# A Global Analysis of Emoji Usage

**Nikola Ljubešić**
Dept. of Knowledge Technologies
Jožef Stefan Institute
Jamova cesta 39
SI-1000 Ljubljana, Slovenia
nikola.ljubesic@ijs.si

**Darja Fišer**
Faculty of Arts
University of Ljubljana
Aškerčeva cesta 2
SI-1000 Ljubljana, Slovenia
darja.fiser@ff.uni-lj.si

## Abstract

Emojis are a quickly spreading and rather unknown communication phenomenon which occasionally receives attention in the mainstream press, but lacks the scientific exploration it deserves. This paper is a first attempt at investigating the global distribution of emojis. We perform our analysis of the spatial distribution of emojis on a dataset of ∼17 million (and growing) geo-encoded tweets containing emojis by running a cluster analysis over countries represented as emoji distributions and performing correlation analysis of emoji distributions and World Development Indicators. We show that emoji usage tends to draw quite a realistic picture of the living conditions in various parts of our world.

## 1 Introduction

Emojis, pictograms that have recently gained a worldwide momentum, are considered to be a further development of emoticons, pictorial representations of facial expressions using punctuation marks. While the first days of emoticons go as far as the 19th century (Fitzgerald, 2016), emojis were developed in the late 1990s by Shigetaka Kurita for Japanese mobile phone providers. The difference between emoticons and emojis is that, while emoticons primarily express emotional states, emojis offer a wider spectrum of concepts such as animals, plants, weather, sports, food etc.

Emojis have been present in the Unicode standard for some time now, with the first Unicode characters explicitly intended as emoji added to Unicode 5.2 in 2009. At that point a set of 722 characters was defined as the union of emoji characters used by Japanese mobile phone carriers (Davis and Edberg, 2015). Additional emoji characters followed in later updates, so that the current version 8.0 comprises 1624 emoji characters (Unicode Consortium, 2016). The current popularity of emojis is primarily due to the inclusion of emoji characters on the iOS and Android mobile platforms.

So far, emojis have primarily attracted mainstream media interest, the most prominent being the Word of the Year nomination handed by Oxford University Press in 2015 for the "Face With Tears of Joy" 😂 emoji. For this nomination Oxford University Press partnered with the company SwiftKey which is the author of the currently most detailed analysis of Emoji usage around the world (SwiftKey, 2015).

Despite their popularity, however, emojis are still a poorly researched communication phenomenon as only a few study have focused on it.

Kralj Novak et al. (2015b) inspect the sentiment of emojis by manually annotating 70,000 tweets written in 13 European languages. Their work has resulted in the Emoji Sentiment Ranking lexicon (Kralj Novak et al., 2015a) consisting of 751 emoji characters with their corresponding sentiment distribution. The data the sentiment distributions were calculated on are also available for download (Mozetič et al., 2016).

Pavalanathan and Eisenstein (2016) investigate the relationship between emojis and emoticons, showing that Twitter users who adopt emojis tend to reduce their usage of emoticons in comparison with the matched users who do not adopt emojis.

In this paper we will try to answer the following questions:

1. How popular are emojis in different parts of the world?
2. Does emoji usage differ in various parts of the world?



*Proceedings of the 10th Web as Corpus Workshop (WAC-X) and the EmpiriST Shared Task*, pages 82–89,
Berlin, Germany, August 7-12, 2016. 

3. Does emoji usage in specific parts of the world reflect local living conditions?

We will answer these questions by performing the following analyses over large collections of geo-encoded tweets:

- estimating the probability of emoji occurrence in a tweet given the country,
- clustering countries represented as emoji probability distributions,
- calculating correlation between World Development Indicators and distributions of specific tweets across countries.

The remainder of the paper is structured as follows: Section 2 describes the two datasets used in the analyses while the remaining sections address our three questions: Section 3 gives an analysis of the popularity of emojis in different parts of the world, Section 4 gives an analysis of the spatial distribution of specific tweets, while in Section 5 we present the results of our correlation analysis over specific emojis and the World Development Indicators.

