International Gender Differences and Gaps in Online Social Networks

Gabriel Magno
Qatar Computing Research Institute (QCRI)
Doha, Qatar
magno@dcc.ufmg.br

Ingmar Weber
Qatar Computing Research Institute (QCRI)
Doha, Qatar
iweber@qf.org.qa

Abstract

Article 1 of the United Nations Charter claims "human rights" and "fundamental freedoms" "without distinction as to [...] sex". Yet in 1995 the Human Development Report came to the sobering conclusion that "in no society do women enjoy the same opportunities as men"[1]. Today, gender disparities remain a global issue and addressing them is a top priority for organizations such as the United Nations Population Fund. To track progress in this matter and to observe the affect of new policies, the World Economic Forum annually publishes its Global Gender Gap Report. This report is based on a number of offline variables such as the ratio of female-to-male earned income or the ratio of women in executive office over the last 50 years.

In this paper, we use large amounts of data from two online social networks, Twitter and Google+, to study gender differences in 45 countries world-wide and to link online indicators of inequality to established offline indicators. Whereas certain differences, such as women tweeting more often, are found globally, differences concerning the frequency of hashtags are varied and often insignificant. Concerning the link to offline variables, we find that online inequality is strongly correlated to offline inequality, but that the directionality can be counter-intuitive. In particular, we consistently observe women to have a higher online status, as defined by a variety of measures, compared to men in countries such as Pakistan or Egypt, which have one of the highest measured gender inequalities.

We believe our findings contribute to ongoing research on using data for development and prove the feasibility of developing an automated system to keep track of changing gender inequality around the globe.

[1] `http://hdr.undp.org/en/content/human-development-report-1995`

1 Introduction

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. Reducing and ultimately erasing the "Gender Gap" in these fields is both an intrinsic, moral obligation but also as crucial ingredient for economic development. 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. 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" 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 paper contributes to this existing line of work by quantifying gender differences around the globe using *online* data, concretely data derived from Twitter and 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, their activity levels 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 meta data of user profiles and tweets. 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 language 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.

Analyzing gender differences with a focus on 45 countries we find both expected and surprising trends. Our main findings are:

- Women have a higher fraction of retweeted content as opposed to original content.
- Women are more active, at least on Twitter, and generate more content.
- Women are more tightly cliqued and more reciprocated, at least on Google+.
- Women have slightly fewer tweets with URLs than men.
- There is little to no difference between women and men regarding their tweet length.
- The gap of the number of users both in Twitter and Google+ correlates with the gender gap index: country with more men than women online are countries with more pronounced gender inequality.
- The gap of the number of followers is negatively correlation with the gender gap index: in countries where the gap index is low, indicating a higher inequality, women are, in general, more followed than men. This result holds both for Google+ and Twitter, either using the mean or the median to calculate the ratios. It holds for other metrics for "status" such as the number of lists a Twitter user is in and PageRank in Google+.
- The gap of the fraction of tweets that are retweets negatively correlates with the gender gap: if we consider tweets that are not retweets as a quantity of original content, then in countries with a low gap score men have a higher fraction of original content than women.
- The offline gender gap can be "predicted" reasonably well using online data, with correlation values of >0.7 across 45 countries, using merely two indicators from Twitter.

Apart from these findings and the overall framework, we also contribute a method for down-weighting users from an activity-based data sample, 10% of the full Twitter stream in our case, to be closer to a uniform user rather than activity sample and to reduce the impact of highly active users.

Generally our analysis is more quantitative and descriptive rather than qualitative and diagnostic. Though we describe the gender differences we find and we comment on whether they agree with (at least our) expectations, we do not attempt to give explanations as to why, say, women have a lower fraction of original content as opposed to retweeting others' content. 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 of political engagement happens online we are convinced that, ultimately, quantifying gender inequality also has to crucially take into account online activity.

2 Related Work

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 paper'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 than have been investigated.

2.1 Offline

Feingold conducted a meta-analysis to investigate differences in personal traits between genders as reported in literature [16]. 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 [39]. 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. [14] 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 [35]. 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. [44] conceived the General Sex Difference Index and observed that sex differences appear to diminish as one moves from Western to non-Western cultures.

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 [23].

2.2 Online

Gender gap. Bimber analyzed data from surveys in the United States, in which people were asked about Internet access and frequency of utilization [5]. 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 [13]. 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. 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 [10]. In our data for the US, we have slightly more male users, both for Google+ and Twitter. 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.

Sentiment and emotion. Thelwall et al. analyzed public comments in MySpace by manually labeling them as having a positive or negative emotion [48]. The authors observed that there is no difference in behavior between genders concerning negative comments, but women tend to give and receive more positive emotions compared to men. Similarly, Kivran-Swaine et al. analyzed Twitter conversations between pairs of users and looked for the presence of positive emotion expressions, and also observed that female users express more positive emotion terms than male [26]. Ottoni et al. collected information from more than 600 thousand user profiles and 200 million items in Pinterest and studied gender differences in relation to activity, network and content shared in the system [34]. Regarding content creation, they observed that females are more generalists, while males are more specialists. Analyzing the profile description text provided by each user, they observe that, while man are more assertive, women tend to use more words of affection and positive emotion.

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 [47]. 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 [24]. 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 asking about their presence, frequency of utilization and the type of images or information they were likely to disclose [8]. They found support for the hypothesis that female participants disclose more images and information on OSN profiles than male participants. 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.

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 [41].

Network. Szell and Thurner analyzed the interactions between players of a massive multiplayer online game [45]. Among other experiments, they constructed the interactions graphs to calculate and analyze network metrics (average degree, average neighbor degree, clustering coefficient, and reciprocity). They observed that there are difference 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 [34]. 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 the homophily, they found out that on average man is almost twice more likely to follow another man than a woman, and, surprisingly, women are also more likely to follow a man [22]. However, such findings depend on how the sample was constructed, e.g., whether users were sampled from an activity stream (thereby biasing towards more active users), whether it came from a partial crawl of the network (thereby biasing it towards better connected users), or whether it came from trying out random, numeric user IDs (thereby including lots of users without any activity). 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"[2] to Twitter [18]. 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.

Gender identification on Twitter. Twitter does not provide an option for users to explicitly declare their gender. Correspondingly, it is impossible to automatically retrieve all female Twitter users in a given country. Since many studies investigate gender behavior, they all require some methodology to identify the gender of a Twitter user. Some studies proposed a gender identification method that relies on language characteristics present in the tweets (such as word n-grams, characters n-grams, punctuation, smile and laughs) and, using a SVM classifier, got accuracy rates between 66% and 70% [43, 36]. Fink et al. utilized LIWC [46] and hashtag features in addition to unigrams from tweet text, and obtained accuracy of 80% [17]. Ciot et al. exploited language-specific characteristics such as popular n-grams in the languages and gender-specific adjectives in French [12] and reported an accuracy of 90%. Another group of studies take advantage of the name, and either using a dictionary-based approach, extracting textual features in the name (e.g. uncle), or counting n-grams in the name, screen name and description, obtained accuracy rates between 84% and 92% [32, 30, 9]. Popularity and activity metrics, such as number of followers and retweet fraction, have also been used together with text or name features to predict the gender [12, 30]. In a different approach, Alowibdi et al. used 5 features related to colors in the profile and got results in the 70-80% accuracy range [1]. Using aggregated information from the neighbors and creating a set of features of textual and activity characteristics extracted from the friends was also studied, and authors show that it is possible to get accuracy up to 80% for gender identification [50]. In this present study we used a high-precision dictionary approach with estimated precision of 96%. See the appropriate section for details. Note that using content features would have been near impossible due to the number of languages covered. Using network features would have introduced biases as users connected to lots of women would have simply been "assumed" to be female by the classifier.

