Chen, K., Corrado, G., and Dean, J. (2013b). Distributed representations of words and phrases and their compositionality. In *Proceedings of the 26th International Conference on Neural Information Processing Systems - Volume 2*, NIPS'13, pages 3111--3119, USA. Curran Associates Inc.

[85] Minkov, M. (2007). *What Makes Us Different and Similar: A New Interpretation of the World Values Survey and Other Cross-Cultural Data*. Klasika y Stil Publishing House.

[86] Miritello, G., Lara, R., Cebrian, M., and Moro, E. (2013). Limited communication capacity unveils strategies for human interaction. *Sci. Rep.*, 3. Article.

[87] mjahr (2016). Never miss important Tweets from people you follow. Twitter. https://blog.twitter.com/official/en_us/a/2016/never-miss-important-tweets-from-people-you-follow.html. [Online; accessed 01-March-2019].

[88] Nations, U. (2013). World Abortion Policies 2013. https://www.un.org/en/development/desa/population/publications/pdf/policy/WorldAbortionPolicies2013/WorldAbortionPolicies2013_WallChart.pdf. [Online; accessed 23-Aug-2020].

[89] Nissim, M., van Noord, R., and van der Goot, R. (2019). Fair is better than sensational: Man is to doctor as woman is to doctor.

[90] O'Connor, B., Balasubramanyan, R., Routledge, B. R., and Smith, N. A. (2010). From tweets to polls: Linking text sentiment to public opinion time series. In *ICWSM*.

[91] Ojanperä, S., Graham, M., and Zook, M. (2019). The digital knowledge economy index: Mapping content production. *The Journal of Development Studies*, 0(0):1–18.

[92] Ottoni, R., Pesce, J. P., Las Casas, D., Franciscani Jr, G., Meira Jr, W., Kumaraguru, P., and Almeida, V. (2013). Ladies first: Analyzing gender roles and behaviors in pinterest. In *Proceedings of the Seventh International Conference on Weblogs and Social Media*, ICWSM '13.

[93] Paul T. Costa Jr., R. R. M. (2008). *The Revised NEO Personality Inventory (NEO-PI-R)*, pages 179–199. SAGE Publications Ltd.

[94] Pearson, K. (1901). LIII. On lines and planes of closest fit to systems of points in space.

[95] Pew Research Center's Forum on Religion and Public Life (2012). *The Global Religious Landscape: A Report on the Size and Distribution of the World's Major Religious Groups as of 2010*.

[96] Pitman, E. J. G. (1937). Significance tests which may be applied to samples from any populations. *Supplement to the Journal of the Royal Statistical Society*, 4(1):119--130. ISSN 14666162.

[97] Pratto, F., Stallworth, L. M., and Sidanius, J. (1997). The gender gap: Differences in political attitudes and social dominance orientation. *The british journal of social psychology*, 36(1):49--68. ISSN 2044-8309.

[98] Preis, T., Moat, H. S., Stanley, H. E., and Bishop, S. R. (2012). Quantifying the advantage of looking forward. *Nature Scientific Reports*, 2:350.

[99] Pride, W. (2017). Publicidad - World Pride Madrid 2017. Web oficial del Orgullo. World Pride Madrid 2017. `http://www.worldpridemadrid2017.com/publicidad`. [Online; accessed 23-Aug-2020].

[100] Quercia, D., Casas, D. B. L., Pesce, J. P., Stillwell, D., Kosinski, M., Almeida, V., and Crowcroft, J. (2012). Facebook and privacy: The balancing act of personality, gender, and relationship currency. In *ICWSM*.

[101] Quercia, D. and Sáez-Trumper, D. (2014). Mining urban deprivation from foursquare: Implicit crowdsourcing of city land use. *IEEE Pervasive Computing*, 13(2):30–36.

[102] Ribeiro, B. and Towsley, D. (2010). Estimating and sampling graphs with multidimensional random walks. In *Proceedings of the 10th ACM SIGCOMM Conference on Internet Measurement*, IMC '10, pages 390--403, New York, NY, USA. ACM.

[103] Rokeach, M. (1973). *The nature of human values*. Free Press.

[104] Sakaki, T., Okazaki, M., and Matsuo, Y. (2010). Earthquake shakes twitter users: Real-time event detection by social sensors. In *Proceedings of the 19th International Conference on World Wide Web*, WWW '10, pages 851--860, New York, NY, USA. ACM.

[105] Schmitt, D. P., Realo, A., Voracek, M., and Allik, J. (2008). Why can't a man be more like a woman? Sex differences in Big Five personality traits across 55 cultures. *Journal of Personality and Social Psychology*, 94(1):168--182.

[106] Schwartz, S. H. (1992). Universals in the content and structure of values: Theoretical advances and empirical tests in 20 countries. In Zanna, M. P., editor, *Advances in Experimental Social Psychology*, volume 25, pages 1 – 65. Academic Press.

[107] Shaban, H. (2019). Twitter reveals its daily active user numbers for the first time. The Washington Post. `https://www.washingtonpost.com/technology/2019/02/07/twitter-reveals-its-daily-active-user-numbers-first-time/`. [Online; accessed 04-Jul-2019].