## 2 The datasets

### 2.1 Data collection

Our analyses in this paper are performed on two datasets of tweets collected through the Public Twitter Stream API[1].

The first dataset consists of tweets that have longitude and latitude encoded, regardless of whether they contain emojis. This dataset's sole purpose was to estimate the probability of an emoji occurrence in a specific part of the world. This dataset was collected during a period of 21 days and contains 12,451,835 tweets. We refer to this dataset as the *Twitter dataset*.

The second dataset consists of tweets that have longitude and latitude encoded and that contain emojis. The purpose of this dataset was to estimate the probability distribution of specific emojis in different parts of the world. Since we need more data to estimate the probability of an occurrence of a specific emoji than the probability of the overall emoji occurrence, this dataset was collected throughout a much longer period of 5 months (and is still running) and currently contains 17,458,001 tweets. We refer to this dataset as the *Emoji dataset*.

[1] https://dev.twitter.com/streaming/public

### 2.2 Removing overrepresented users

A frequent problem when analysing data from social networks is the problem of bias towards users with higher productivity, especially since the most productive users tend to be bots with a frequent and specific, if not static, content production.

We apply three methods of removing users with frequent or temporally regular activity. All three methods are run on our Emoji dataset which contains tweets of 2,623,645 users. The identified overrepresented users are then removed both from the Twitter and the Emoji dataset.

The first method removes users who produced on average more than 10 tweets with emojis per day. With that approach we removed 42 users, the user with the highest emoji productivity posting on average 509 tweets per day, the second one posting 72 tweets per day.

Given that most of our later analyses are based on comparing emoji distributions on country level, our second method removes tweets of users that have contributed more than 10% of the tweets that contain emojis in a specific country. Through this procedure we assure that the emoji distribution in a specific country is not heavily influenced by a single user.

We perform this procedure in an iterative manner, removing in each iteration all users that contribute to a specific country more than 10% of all its data points. After each iteration the distributions of user contributions given the country are recalculated. We should note that with this procedure we remove all users from countries that had ten or fewer contributors. With this method we removed 260 users.

The third method focuses primarily on removing bots by calculating the time between two postings and removing all users for which the three most frequent time spans between postings, calculated in seconds, cover more than 90% of their overall production. This method removed overall 16 users from our datasets.

While the precision of all the three presented methods is very high, our assumption is that we still suffer from recall issues. Our plan is to focus on the problem of removing overrepresented / non-human users in more detail in future work.

## 3 Overall emoji popularity

The analyses in this section are primarily focused on how popular emojis are on Twitter. The first part of the analyses looks at the world as a whole, while the second one focuses on the distribution across countries.

Given that for these analyses we need both tweets with and without emojis, we perform all analyses in this section on our Twitter dataset.

### 3.1 Global analysis

Emojis are present in nearly a quarter of the tweets in the dataset (19.6%) and are used by well over a third of the users (37.6%). In this and the following analyses that are focused on users we take under consideration only the users with 100 or more tweets in our dataset as for the remaining users we do not have enough data gathered to produce stable estimates. There are 8,489 such users in our Twitter dataset.

While we have already reported that 62.4% of the users do not use emojis, investigating the probability distribution of using emojis in a tweet among the remaining users shows that half of them use emojis in up to 10% of the tweets while 75% use them in not more than 30% of the tweets. However, the distribution shows a surprisingly thick tail: while 5% of emoji users insert them in every second tweet, 2% of users post less than one emoji-less tweet in ten.

In the following analyses we investigate the differences between the emoji-using and emoji-abstaining users regarding their number of tweets, the number of tweets they have favourited, their number of followers and friends (users that a user follows). We compare the distributions of the four variables among the two types of users with the Wilcoxon test as neither of the variables is normally distributed. The null hypothesis assumes that the median of the two distributions is zero. We always perform a one-tailed test.