Socio-economic indicators from online data. Putting aside the concrete issue of gender inequality, 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 [2], consumer confidence [33], migration rates [49, 21], values of stock market and asset values [7, 6, 51] and measures of social deprivation [42]. Work in [40] 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.

3 Data Sets

For our analysis of online social networks we used two data sets: one data set built around Twitter messages (tweets) and one around a crawl of the Google+ social network.

[2] `http://en.wikipedia.org/wiki/Bechdel_test`

3.1 Twitter

Our Twitter dataset is consists of tweets from the Twitter "Decahose", which is a 10% uniform random sample of all public tweets produced by users in Twitter. We used the Decahose data for a two-week period from March 24th to April 6th which contained public tweets for 59,353,862 users.

Mapping Users to Countries. In order to identify the country of the users we examine the self-declared *location* field of the Twitter profile. This is a free-text field where the user can write anything, including non-sensible information. Considering that, we use a local installation of Nominatim[3] to geo-code these noisy strings and to identify the country mentioned (if any). Nominatim has a search feature that looks up a textual string and associates it with a indexed location containing information such as street, city, state and country, depending on how specific the written text is. The string may be ambiguous and match more than one place. For example, "St Petersburg" might refer to St. Petersburg in Florida rather than in Russia. In such cases we consider the top, most relevant result given by Nominatim. Since popular strings (e.g. "New York") are repeated among users, we create a list of unique strings present in our dataset. In order to minimize the number of requests, we remove location strings that occurred only once. Among the 59 million users 29,749,320 had a non-empty location string. Considering these users, we have 9,483,510 unique strings, from which 1,141,861 appeared at least twice. From these 562,559 (49.3%) could be translated into a valid country, resulting in 18,278,271 users with an identified country.

Gender Identification. Twitter does not provide a gender field in the profile of the users. Therefore it is necessary to use an automated technique to infer the gender of a user. To accomplish the gender identification, we use the name shown in the Twitter profile and in combination with a gender-specific first name dictionary. For this purpose, we first calculate the frequency of first names (first non-whitespace string sequence of the full name) for both male and female users in Google+. Then, we create a name-to-gender table, including only names used by at least 10 Google+ users and for which at least 95% of the users with that name were from the same gender. This table has 160,740 names, from which 54,987 are female names and 105,753 are male names. To increase our coverage we also use the `gender.c` tool[4] to include more names not present in our Google+ table and to remove names for which there is a disagreement between the dictionary built from Google+ and gender.c. In order to evaluate our technique we test the combined table against the Google+ users. We were able to identify the gender of 80.1% of the users (100 out of 124 million) with 96.3% precision.

Unbiasing the User Sample. The Decahose represents a 10% sample of tweets, not of users. Highly active users with a large number of tweets are more likely to be included in the Decahose sample and hence in our data set. This creates a selection bias and our user set is not representative of a uniform random sample across users. To reduce this bias, we down-weight users according to the probability that they were included in our sample. Concretely, we calculate for each user a factor of activity defined as

$$w_u = \frac{0.1}{1 - 0.9^{ntweets_u}},$$

where the number of tweets $nweets_u$ is estimated as

$$max\left(\text{\# tweets in Decahose}, \frac{\text{\# tweets in total}}{\text{\# days since registration}} \times 14\right).$$

Though this re-weighting goes a long way in unbiasing the user sample, it is still not perfect. First, the procedure requires that a user tweets at least once during a 14-day interval. Otherwise, the user cannot be detected from the activity stream of the Decahose. In practice, all online companies only report "active users" where the definition of "active" usually requires a single log-in during a period of one month. This means that neither our procedure nor common definitions would count inactive users as actual users and so disregarding these users is in line with common practices. Second, the procedure assumes a uniform activity profile where a user's average tweet account can be taken as representative for our 14-day interval. To get the actual tweet count during the 14-day interval for all of our users is, unfortunately, not feasible as it would require having access to the full users' activity streams. However, we do not believe this shortcoming to have a dramatic effect as, in expectation, our tweet count estimate is correct even though it does not catch temporal variations.

Our Twitter user weights are used throughout our analysis and all averages or medians are *weighted* accordingly.

3.2 Google+

The Google+ dataset was created by collecting public information available in user profiles in the network. We inspected the *robots.txt* file and followed the sitemap to retrieve the URL's 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 pro-

[3] **http://wiki.openstreetmap.org**
[4] **http://www.heise.de/ct/ftp/07/17/182/**

files in Google+ by the time of the second 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+, contrary to Twitter, 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 1.85% selected "other". In this work we only consider users from the female and male categories.

Details of the Google+ platform and a data characterization of an early version of the dataset are discussed in (anonymous).

3.3 Online Variables

In Twitter we have two types of variables for each user: (1) variables extracted from the profile and (2) those extracted from the tweets. The profile metrics are extracted from the profile information contained in the most recent tweet of the user in the dataset.[5] The tweet metrics are calculated by aggregating information from all the tweets of a user if more than one tweet was present in the Decahose sample. The variables are:

- *Profile*
  - Number of tweets (total).
  - Number of followers.
  - Number of friends.
  - Number of tweets the user has favorited.
  - Number of lists the user is a member of.
  - Time since registration.
  - Number of tweets per day (average).
  - Has profile URL. (true/false)
  - Profile URL is linkedin.com. (true/false)
  - Profile URL is facebook.com. (true/false)
  - Uses interface in English. (true/false)
- *Tweets*
  - Average length of tweet.
  - Fraction of tweets with URL.
  - Fraction of tweets with hashtag.
  - Fraction of tweets with mention (disregarding retweets).
  - Average number of URLs in a tweet.
  - Average number of hashtags in a tweet.
  - Average number of mentions in a tweet.
  - Fraction of tweets that are retweet.
  - Fraction of tweets that are reply.
  - Fraction of tweets that have GPS coordinates.
  - Fraction of tweets from an iPhone device.
  - Uses GPS (true/false).
  - Uses iPhone (true/false).

Note that even though the tweets variables are computed over a sample of only 10% of a user's tweets, all variables related to fractions or averages are unbiased estimators in a statistical sense[6]. Though for a user with only a single tweet in the sample the fraction of their tweets with, say, a URL will be 100% or 0%, the expected fraction across all users is unbiased and, for a sufficient data size, will converge to the true sample fraction.

For Google+ we had the possibility to calculate additional *network* measures as we had the complete social graph. The variables are:

- *Google+*
  - In-degree (number of followers).
  - Out-degree (number of followees).
  - Reciprocity: 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.
  - Clustering coefficient: for a particular node, it 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.
  - PageRank: measures the relative importance of

[5]Each individual tweet in the Decahose comes with a rich set of meta information for the respective user, including among many other attributes the number of their followers, the date they registered and even their background color.

[6]**http://en.wikipedia.org/wiki/Bias_of_an_estimator**

a user in the network, considering only the 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.
- Differential assortativity[7]: "lift" of the fraction of users of the same gender followed by a particular user, 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 in an online population of, say, 80% men males are trivially more likely to follow other males even without any same-gender homophily.
- Assortativity: fraction of the links to the same gender. For this variable we calculate the average of the fraction among all users of a country, regardless of gender, instead of calculating a gender ratio. A large value can be indicative of either strong same-gender linkage preference or a strong gender imbalance with respect to the gender distribution of the users.