[108] Silva, T. H., de Melo, P. O. S. V., Almeida, J. M., Musolesi, M., and Loureiro, A. A. F. (2014). You are what you eat (and drink): Identifying cultural boundaries

by analyzing food and drink habits in foursquare. In *Proceedings of the Eighth International Conference on Weblogs and Social Media, ICWSM 2014, Ann Arbor, Michigan, USA, June 1-4, 2014*.

[109] Sloan, L. and Morgan, J. (2015). Who tweets with their location?: Understanding the relationship between demographic characteristics and the use of geoservices and geotagging on twitter. *PloS one*, 10(11):e0142209–e0142209. ISSN 1932-6203.

[110] Spencer-Oatey, H. (2008). *Culturally speaking: culture, communication and politeness theory*. Continuum. ISBN 9780826493101.

[111] Spencer-Oatey, H. (2012). What is culture?: A compilation of quotations. Recommended.

[112] Staff, I. P. (2018). How The Internet has Changed the Way we Communicate. Incredible Planet. `https://incredibleplanet.net/internet-changed-way-communicate/`. [Online; accessed 24-Aug-2019].

[113] Studios, F. (2014). I, Fashion: How technology is changing the way we dress. Fast Company. `https://www.fastcompany.com/3038001/how-the-internet-has-changed-the-way-we-eat/`. [Online; accessed 24-Aug-2019].

[114] Szell, M. and Thurner, S. (2013). How women organize social networks different from men. *Scientific Reports*, 3. ISSN 2045-2322.

[115] Thelwall, M. (2008). Social networks, gender, and friending: An analysis of myspace member profiles. *JASIST*, 59(8):1321–1330.

[116] Tornes, A. (2017). Introducing Twitter premium APIs. Twitter. `https://blog.twitter.com/developer/en_us/topics/tools/2017/introducing-twitter-premium-apis.html`. [Online; accessed 04-April-2019].

[117] van der Maaten, L. and Hinton, G. (2008). Visualizing data using t-SNE. *Journal of Machine Learning Research*, 9:2579--2605.

[118] Wikipedia (2019). World Values Survey — Wikipedia, the free encyclopedia. `http://en.wikipedia.org/w/index.php?title=World%20Values%20Survey&oldid=885226660`. [Online; accessed 08-May-2019].

[119] Wikipedia contributors (2020a). Languages with official status in india — Wikipedia, the free encyclopedia. `https://en.wikipedia.org/w/index.php?`

title=Languages_with_official_status_in_India&oldid=938502640. [Online; accessed 02-Feb-2020].

[120] Wikipedia contributors (2020b). Lgbt rights in argentina — Wikipedia, the free encyclopedia. https://en.wikipedia.org/w/index.php?title=LGBT_rights_in_Argentina&oldid=972196525. [Online; accessed 23-August-2020].

[121] Wikipedia contributors (2020c). Lgbt rights in belarus — Wikipedia, the free encyclopedia. https://en.wikipedia.org/w/index.php?title=LGBT_rights_in_Belarus&oldid=973246568. [Online; accessed 23-August-2020].

[122] Wikipedia contributors (2020d). Lgbt rights in nigeria — Wikipedia, the free encyclopedia. https://en.wikipedia.org/w/index.php?title=LGBT_rights_in_Nigeria&oldid=967925279. [Online; accessed 23-August-2020].

[123] Wikipedia contributors (2020e). Lgbt rights in poland — Wikipedia, the free encyclopedia. https://en.wikipedia.org/w/index.php?title=LGBT_rights_in_Poland&oldid=973863469. [Online; accessed 23-August-2020].

[124] Wikipedia contributors (2020f). Lgbt rights in uzbekistan — Wikipedia, the free encyclopedia. https://en.wikipedia.org/w/index.php?title=LGBT_rights_in_Uzbekistan&oldid=957672101. [Online; accessed 23-August-2020].

[125] Wikipedia contributors (2020g). List of largest lgbt events — Wikipedia, the free encyclopedia. https://en.wikipedia.org/w/index.php?title=List_of_largest_LGBT_events&oldid=967617014. [Online; accessed 23-August-2020].

[126] Wikipedia contributors (2020h). South Africa — Wikipedia, the free encyclopedia. https://en.wikipedia.org/w/index.php?title=South_Africa&oldid=938819999. [Online; accessed 02-Feb-2020].

[127] WIN/Gallup International (2015). Losing our religion? two thirds of people still claim to be religious. https://www.gallup-international.bg/en/33531/losing-our-religion-two-thirds-of-people-still-claim-to-be-religious/. [Online; accessed 22-Aug-2020].

[128] Youyou, W., Kosinski, M., and Stillwell, D. (2015). Computer-based personality judgments are more accurate than those made by humans. *Proceedings of the National Academy of Sciences*, 112(4):1036--1040. ISSN 0027-8424.

[129] Z., Y. (2014). Building a complete Tweet index. Twitter. https://blog.twitter.com/engineering/en_us/a/2014/building-a-complete-tweet-index.html. [Online; accessed 04-April-2019].

[130] Zagheni, E., Garimella, V. R. K., Weber, I., and State, B. (2014). Inferring international and internal migration patterns from twitter data. In *WWW (Companion Volume)*, pages 439–444.

[131] Zhang, X. and Gloor, H. F. P. A. (2012). Predicting asset value through twitter buzz. *Advances in Intelligent and Soft Computing*, 113:23–34.