By performing our tests on the median we additionally eliminate the impact of outliers which is very beneficial given that our procedures for removing highly active and temporally regular users described in Section 2.2 were focused on emoji-producing users only.

The emoji-producing users have significantly more followers (median 595 vs. 402) and friends (median 438 vs. 288), produce more tweets (median 18280 vs. 12020) and favourite more tweets (median 1760 vs. 1). All the obtained p-values lie in the range $p < 0.001$. One should bear in mind that all the users taken under consideration are highly active on Twitter, producing in the time span of 21 days on average five or more tweets per day.

We have also investigated the dependence of the amount of emojis a user produces and the remaining four variables we have at our disposal, but none of the correlations were strong enough to be worth reporting.

Finally, looking into the number of emojis per tweet we find that single emojis occur in 45% of the emoji-containing tweets, two emojis make for 25% of the tweets, three emojis 15%, four emojis 7%, five emojis 3% and tweets with more than five emojis make 5% of all emoji-containing tweets. This distribution shows that in more than half of the tweets emojis occur with other emojis which makes a co-occurrence analysis as a method for obtaining an insight in the meaning of emojis (or rather the similarity of their meanings) very appealing.

### 3.2 Per-country analysis

In this subsection we investigate the popularity of emojis on a per-country basis. We quantify the emoji popularity in a specific country by calculating the percentage of geo-encoded tweets that contain emojis. By calculating the percentage of the tweets containing emojis, and not the overall amount of the emojis produced on Twitter, we neutralise the differences in popularity of Twitter among different countries.

Emoji density by country is given in Figure 1. The highest density of tweets can easily be observed in Indonesia (46.5% of tweets containing emojis) and the neighbouring third-ranking Philippines (34.6%). In South America the king of emojis, overall ranking second, is Paraguay (37.6%), followed by Argentina, overall ranking sixth (30.7%). In Africa emojis are most popular in the north, with Algeria ranking fourth (33.5%), Egypt ranking seventh (30.4%) and Libya ranking eight (29.7%). In the Arab peninsula Qatar comes first (overall ranking fifth, 32.6%), followed by UAE (ranking 10th, 27.1%). The two highest ranking European countries are Latvia (24.4%) and Spain (24.1%), followed by the Czech Republic, Portugal and the Russian Federation. Interestingly, Japan, the home of emojis, is ranked 163rd