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, described below. 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. For the Boolean (true/false) variables we calculate the fraction of users with a "true" value. For the Twitter variables we use the weighted average or weighted median[8] using the unbiasing weights (explained in the Datasets section). 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.

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.

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

[8]The weighted median is the 50% weighted percentile, where the percentage is calculated in relation to the sum of the weights instead to the total number of values. See **http://en.wikipedia.org/wiki/Weighted_median**.

3.4 Offline Variables

The Global Gender Gap Index [9] 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 un-weighted average of the four subindexes to create the overall index. The scores range from 0 (total inequality) to 1.0 (total equality). For this study we use the 2013 Global Gender Gap report [20].

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 [10] on internet usage for 2012. The GDP per capita information was collected from the World Bank website [11] and is for 2011. Information for more recent years was missing for some countries which is why we selected data from 2011.

3.5 Selection of Countries

For our analysis we only select countries for which we have a reasonable coverage. Concretely, we consider only countries with a gender coverage higher than 50% on Twitter[12], to make sure we are not biased towards a smaller specific group of a country, and that have at least 1,000 users in both gender groups, both for Twitter and Google+. Table 4 lists all the countries and the respective number of users and gender coverage.

4 Gender Differences Online

Before we link online variables to offline indicators of gender gaps, we first describe how men and women in 45 countries differ in their usage of online social networks.

Figure 1 shows the gender ratio of the variables for each country. We observe that for some variables there's a female predominance (such as "Number of tweets per day",

[9]**http://www.weforum.org/issues/global-gender-gap**

[10]**http://www.internetworldstats.com**

[11]**http://www.worldbank.org**

[12]Recall that whereas for Google+ a user's self-declared gender is explicit, we inferred gender for users on Twitter using a dictionary-based approach described above.

"Fraction of tweets that are retweet", "Reciprocity" and "Clustering Coefficient"), while for others there's a male predominance (such as "Fraction of tweets with URL" and "Differential assortativity"). Interestingly, there are also variables with no difference between the gender groups (such as "Size of a tweet"). In most cases, the gender predominance is the same across countries, but for some variables ("Number of followers" in Twitter and Google+) there are divergences.

5 Online and Offline Gender Gaps

Maybe we could structure this section by indicator type. Such as economy, education, social status, ...

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 distribution of the variables.[13] First, for each country we compute the average (or weighted average, or median, or weighted median) 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 we present the significance test result for some variables. For most countries and most variables the difference between female and male is significant.

Figure 2 shows the linear regression between online variables and the Global Gender Gap scores. Both for Twitter and Google+ we observe that the gap score for the number of users 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, at least to us, we also find that the number of followers and other measures of "status" 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 for both of our online social networks. We discuss potential reasons later in the paper.

Figure 3 shows the linear regression plots of the assor-

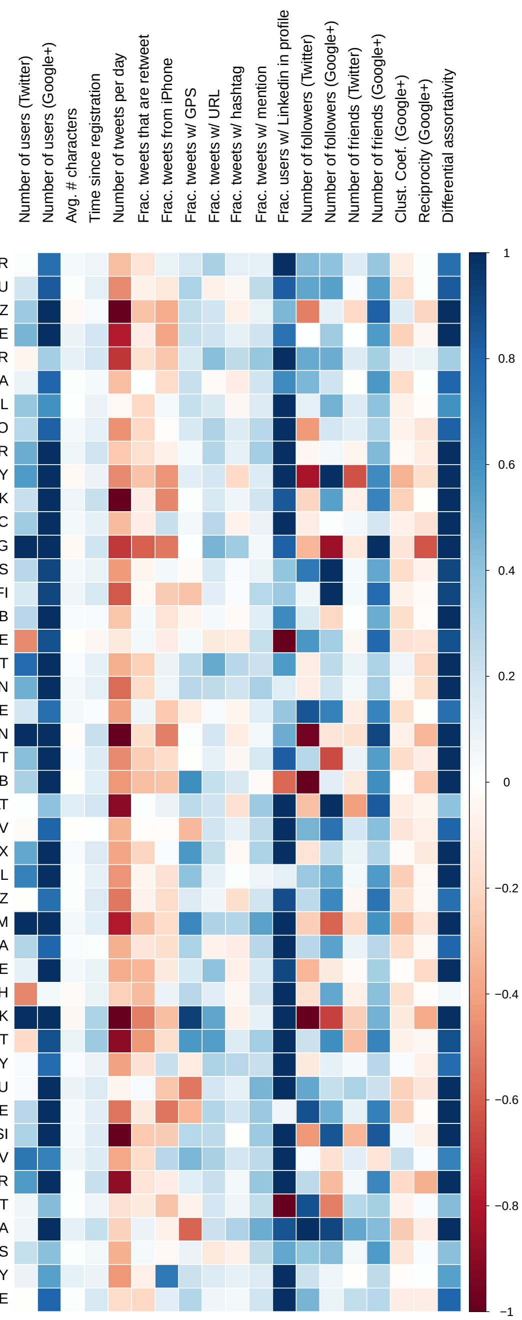


Figure 1: Ratio (log) of female by male value for the variables in each country (ratio is truncated at 0.5 or 2.0 accordingly). A value higher than 0 (blue) means male predominance, and lower than 0 (red) means female predominance.

[13] See **http://en.wikipedia.org/wiki/Resampling_%28statistics%29#Permutation_tests** for background information on permutation tests in statistics.