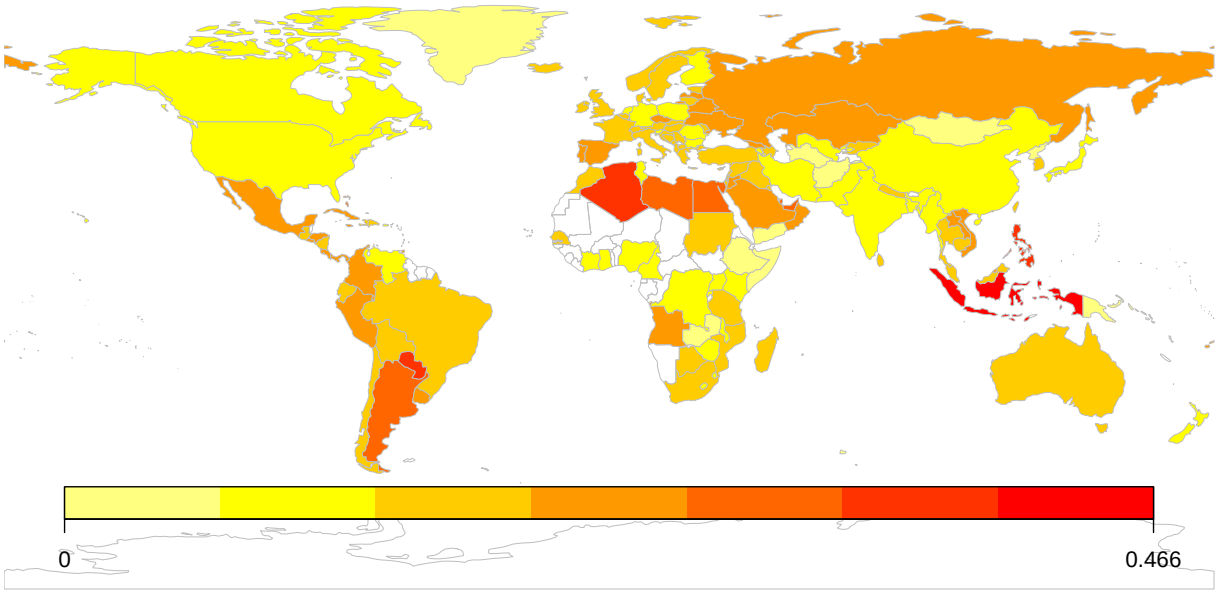


Figure 1: Emoji density per country measured as the percentage of tweets containing emojis

with only 7% of tweets containing emojis. The United States of America, the country responsible for making the pictograms widely popular, is just doing slightly better, ranking 152nd with 10% of tweets containing emojis. The highest ranking North American state is Mexico (21.8%) in 37th position.

Regarding the density of tweets on the continent level, Asia has the highest density with 26.3% tweets containing emojis, South America comes second with 20.9%, followed by Europe (16.7%), Africa (14.9%), Australia (13.7%) and North America (11.5%).

One has to stress right here that although the dataset used for estimating this distribution is rather large, it is still collected from one source only and therefore reflects the sociodemographic specificities of Twitter users of a specific country. Investigating the reliability of these estimates calculated on one social network only is left for future work.

## 4 Popularity of specific emojis

In this section we move from analysing the overall popularity of emojis to analysing the popularity of specific emojis. Again we start with a global analysis, continuing with a per-country one.

This set of analyses is performed on the Emoji dataset as here we are not interested in the probability of emoji occurrence, but the probability of specific emojis among all of them. To estimate these probabilities we do not require tweets that do not contain emojis.

### 4.1 Global analysis

The overall frequency distribution of emojis shows that the most frequent emoji on Twitter since December 2015, with around 2.6 million occurrences in our Emoji dataset, is the "Face with tears of joy" 😂, representing 6.7% of all emoji occurrences. The second most frequent emoji is the "Smiling face with heart-shaped eyes" 😍 (3.72%), on third place we find the "Emoji modifier Fitzpatrick type-1-2" [2] (2.3%), position 4 is taken by "Smiling face with smiling eyes" 😊 (2.1%), and position 5 by "Face throwing a kiss" 😘 (2.1%).

We give a full list of encountered emojis with their frequency and popularity across countries in a separate publication we call *The Emoji Atlas*.[3]

### 4.2 Per-country analysis

In this set of analyses we are interested in how popular specific emojis are in individual countries. We therefore calculate the probability distribution of specific emojis for each country. We discard all the countries having less than 5000 data

[2]There are 5 emoji modifiers that define the skin tone of the emoji. In our analyses we consider these modifiers to be entities by themselves to achieve better generalisation both among modifiers and emojis.