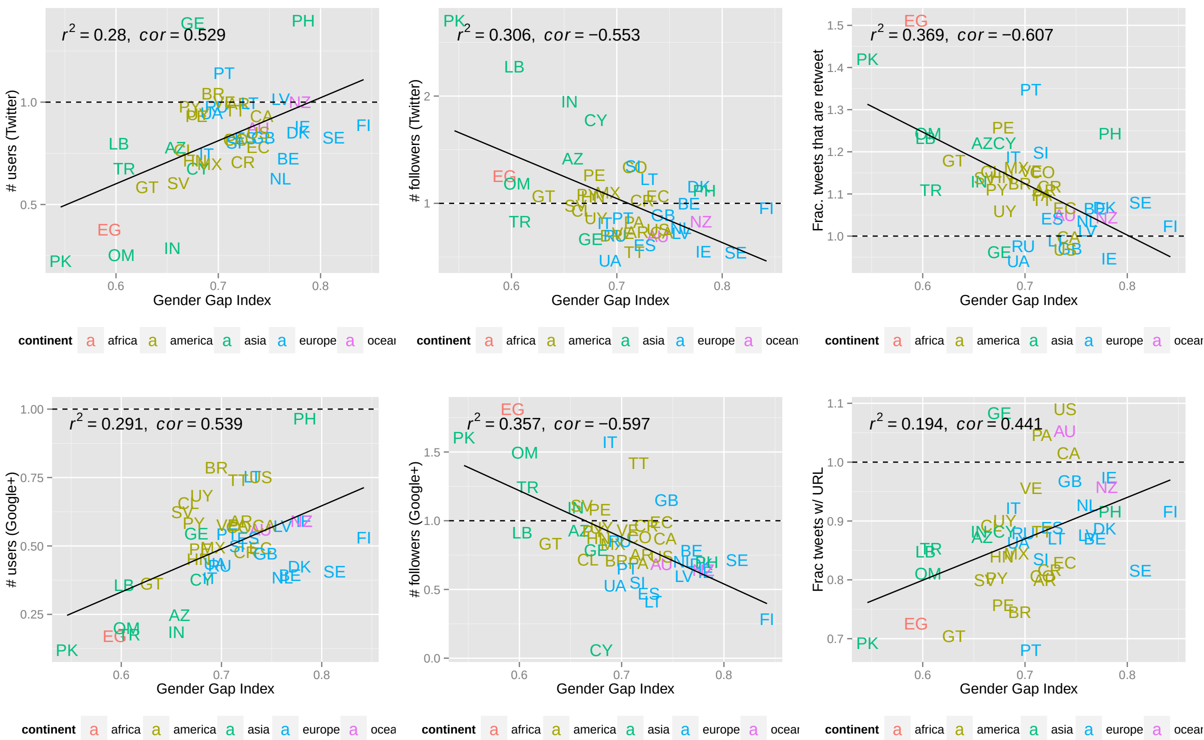


Figure 2: Linear regression and correlation between online social network metrics and the Global Gender Gap score.

tativity variables in Google+. When we analyze the Differential assortativity we observe that most countries have similar values for female and male, meaning that the level of gender assortativity is the same for women and men, whereas 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. When we analyze not the gap but the actual assortativity (average) of a country we observe a positive correlation with the gap score, meaning that in countries with higher Gender Gap score, there's higher assortativity, i.e. more segregation.

Figure 4 presents the matrix of correlation between the online and offline variables. First, we notice that the gender gap of the number of users in both Twitter and Google+ is positively correlated with the Gender Gap score, as expected. Surprisingly, the gap of the number of followers have a negative correlation with the Global Gender Gap score, meaning that countries with a low Gender Gap score female users have relatively more followers compared to other countries. The negative correlation also holds for Reciprocity in Google+. The gap of the Fraction of Tweets that are Retweet negatively correlates with the gender gap, meaning that in countries with a low gap score there's a big gap between women and men in terms of producing original content. 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.

5.1 Regression

In this experiment we analyzed the plausibility of inferring the Global Gender Gap score of a country using information retrieved from users of an Online Social Network. For each one of the 45 countries we create a list of features consisting of the female by male ratio of 25 variables from Twitter and use a linear regression model to estimate the overall Global Gender Gap score. We choose to exclusively use data from Twitter, rather than Google+, as Twitter data is more readily available and hence more appropriate to use for the type of "real-time" index we are interest in building.

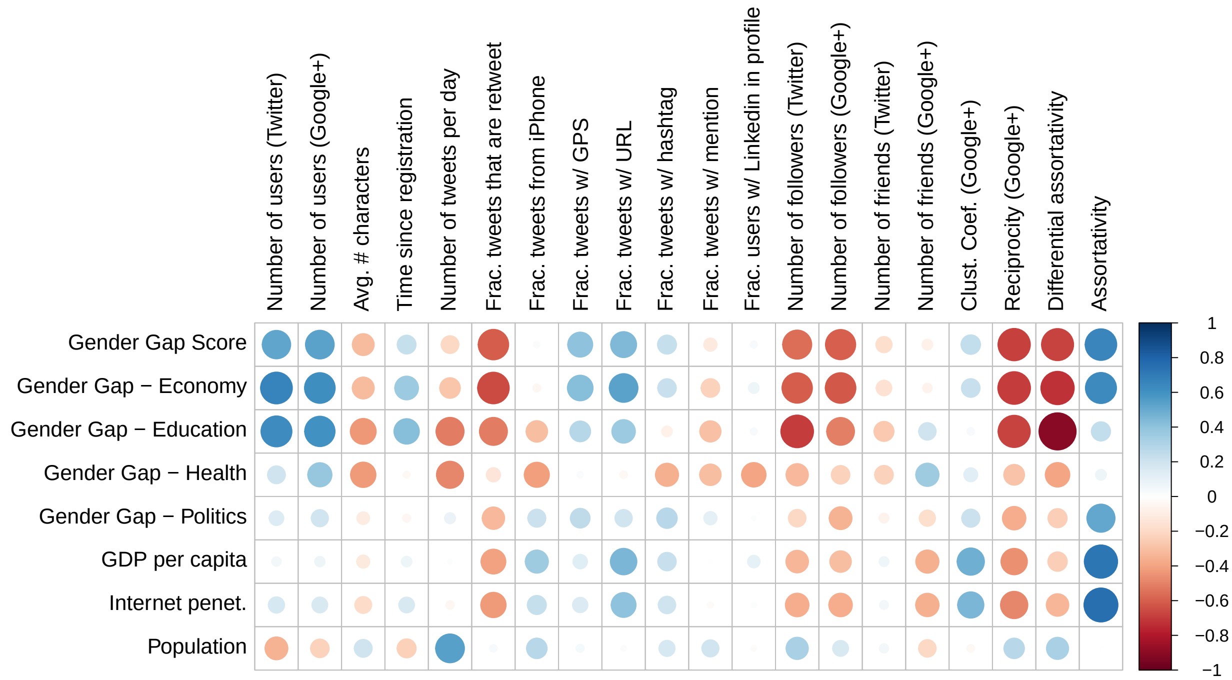


Figure 4: Correlation between offline variables and the ratio of online variables of the countries.

We conduct a greedy forward approach for feature selection. In each step we select the feature to be included next which has the biggest reduction in the MSE (mean squared error), assuming this reduction is actually positive. To train and test the regression model we use a three fold cross-validation with the 45 countries. Since the best feature varies depending on the train/test split, for each step we repeat the experiment 100 times and choose the feature that performed best in most of the runs.

The three best features, in order, are: n_list (number of lists the user is in), f_hash (fraction of tweets that contain hashtag) and f_ret (fraction of tweets that are retweet). Considering the 100 runs to vote and choose the best feature, n_list was the best 98% of the times among 2 features, $frac\_hash$ was 58% among 7 features and $frac\_ret$ was 22% among 20 features.

We then create three linear models considering one, two and three features, now training with all the 45 countries. We test the models interpreting them as meta-features and calculating the correlation with the predicted value and the real gender gap score. Besides testing with the 45 countries we also test with the 88 countries set (which disregards the filter of 50% gender coverage) to gain additional insights into the generalization performance. Table 1 summarizes the results, and Table 2 show the significance of the variables for each model.

Model	Cor. 45	Cor. 88
$-0.097 \times n\_list + 0.800$	0.658	0.519
$-0.098 \times n\_list + 0.163 \times f\_hash + 0.642$	0.705	0.573
$-0.075 \times n\_list + 0.129 \times f\_hash - 0.119 \times f\_ret + 0.787$	0.727	0.612

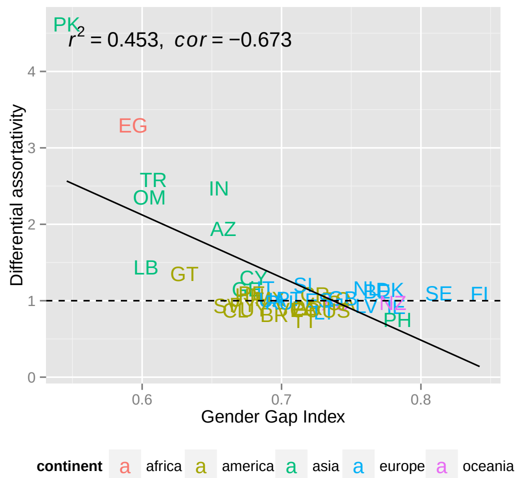
Table 1: Correlation scores between a combination of Twitter online variables and the offline Gender Gap Score for (i) 45 countries (also used for training) and (ii) 88 countries (including 43 previously not used).

Model	(Interc.)	n_list	f_hash	f_ret
1	< 0.001	< 0.001	-	-
2	< 0.001	< 0.001	< 0.05	-
3	< 0.001	< 0.001	< 0.1	0.1

Table 2: Significance of the variables for each regression model (p-value).

6 Discussion

Our main motivation for this work was to see if freely available online data, in particular from Twitter, 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. We have largely shied



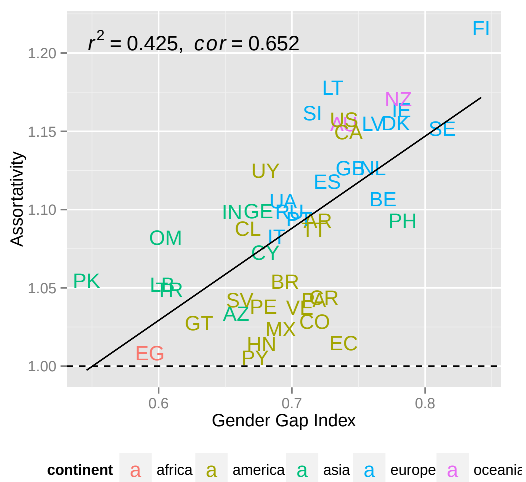


Figure 3: Linear regression and correlation between the assortativity variables in Google+ and the gender gap scores.

away from hypothesizing about the reasons underlying the online differences we observe.

Surprisingly, the directionality of important indicators was *reversed*. Concretely, we found that all indicators of gaps in online social status such as the average/median number of followers on Twitter or Google+, the Pagerank on Google+ or the number of Twitter lists a user is contained in all had noticeable (.50 -.65) *negative* correlations 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 more than 50% more followers on Google+ and 100% more followers on Twitter than men. Note that the number of followers is typically heavy-tailed [27] and for such distributions it is known that the observed average will increase as the sample size increases[14]. 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.

Our current hypothesis is that this unexpected result might be due to an instance of the so-called "Jackie Robinson Effect"[15]. Jackie Robinson was a baseball player who 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" [3]. A related concept is that of the "smurfette principle" [37], which basically says that being a woman is an identity feature, while being a man is the background, the ground assumption.

Of course, our current data set and methodology are by no means perfect. Clearly, our user set is by no means 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 4-5 times as many male Twitter users as female Twitter users 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 [4, 28, 15].

The example of monitoring flu activity also points to another limitation of our study: the use of only two data sources. For flu monitoring using online data, Google Flu Trends [19] is the de-facto standard and baseline to beat. Recently, its use as a figurehead has however been

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

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

questioned [29]. 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. Our current choice of data sets, Twitter and Google+, is mostly dictated by (i) the possibility to obtain large amounts of data and (ii) its global coverage without having to develop custom tools for each country considered.

Another big limitation is our decision to ignore the content/topics that are discussed. As previously mentioned, the main reasons for this are 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 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. In future work we plan 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 potentially shed light or at least provide hypotheses for the root causes of the differences we observe.

A technical limitation in our study is the gender identification on Twitter. Here we decided to use a precision-focused dictionary-based approach and to drop countries from our analysis where the recall was low. Of course the set of names in our dictionary could be a cause for certain biases. For example, our coverage is better for more common names and women or men with unusual names might have a different characteristic. Similarly, our coverage is generally better for Latin-based alphabets and users that use English variants of their names in, say, South Korea might again have different characteristics than "normal" users. Using content and language-specific features could help with the coverage of users whose gender can be identified as in many languages adjectives and verbs in the first person singular ("I") have gender-specific pre- or suffixes. Additional gains could be obtained by obtaining the profile photos and, when these contain a face, applying automatic gender detection[16], or by using textual features from the tweets to discriminate gender [17]. Though promising, such techniques have limitations themselves such as the requirement for language-specific adaptation or scalability problems when tens if not hundreds of millions of photos would need to be downloaded and processed. Similar limitations apply to the inference of a user's country of residence, in particular when there is no explicit location provided in their profile [25, 31, 38, 11].

Our current analysis is based on a single, static snapshot of time. Given the ease with which large amounts of online data can be collected in short amounts of time, our declared goal is to design a system frequently calculates the latest online indicators of gender gaps and makes these publicly available. This is done with initiatives such as the United Nations Global Pulse in mind. "The Global Pulse initiative is exploring how new, digital data sources and real-time analytics technologies can help policymakers understand human well-being and emerging vulnerabilities in real-time."[17] Similarly, the United Nations Population Fund supports use of Data for Development and "women's roles and status, spatial mobility of populations and differentials in morbidity and mortality within population subgroups were singled out as pressing concerns"[18]. At a broader level, more and more non-profit organizations are advocating the use of data mining "for good" and, as an example, the US Center for Disease Control and Prevention is organizing a competition to encourage the use of social media to predict flu activity[19].

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. 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.

7 Conclusion

We presented a large-scale study of gender differences and gender gaps around the world in two large online social networks, Twitter and Google+. Our analysis is based on 7,831,229/13,907,340 users of Twitter/Google+ from 45 countries with an identified gender and, to the best of

[16]See **http://www.faceplusplus.com/** for one such tool.

[17]**http://www.unglobalpulse.org/**

[18]**http://www.unfpa.org/public/datafordevelopment**

[19]**http://www.cdc.gov/flu/news/predict-flu-challenge.htm**

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, including activity levels and online social status. 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 0.7 correlation across 45 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, Pagerank or number of Twitter lists to be included in. 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.

Whereas our current work looked only at well-defined, numerical attributes from meta data, such as the number of followers, the number of tweets or the fraction of tweets that are retweets, integrating content in the analysis would almost surely lead to additional insights. To address the challenges involved in dealing with dozens of different languages and character sets, while ensuring comparability across countries, we plan to focus on a smaller set of countries and languages in future work.

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.

8 Acknowledgments

We thank Ricardo Hausmann at the Harvard Center for International Development and Martina Viarengo at Graduate Institute of International and Development Studies of Geneva for their valuable input.

References

[1] Jalal S. Alowibdi, Ugo A. Buy, and Philip S. Yu. Language independent gender classification on twitter. In *ASONAM*, pages 739–743, 2013.

[2] Dolan Antenucci, Michael Cafarella, Margaret Levenstein, Christopher R, and Matthew D. Shapiro. Using social media to measure labor market flows. Technical Report 20010, National Bureau of Economic Research, March 2014.

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

[4] Eiji Aramaki, Sachiko Maskawa, and Mizuki Morita. 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, 2011. Association for Computational Linguistics.

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

[6] Johan Bollen, Huina Mao, and Alberto Pepe. Modeling public mood and emotion: Twitter sentiment and socio-economic phenomena. In *ICWSM*, 2011.

[7] Johan Bollen, Huina Mao, and Xiao-Jun Zeng. Twitter mood predicts the stock market. *J. Comput. Science*, 2(1):1–8, 2011.

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

[9] John D. Burger, John Henderson, George Kim, and Guido Zarrella. Discriminating gender on twitter. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing*, EMNLP '11, pages 1301–1309, Stroudsburg, PA, USA, 2011. Association for Computational Linguistics.