[3]http://nlp.ffzg.hr/data/emoji-atlas/

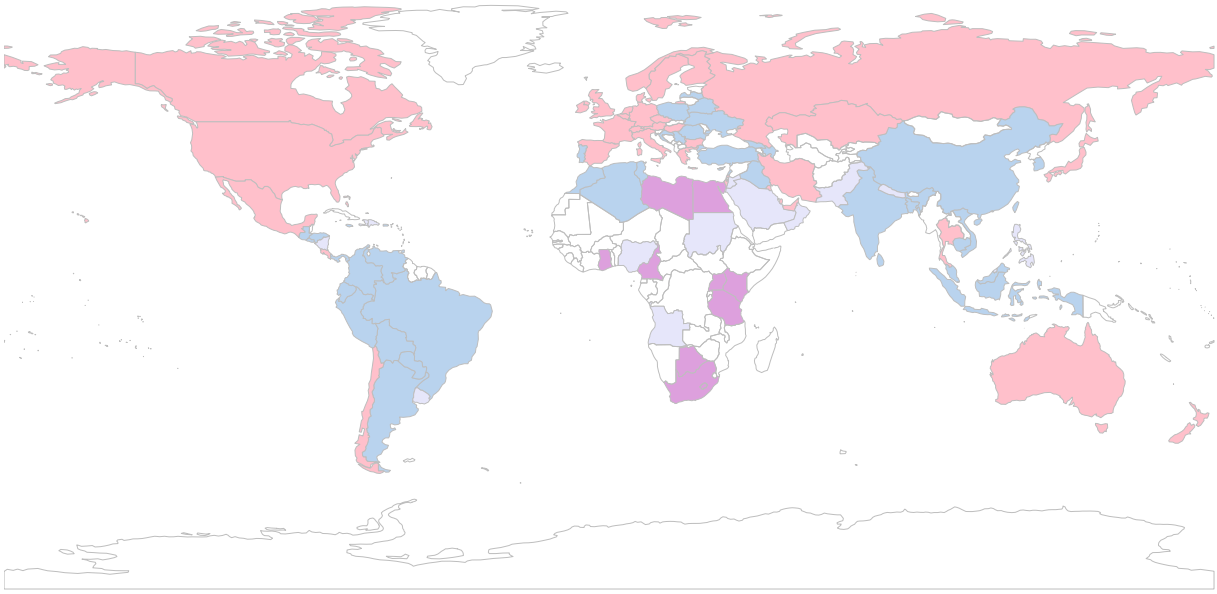

Figure 2: Results of the k-means algorithm on countries represented through the emoji probability distribution

points from our analyses as the estimated distribution of the 1282 emojis found in our data below this threshold would be quite unreliable. While defining this frequency threshold we were not only lead by the number of variables to be determined, but also by the percentage of countries left for our analysis, aiming at a decent global coverage. By applying the defined threshold we were left with 108 out of 233 countries from which we collected tweets in the 5-month period.

To obtain a first insight into the similarities and differences of emoji distributions among countries we ran the K-means clustering algorithm on countries, each country represented by the emoji probability distribution only. We ran the algorithm multiple times on different numbers of clusters and concluded for the 4-cluster division as presented in Figure 2 to be most explanatory. Additionally, this clustering result has proven to be very stable.

We refer to the light red cluster covering North America, Western Europe, the Russian Federation, and Australia as the "first world" cluster.

We call the blue cluster, covering most of South America, India and China, Eastern Europe, Morocco, Algeria and Tunisia the "second world" cluster.

The light blue cluster covering three African states (Angola, Nigeria and Sudan), Jordan, Saudi Arabia, Yemen, Pakistan, Nepal and the Philippines is referred to as the "third world" cluster.

The lilac cluster covering the remaining African states with enough coverage we call the "fourth world" cluster.

While most of the clustering decisions, besides a few that should be inspected more carefully (like Chile belonging to the "first world" cluster), are self-explanatory, we were quite puzzled by the clustering algorithm to pick out Angola, Nigeria and Sudan from the Sub-Saharan Africa and attach them to the cluster of less-fortunate Arab and Asian states. A short online search pointed to their common attribute: they have oil. The question remains whether the shift in the emoji distribution is due to better living conditions of the local population in comparison to most other African states or to the impact of the oil exploiters on the Twitter emoji production.

We analyse the difference between each cluster and the remaining world by calculating one arithmetic mean emoji vector for the cluster in question and another arithmetic mean emoji vector for the remaining clusters. We then subtract the cluster vector from the remaining world vector and inspect the 20 lowest dimensions, i.e. emojis that are most distinctive for the cluster in question. The twenty most distinctive emojis per cluster are given in Table 1.