[10] Women's Media Center. The status of-women in the u.s. media 2014, 2014. `http://www.womensmediacenter.com/page/-/statusreport/WMC-2014-status-women-with-research.pdf`.

[11] Zhiyuan Cheng, James Caverlee, and Kyumin Lee. You are where you tweet: A content-based approach to geo-locating twitter users. In *CIKM*, pages 759–768, 2010.

[12] Morgane Ciot, Morgan Sonderegger, and Derek Ruths. Gender inference of twitter users in non-english contexts. In *EMNLP*, pages 1136–1145, 2013.

[13] Benjamin Collier and Julia Bear. Conflict, criticism, or confidence: an empirical examination of the gender gap in wikipedia contributions. In *CSCW*, pages 383–392, 2012.

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

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

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

[17] Clayton Fink, Jonathon Kopecky, and Maksym Morawski. Inferring gender from the content of tweets: A region specific example. In *ICWSM*, 2012.

[18] David Garcia and Ingmar Weber Venkata Rama Kiran Garimella. Gender asymmetries in reality and fiction: The bechdel test of social media. In *ICWSM*, pages 131–140, 2014.

[19] Jeremy Ginsberg, Matthew Mohebbi, Rajan Patel, Lynnette Brammer, Mark Smolinski, and Larry Brilliant. Detecting influenza epidemics using search engine query data. *Nature*, 457:1012–1014, 2009. doi:10.1038/nature07634.

[20] Ricardo Hausmann, Laura D. Tyson, Saadia Zahidi, and Editors. The global gender gap report 2013, 2013. **http://www3.weforum.org/docs/WEF_GenderGap_Report_2013.pdf**.

[21] Bartosz Hawelka, Izabela Sitko, Euro Beinat, Stanislav Sobolevsky, Pavlos Kazakopoulos, and Carlo Ratti. Geo-located twitter as proxy for global mobility patterns. *Cartography and Geographic Information Science*, 41:260–271, 2014.

[22] Bill Heil and Mikolaj Piskorski. New twitter research: Men follow men and nobody tweets, June 2009. **http://blogs.hbr.org/2009/06/new-twitter-research-men-follo/**.

[23] Janet S. Hyde. The Gender Similarities Hypothesis. *American Psychologist*, 60(6):581–592, September 2005.

[24] Adam N. Joinson. 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, 2008.

[25] Clodoveu Augusto Davis Jr., Gisele Lobo Pappa, Diogo Renn Rocha de Oliveira, and Filipe de Lima Arcanjo. Inferring the location of twitter messages based on user relationships. *T. GIS*, 15(6):735–751, 2011.

[26] Funda Kivran-Swaine, Sam Brody, Nicholas Diakopoulos, and Mor Naaman. Of Joy and Gender: Emotional Expression in Online Social Networks. In *Proceedings of the ACM 2012 Conference on Computer Supported Cooperative Work Companion*, CSCW '12, pages 139–142. ACM, 2012.

[27] Haewoon Kwak, Changhyun Lee, Hosung Park, and Sue B. Moon. What is twitter, a social network or a news media? In *WWW*, pages 591–600, 2010.

[28] Vasileios Lampos and Nello Cristianini. Nowcasting events from the social web with statistical learning. *ACM Trans. Intell. Syst. Technol.*, 3(4):72:1–72:22, September 2012.

[29] David Lazer, Ryan Kennedy, Gary King, and Alessandro Vespignani. The parable of google flu: Traps in big data analysis. *Science*, 343:1203–1205, 2014.

[30] Wendy Liu and Derek Ruths. What's in a name? using first names as features for gender inference in twitter. In *AAAI Spring Symposium: Analyzing Microtext*, 2013.

[31] Jalal Mahmud, Jeffrey Nichols, and Clemens Drews. Where is this tweet from? inferring home locations of twitter users. In John G. Breslin, Nicole B. Ellison, James G. Shanahan, and Zeynep Tufekci, editors, *ICWSM*. The AAAI Press, 2012.

[32] scar Muoz-Garca, Jess Lanchas Sampablo, and David Prieto Ruz. Characterising social media users by gender and place of residence. *Procesamiento del Lenguaje Natural*, 51:57–64, 2013.

[33] Brendan O'Connor, Ramnath Balasubramanyan, Bryan R. Routledge, and Noah A. Smith. From tweets to polls: Linking text sentiment to public opinion time series. In *ICWSM*, 2010.

[34] Raphael Ottoni, Joao Paulo Pesce, Diego Las Casas, Geraldo Franciscani Jr, Wagner Meira Jr, Ponnurangam Kumaraguru, and Virgilio Almeida. Ladies first: Analyzing gender roles and behaviors in pinterest. In *Proceedings of the Seventh International Conference on Weblogs and Social Media*, ICWSM '13, 2013.

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

[36] Claudia Peersman, Walter Daelemans, and Leona Van Vaerenbergh. Predicting age and gender in online social networks. In *Proceedings of the 3rd International Workshop on Search and Mining User-generated Contents*, SMUC '11, pages 37–44, New York, NY, USA, 2011. ACM.

[37] Katha Pollitt. The smurfette principle. *New York Times Magazine*, 7:2224, 1991.

[38] Tatiana Pontes, Gabriel Magno, Marisa Vasconcelos, Aditi Gupta, Jussara Almeida, Ponnurangam Kumaraguru, and Virgilio Almeida. Beware of what you share: Inferring home location in social networks. In *Proceedings of the Workshop on Privacy in Social Data*, PinSoDa, 2012.

[39] Felicia Pratto, Lisa M. Stallworth, and Jim Sidanius. The gender gap: Differences in political attitudes and social dominance orientation. *The british journal of social psychology*, 36(1):49–68, March 1997.

[40] Tobias Preis, Helen Susannah Moat, H. Eugene Stanley, and Steven R. Bishop. Quantifying the advantage of looking forward. *Nature Scientific Reports*, 2:350, 2012.

[41] Daniele Quercia, Diego B. Las Casas, João Paulo Pesce, David Stillwell, Michal Kosinski, Virgilio Almeida, and Jon Crowcroft. Facebook and privacy: The balancing act of personality, gender, and relationship currency. In *ICWSM*, 2012.

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

[43] Delip Rao, David Yarowsky, Abhishek Shreevats, and Manaswi Gupta. Classifying latent user attributes in twitter. In *Proceedings of the 2Nd International Workshop on Search and Mining User-generated Contents*, SMUC '10, pages 37–44, New York, NY, USA, 2010. ACM.

[44] David P. Schmitt, Anu Realo, Martin Voracek, and Jüri Allik. 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, January 2008.

[45] Michael Szell and Stefan Thurner. How women organize social networks different from men. *Scientific Reports*, 3, July 2013.

[46] Yla R. Tausczik and James W. Pennebaker. The psychological meaning of words: Liwc and computerized text analysis methods. *Journal of Language and Social Psychology*, 29:24–54, 2010.

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

[48] Mike Thelwall, David Wilkinson, and Sukhvinder Uppal. Data mining emotion in social network communication: Gender differences in myspace. *JASIST*, 61(1):190–199, 2010.

[49] Emilio Zagheni, Venkata Rama Kiran Garimella, Ingmar Weber, and Bogdan State. Inferring international and internal migration patterns from twitter data. In *WWW (Companion Volume)*, pages 439–444, 2014.