Interestingly, different to all other clusters, the most distinctive emojis in the "first world" cluster are not face emojis, the first one occurring on

| cluster | most distinctive emojis |
|---|---|
| "first world" | 🏻 🏼 ❤ ✨ 💙 ☀ ❄ 💕 👍 🌴 🎉 🎄 🏡 🌞 🌸 🍀 💖 🌺 😴 ☕ |
| "second world" | 😘 🏼 🎉 😎 💋 😄 😍 👍 💕 💪 💖 🌞 😉 👑 💞 🍻 🌴 🌊 😚 😁 |
| "third world" | 😍 😂 😁 😭 💔 🌹 🙏 😘 😪 😒 👏 💃 💙 😌 😴 🚶 😑 😔 😏 😢 |
| "fourth world" | 😂 🔥 🙌 😭 😅 💃 😒 🏃 ✋ 😢 🏾 🙏 👇 👌 👉 🎶 🏾 🎼 👆 😩 |

Table 1: Twenty most distinctive emojis by cluster

position 19, the "Sleeping face" emoji 😴. The two most distinctive emojis in the "first world" cluster are Emoji modifiers Fitzpatrick type 1-2 🏻 and 3 🏼. The list is continued with a series of weather conditions like the sun ☀ and snow ❄, natural occurrences like "Palm tree" 🌴, different types of flowers ("Cherry blossom" 🌸, "Hibiscus" 🌺), "Sparkles" ✨, and celebration symbols such as "Party popper" 🎉 and "Christmas tree" 🎄. The interesting feature of this cluster is that it lacks direct emotions represented by face emojis that make the most distinctive emojis in all the remaining clusters.

The most distinctive "second world" cluster emojis are a series of happy faces like "Face throwing a kiss" 😘, "Smiling face with sunglasses" 😎, "Smiling face with open mouth and smiling eyes" 😄, "Smiling face with heart-shaped eyes" 😍, "Winking face" 😉 and "Kissing face with closed eyes" 😚. They are interrupted with a "Kiss mark" 💋, "Thumbs up sign" 👍, "Sparkling heart" 💖 etc.

The 20 emojis most typical of the "third world" cluster contain some happy faces such as "Smiling face with heart-shaped eyes" 😍 and "Face with tears of joy" 😂, but also unhappy ones, such as "Loudly crying face" 😭, "Sleepy face" 😪, "Unamused face" 😒 and "Expressionless face" 😑. Here the "Person with folded hands" 🙏 makes its first appearance in our analyses.

Besides two rather happy emojis, "Face with tears of joy" 😂 and "Smiling face with open mouth and cold sweat" 😅, the "fourth world" cluster is specific for a series of unhappy faces such as "Loudly crying face" 😭, "Unamused face" 😒, "Crying face" 😢 and "Weary face" 😩. In addition to these face emojis the "Fire" 🔥, "Person raising both hands in celebration" 🙌 and "Dancer" 💃 emojis are very distinctive as well. A series of emojis depicting people's hands like "Raised hand" ✋, "White down pointing backhand index" 👇, "White right pointing backhand index" 👉 and "White up pointing backhand index" 👆 are also very specific. The list then continues with a series of music-related emojis like "Multiple musical notes" 🎶, "Headphone" 🎧, "Microphone" 🎤 and "Musical note" 🎵 (the last two not making the top 20 emojis), the Emoji modifiers Fitzpatrick type 5 🏾 and 6 🏿 and "Person with folded hands" 🙏.

While skin modifiers make the list of most distinctive emojis in three out of four clusters, their presence does not significantly affect the clustering results as the same cluster structure was obtained when these modifiers were removed from our dataset.