[50] Faiyaz Al Zamal, Wendy Liu, and Derek Ruths. Homophily and latent attribute inference: Inferring latent attributes of twitter users from neighbors. In *ICWSM*, 2012.

[51] Xue Zhang and Hauke Fuehres Peter A. Gloor. Predicting asset value through twitter buzz. *Advances in Intelligent and Soft Computing*, 113:23–34, 2012.

	Twitter					Google+		
Country	Frac. hash.	Frac. retweet	# tweets / day	N. followers (mean)	N. followers (median)	In-degree (mean)	In-degree (median)	Differential assortativity
Argentina	.16/.18**	.30/.27**	3.1/2.5**	353/480**	55/63**	13/18**	3.0/3.0	1.0/1.1**
Australia	.29/.28**	.24/.23**	1.9/1.4**	388/558**	51/48**	15/22**	3.0/3.0	1.1/1.2**
Azerbaijan	.19/.18	.27/.22**	3.4/1.7**	301/212*	49/31**	18/19	3.0/2.0	1.7/0.9**
Belgium	.31/.34**	.27/.25*	1.9/1.1**	275/275	55/51*	14/18*	3.0/3.0	1.2/1.1**
Brazil	.14/.17**	.20/.18**	5.3/3.2**	533/758**	81/56**	20/29**	7.0/7.0	0.9/1.1**
Canada	.31/.29**	.27/.27	1.7/1.4**	347/475**	64/61**	34/39	3.0/3.0	1.1/1.2**
Chile	.22/.21	.33/.28**	1.6/1.6	378/405	52/53	10/14**	3.0/4.0**	1.0/1.1**
Colombia	.14/.16**	.31/.27**	1.8/1.3**	494/368	49/43**	10/11**	3.0/3.0	0.9/1.1**
Costa Rica	.18/.20	.26/.23**	2.1/1.8**	288/280	50/42**	15/15	4.0/4.0	1.1/1.0**
Cyprus	.20/.17	.29/.24**	2.1/1.5**	464/261**	61/57	14/231	3.0/3.0	1.3/1.0**
Denmark	.30/.31	.27/.25*	2.5/1.2**	347/300	39/35*	13/18**	3.0/4.0*	1.3/1.1**
Ecuador	.20/.19*	.32/.30**	1.6/1.3**	334/312	46/43**	9/9	2.0/2.0	1.0/1.0**
Egypt	.19/.24**	.29/.19**	3.6/2.2**	740/589	66/28**	34/19**	3.0/2.0**	2.6/0.8**
El Salvador	.12/.14*	.32/.29**	3.4/2.6**	300/307	61/54*	13/12	5.0/4.0**	1.0/1.1**
Finland	.36/.36	.28/.27	2.2/1.4**	233/244	42/44	13/47**	3.0/4.0**	1.3/1.2**
Georgia	.22/.21	.22/.23	2.2/2.0	210/315	42/40	20/26	3.0/3.0	1.2/1.0**
Guatemala	.10/.12**	.32/.27**	3.6/2.8**	277/259	70/52**	10/12	3.0/3.0	1.3/0.9**
Honduras	.12/.14*	.33/.29**	2.6/1.8**	237/223	68/57**	11/12	3.0/2.0	1.1/1.0*
India	.22/.21**	.26/.23**	2.4/0.8**	627/322**	32/18**	25/23	8.0/7.0**	2.3/0.9**
Ireland	.26/.25*	.29/.31**	2.0/1.5**	320/580	95/83**	14/22**	3.0/4.0**	1.1/1.2**
Italy	.29/.28*	.25/.21**	2.0/1.4**	345/422	42/40**	35/22	3.0/3.0	1.2/1.0**
Latvia	.11/.12	.33/.33	2.1/1.6**	195/269**	87/78**	12/20**	2.0/3.0	1.1/1.2**
Lebanon	.21/.24*	.34/.28**	2.6/1.9**	1740/765	86/59**	18/20	4.0/3.0**	1.4/1.0**
Lithuania	.22/.19	.18/.18	1.9/1.0**	188/153	32/18**	8/19**	2.0/3.0	1.1/1.3**
Mexico	.20/.20*	.33/.28**	2.3/1.8**	555/504	56/48**	10/13**	3.0/3.0	1.0/1.0**
Netherlands	.29/.29	.30/.29**	2.5/1.8**	277/360**	79/80	16/22**	3.0/3.0	1.3/1.1**
New Zealand	.28/.24**	.23/.22	2.2/1.5**	311/374	52/50	14/22**	3.0/4.0	1.1/1.2**
Oman	.14/.18	.36/.29**	3.2/1.9**	358/302	77/44**	25/16*	4.0/4.0	2.2/0.9**
Pakistan	.17/.17	.23/.16**	2.7/1.3**	600/222**	24/15**	25/16**	3.0/2.0	3.7/0.8**
Panama	.21/.20	.33/.30**	2.6/2.0**	334/405	68/65	10/15	2.0/3.0	1.0/1.1**
Paraguay	.11/.14**	.42/.38**	3.0/2.2**	442/409	121/97**	17/18	5.0/4.0**	1.0/1.0
Peru	.23/.22**	.24/.19**	1.3/1.0**	305/241	30/24**	12/11	3.0/2.0**	1.1/1.0**
Philippines	.20/.19*	.17/.14**	3.4/2.9**	528/472	64/60**	12/17**	3.0/3.0	0.9/1.2**
Portugal	.15/.16*	.21/.15**	4.6/2.5**	240/278	55/35**	13/20**	4.0/4.0	1.1/1.1**
Russian Feder.	.14/.15**	.18/.19	2.6/2.5**	296/422**	17/18	18/21**	2.0/2.0	1.1/1.1**
Slovenia	.21/.21	.28/.24**	5.5/2.3**	312/231**	88/50**	10/18**	5.0/5.0	1.3/1.1**
Spain	.20/.20**	.42/.40**	2.8/2.1**	327/536**	98/89**	14/29**	3.0/4.0*	1.1/1.1**
Sweden	.21/.24**	.30/.28**	2.1/1.4**	243/447**	46/43	17/24**	3.0/4.0	1.2/1.1**
Trinidad & Tobago	.20/.21	.21/.19	2.4/2.1	176/322	40/43	20/14	2.0/3.0	0.9/1.2**
Turkey	.12/.13**	.36/.33**	2.8/1.5**	545/656	132/74**	19/15**	3.0/2.0**	2.3/0.9**
Ukraine	.12/.15**	.18/.20*	1.6/1.4**	152/325**	13/13	20/38**	2.0/2.0	1.1/1.1**
United Kingdom	.24/.24**	.26/.27**	2.2/1.8**	565/636	87/82**	31/27	3.0/3.0	1.2/1.1**
United States	.26/.24**	.24/.25**	2.3/1.8**	543/716**	63/65**	35/47**	3.0/4.0**	1.1/1.2**
Uruguay	.15/.16	.27/.25	2.5/1.9**	243/281	46/51*	13/14	3.0/4.0	1.0/1.2**
Venezuela	.19/.20**	.49/.43**	2.0/1.7**	386/534**	52/58**	13/14	3.0/3.0	0.9/1.1**

Table 3: Significance test results for variables in Twitter and Google+. The value on the left is the aggregated female value and the value on the right is the male value, followed by the significance result ('*' is 95% significant, '**' is 99% significant).