To wrap-up, the "first world" cluster seems emotionally empty (or at least inexplicit) with almost no face emojis and only two general celebratory emojis, the "second world" cluster is highly positive with most distinctive emojis being the happy face emojis. The dominating emojis in the "third world" cluster are a series of unhappy faces and the praying emoji, while the "fourth world" cluster adds a series of hand symbols (for which we do not have an explanation yet) and rather basic symbols like fire, dance, music and celebration.

The already mentioned *Emoji Atlas* contains a full list of countries with their corresponding emoji distributions as well as a full list of emojis with their country distributions. We are confident that this resource will be of great use for less technically-oriented researchers who do not have the opportunity to compile and process such big datasets. Additionally, as we continue to run the data collection procedure, we will regularly update the resource and, with time, add the temporal component as well.

## 5 Emojis and development indicators

In the last set of experiments we look into the correlation between the spatial representation of each emoji and a selection of World Development Indicators of the World Bank.[4] Our spatial repre-

[4] http://data.worldbank.org

sentation of an emoji consists of probabilities of the emoji given a country which makes it comparable to the World Development Indicators since they are calculated by country as well.

For this initial analysis we have selected World Development Indicators for which we were intuitively expecting results with a straight-forward explanation: "Life expectancy at birth, total (years)", "Total tax rate (% of commercial profits)", "Trade in services (% of GDP)" and "GDP per capita (current US$)". Future work should include a wider set of Indicators.

For each indicator we calculate the Pearson correlation coefficient with each of the emojis and rank them by absolute value, inspecting all emojis with a correlation higher than 0.4.

We again remove data from countries with less than 5000 tweets with emojis as we consider the probability distribution of 1282 emojis calculated on such little data to be insufficient for a good estimate.

### 5.1 Life expectancy

The first indicator we take into account is the "Life expectancy at birth, total (years)" indicator.[5]

The emoji with absolutely the highest correlation with this indicator is the frequently mentioned "Face with tears of joy" emoji 😂 (-0.675), surprisingly with a negative sign, meaning that the higher the life expectancy, the lower the usage of the emoji. We have already observed this emoji to be heavily used in our "third world" and "fourth world" clusters.

The second and fourth absolutely highest correlations are the Emoji modifiers Fitzpatrick type 3 (0.596) and type 1-2 (0.578), both occurring more frequently as life expectancy rises. The third position is taken by the "Confused face" emoji 😕 (-0.585), the fifth by the "Person with folded hands" 🙏 (-0.549), both occurring, as expected, more frequently as life expectancy shrinks.

"Dog face" 🐶 and "Hot beverage" ☕ are following emojis with positive correlation, while the strong ones with negative correlation are "Dancer" 💃, "Fire" 🔥, "Baby symbol" 🚼 and "Person raising both hands in celebration" 🙌, all of which have a correlation coefficient higher than 0.5 which is considered to be a strong correlation.

[5] http://data.worldbank.org/indicator/SP.DYN.LE00.IN

### 5.2 Tax rate

The second indicator we consider is the "Total tax rate (% of commercial profits)" indicator.[6]

The only two emojis with a correlation above 0.4 are "Thumbs down sign" 👎 (0.467) and "Pouting face" 😡 (0.461).

### 5.3 Trade

Our third indicator is the "Trade in services (% of GDP)" indicator.[7]

The three emojis with the highest correlation to this indicator are "Slot machine" 🎰 (0.626), "Game die" 🎲 (0.584) and "Speedboat" 🚤 (0.579). Interestingly, there are no emojis with a high and negative correlation with this indicator.

### 5.4 GDP per capita

Our last indicator is the "GDP per capita (current US$)" indicator.[8]

The three emojis with the strongest correlation are "Emoji modifier Fitzpatrick type-3" 🏽 (0.593), "Fork and knife with plate" 🍽 (0.592) and "Bottle with popping cork" 🍾 (0.565). Further positively strongly correlating emojis are "Airplane" 🛬 and "Cooking" 🍳.

The emojis with the strongest negative correlation are "Unamused face" 😒 (-0.428) and "Crying face" 😢 (-0.419).

## 6 Conclusion

In this paper we presented a worldwide spatial study of emoji usage by analysing a large dataset of geo-encoded tweets containing emojis. We depicted the popularity of emojis on Twitter around the world showing that they are most popular in South-Eastern Asia and South America, while in the USA (that technically enabled the rise of emojis) and Japan (the origin of emojis) the usage frequency on Twitter is multiple times lower.

Inspecting the specificities of the countries regarding the usage of different emojis, our country clustering results differentiate between the "first world" cluster the most distinctive features of which are rather emotionally empty, the "second world" cluster which is specific for highly positive emotions, the "third world" cluster which

[6] http://data.worldbank.org/indicator/IC.TAX.TOTL.CP.ZS

[7] http://data.worldbank.org/indicator/BG.GSR.NFSV.GD.ZS

[8] http://data.worldbank.org/indicator/NY.GDP.PCAP.CD

contains both positive and negative emotions, and the "fourth world" cluster which is predominantly negative with additional, rather basic concepts like fire, dance, music and hand gestures.

Finally, by performing a correlation analysis between emoji distributions across countries and a series of the World Development Indicators we have shown that emojis with the strongest correlation clearly describe the indicator in question which allows us to conclude that emoji usage is indicative of the living conditions in different parts of the world.

However, all our results are to be perceived by having in mind that only one social network was used for building our datasets which opens the natural question of data representativeness as (1) not all people use a specific social network and (2) different sociodemographic groups use the same social network in different countries. Nevertheless, this study objectively depicts the state in our social network of choice.

Our future work goes in three directions. The first one is investigating the impact of using only one social network on the final results.

The second direction goes towards the understanding of the meaning of emojis and using them for tasks like sentiment identification, emotion detection etc. For unsupervised modelling of the emoji meaning we primarily consider distributional models and emoji co-occurrence. We also wish to investigate semantic shifts of emojis across space. By continuous data collection, the temporal dimension becomes a relevant focus of interest with a series of similar research questions.

The third direction is aimed at understanding how emojis are included in natural language syntax.

## Acknowledgments

The research leading to these results has received funding from the Swiss National Science Foundation grant IZ74Z0 160501 (ReLDI), and the Slovenian Research Agency within the national basic research project "Resources, Tools and Methods for the Research of Nonstandard Internet Slovene" (J6-6842, 2014-2017).

## References

Mark Davis and Peter Edberg. 2015. Unicode emoji. Technical report, Unicode Consortium. http://unicode.org/reports/tr51/.

Britney Fitzgerald. 2016. Did Abraham Lincoln pioneer emoticons? 1862 speech may offer clues. http://www.huffingtonpost.com/2012/09/19/abraham-lincoln-emoticons_n_1893411.html.

Petra Kralj Novak, Jasmina Smailović, Borut Sluban, and Igor Mozetič. 2015a. Emoji Sentiment Ranking 1.0. Slovenian language resource repository CLARIN.SI.

Petra Kralj Novak, Jasmina Smailović, Borut Sluban, and Igor Mozetič. 2015b. Sentiment of Emojis. *PLoS ONE*, 10(12).

Igor Mozetič, Miha Grčar, and Jasmina Smailović. 2016. Twitter sentiment for 15 European languages. Slovenian language resource repository CLARIN.SI.

Umashanthi Pavalanathan and Jacob Eisenstein. 2016. Emoticons vs. Emojis on Twitter: A Causal Inference Approach. In *AAAI Spring Symposium on Observational Studies through Social Media and Other Human-Generated Content*.

SwiftKey. 2015. SwiftKey Emoji Report, April 2015. Technical report. https://goo.gl/9QXoEn.

The Unicode Consortium. 2016. Full emoji data. http://unicode.org/emoji/charts/full-emoji-list.html.