Country		Selected	Twitter			Google+	
Code	Name		% w/ gender	Female	Male	Female	Male
US	United States	yes	60.7	1,243,813	1,461,900	2,186,509	2,910,470
GB	United Kingdom	yes	62.3	450,739	544,063	210,801	445,343
ID	Indonesia		39.1	285,577	301,156	136,013	396,028
ES	Spain	yes	57.5	225,256	272,886	116,997	221,343
BR	Brazil	yes	64.0	249,403	239,298	563,173	716,455
TR	Turkey	yes	68.8	130,495	193,248	25,974	147,023
CA	Canada	yes	59.3	154,268	165,309	147,247	255,750
MX	Mexico	yes	61.0	101,909	146,075	129,566	261,958
AR	Argentina	yes	53.0	136,382	137,370	68,877	116,617
CO	Colombia	yes	64.6	92,473	112,960	62,590	110,004
IN	India	yes	59.7	31,316	108,575	363,956	1,964,070
VE	Venezuela	yes	60.9	104,522	104,413	32,623	56,556
FR	France		45.2	85,873	97,994	98,628	211,602
RU	Russian Federation	yes	57.8	91,535	93,515	140,024	326,464
SA	Saudi Arabia		44.8	31,590	79,898	15,173	85,416
NL	Netherlands	yes	56.9	42,872	68,424	40,074	104,336
PH	Philippines	yes	51.0	90,156	64,414	78,760	81,601
IT	Italy	yes	55.8	47,671	63,930	87,028	226,777
JP	Japan		13.9	71,415	61,727	57,234	221,049
AU	Australia	yes	55.5	53,271	60,915	87,605	156,493
DE	Germany		49.0	45,159	58,122	98,500	275,813
CL	Chile	yes	65.1	40,231	52,416	53,286	81,165
EG	Egypt	yes	55.2	14,576	38,475	19,414	113,495
ZA	South Africa		44.1	33,375	38,096	34,153	66,871
MY	Malaysia		35.4	27,778	34,738	60,607	95,842
IE	Ireland	yes	66.1	28,593	32,465	21,277	35,959
EC	Ecuador	yes	63.6	25,308	32,448	15,611	31,654
PE	Peru	yes	67.5	28,436	30,486	32,296	66,141
SE	Sweden	yes	63.7	21,986	26,544	22,342	54,815
UA	Ukraine	yes	53.8	19,470	20,552	46,132	105,582
PK	Pakistan	yes	51.7	4,596	20,548	15,420	128,150
NG	Nigeria		34.7	7,561	18,754	5,050	23,523
BE	Belgium	yes	55.1	13,061	18,011	21,755	55,223
GT	Guatemala	yes	52.2	9,877	16,846	7,342	20,189
KW	Kuwait		40.7	5,322	16,596	3,234	14,674
AE	United Arab Emirates		44.3	9,833	15,884	12,250	57,399
PY	Paraguay	yes	60.6	14,153	14,475	6,273	10,730
SG	Singapore		43.1	10,207	13,642	20,798	43,515
KE	Kenya		48.2	5,900	12,893	6,868	22,522
PA	Panama	yes	61.1	9,700	11,875	4,936	8,565
SV	El Salvador	yes	55.7	6,904	11,424	11,891	19,049
PT	Portugal	yes	51.3	12,864	11,255	32,218	59,238
CN	China		21.1	8,931	11,067	45,551	199,300
FI	Finland	yes	57.1	9,814	11,048	21,831	41,072
PL	Poland		47.1	9,367	10,855	48,381	102,802
NZ	New Zealand	yes	50.8	10,523	10,516	17,462	29,547
DK	Denmark	yes	56.5	8,478	9,942	20,219	47,470
AT	Austria		48.0	7,439	9,491	15,487	37,185
UY	Uruguay	yes	52.9	8,815	9,360	9,966	14,552
CH	Switzerland		45.6	5,465	8,867	14,255	42,085
DO	Dominican Republic		45.2	7,740	8,676	10,750	23,303
CR	Costa Rica	yes	64.0	5,429	7,661	9,632	20,186
GR	Greece		49.5	7,167	7,301	17,578	41,393
IQ	Iraq	no gp	41.4	2,021	7,226	2,101	21,634
HN	Honduras	yes	52.5	4,919	6,882	4,121	9,101
LV	Latvia	yes	58.8	6,868	6,756	5,722	9,979
TH	Thailand		15.5	6,669	6,491	80,655	117,904
KR	Korea (republic of)		25.3	8,173	6,195	16,570	60,696
RS	Serbia		44.7	5,790	5,688	16,458	40,241
JO	Jordan		47.4	2,794	5,351	4,609	20,795
CZ	Czech Republic		47.7	3,243	5,282	19,409	46,548
LB	Lebanon	yes	51.5	4,080	5,111	3,736	10,433
QA	Qatar		46.0	2,308	4,939	2,230	13,176
BY	Belarus	no GP	51.1	6,406	4,746	9,131	21,173
NI	Nicaragua		47.1	3,629	4,604	2,898	6,351
RO	Romania		41.4	3,933	4,449	28,907	63,982
GE	Georgia	yes	61.9	5,772	4,160	3,622	6,642
IL	Israel		42.4	2,520	4,147	15,101	33,752
OM	Oman	yes	50.1	1,041	4,090	1,946	9,708
GH	Ghana		42.0	1,606	3,989	2,821	13,275
IR	Iran		39.9	1,223	3,772	27,153	112,444
PS	Palestinian Territory	no GP	52.8	1,677	3,656	4,278	12,743
AM	Armenia		33.3	2,394	3,615	2,018	5,299
MA	Morocco		39.4	1,949	3,312	7,170	29,434
TW	Taiwan	no GP	15.9	4,064	3,116	43,640	86,235
HU	Hungary		35.2	3,166	3,086	30,525	48,858
NP	Nepal		36.4	1,393	2,926	3,113	22,026
BG	Bulgaria		41.3	2,267	2,839	13,136	25,260
CU	Cuba		41.8	2,351	2,802	2,228	4,408
JM	Jamaica		36.0	2,200	2,797	3,716	5,806
VN	Vietnam		27.9	2,651	2,716	64,539	152,459
BO	Bolivia		41.6	1,644	2,641	4,510	12,093
SI	Slovenia	yes	50.1	2,072	2,581	5,644	11,269
AZ	Azerbaijan	yes	58.9	1,973	2,534	1,840	7,439
KZ	Kazakhstan		40.7	2,543	2,532	5,727	12,555
CY	Cyprus	yes	58.7	1,578	2,334	2,401	6,368
HR	Croatia		40.0	2,273	2,312	13,612	23,944
TN	Tunisia	no GP	53.0	1,757	2,196	5,246	17,805
TT	Trinidad and Tobago	yes	50.3	1,926	2,010	3,408	4,595
SK	Slovakia		46.8	2,162	1,998	16,061	27,749
DZ	Algeria		31.6	1,144	1,962	5,176	24,887
EE	Estonia		39.7	2,001	1,915	5,337	8,337
LT	Lithuania	yes	55.7	1,001	1,004	10,416	13,801

Table 4: List of countries with their respective country codes, whether they were selected for analysis or not, percentage of gender identification in Twitter and the total number female and male users, both in Twitter and Google+. We select only countries with at least 50% of gender coverage in Twitter and at least 1,000 females and males. A country marked as "no GP" is a country that could have been included in terms of the online data available, but that was not included as it is not included in the (offline) Global Gender report.