

Assignment 3 - Social Web

Arthur-Ervin Avramiea
2517642
a.e.avramiea@student.vu.nl

**Mihnea
Dobrescu-Balaur**
2549278
mihnea@linux.com

Zilvinas Kucinskas
zks300
zil.kucinskas@gmail.com

1. DISCUSS AN EXISTING ANALYSIS

1.1 About the data

Report: <http://www.emarketer.com/Article/Where-World-Hottest-Social-Networking-Countries/1008870>

Where in the World Are the Hottest Social Networking Countries?

The report by *emarketer* quantifies the use of social networking websites throughout the world, measuring their spread and predicting future numbers.

There is data about Facebook users and Social Networking websites in general, on all devices, including mobile.

According to the report, the numbers are based on survey and traffic data from research firms and regulatory agencies. The predictions are based on trends observed by *emarketer* and other country-specific socio-economic factors.

The demographics for this report are broad - people who access the internet from any kind of device. The data was clustered by country and region for one of the reports. Its geographic spread includes all the world, restricted to places that have Internet connectivity.

The report was released in February 2012 and covers the 2011-2014 period, meaning half of the data is predicted.

1.2 Types of analyses

The report contains three key points:

- the number of social users worldwide

- the number of social users by region and country
- the number of Facebook users worldwide

All of the data covers the 2011-2014 timeframe.

Considering that the report came out at the beginning of 2012, it means that the numbers for 2012, 2013 and 2014 are based on predictions formed by trend analysis.

In addition, the second point (number of social users by region and country) uses clustering to aggregate the numbers by country and region.

The three analyses don't look only at the absolute value of data, but also at the rate at which it is changing year by year. So, for example, the report predicts that Facebook's yearly increase of users will drop from 44% in 2011 to 13.9% in 2014.

1.3 Limitations of the analyses

The report is released by a research firm, and they favor their corporate subscribers. Because of this, the data that the report is based on is not released to the public. Moreover, there are no specifics mentioned about the size of the data, concrete sources etc. All that is mentioned is that the data is based on traffic data from "research firms" and other surveys. Because of this, the report can not be compared directly with others from the same sources, since those sources are unknown.

It is worth mentioning that the predictions it made were reasonably accurate - for example, the report predicted that Facebook will have 1.14 billion users in 2014, and the latest report says Facebook has 1.19 billion users.

While the report includes social users from Asia and makes claims about users in China, there are no mentions about other social networks. This is important, because Facebook was banned in China at the time of the report. It would have been useful for the report to include data about other social networks as well, especially since there are countries around the world (Brazil, Russia) where Facebook is not the most popular social network.

2. ANALYSIS OF THE GEOGRAPHY OF MEDIA INFLUENCE

Media, especially in the form of news sources, has a complex role in defining the way in which we perceive the side of the world that is near us, as well as the "other side" of the world. First of all, it can determine the priorities of the society, and influence our affinities. For example, as [2] notes, the news media has a major influence on the topics around which a political campaign revolves. Second of all, news media can modulate our interest in specific national or international events [3]. Moreover, consumers judge the quality of the news by the extent to which the news meet their preconceptions and expectations[1].

In effect, the more homogenous the news sources to which a person is exposed, the less educated he/she may be about the topic. To resolve this issue, one can study various news sources when choosing which candidate to vote for in order to have a better impression of the advantages and disadvantages of each candidate choice. When one takes into account various international perspectives on external events, one may have a more complete and balanced image of the parties involved and the issues at stake. And even it involves an additional effort, reading news from sources that do not fit the individuals' preconceptions, or even contradict them, helps the individual understand the basis on which perspectives other than his own are grounded.

With the increased usage of the internet, users have access to sources other than the local or national news at the tip of their fingers. In this context, our research question is whether internet users take advantage of this opportunity, and access news originating from different countries, when informing themselves on an international topic. Although we are aware that different news agencies from the same country may reflect different perspectives on the same topic, we use the simplifying assumption that in sum, the news sources in a country reflect the country's stance on the topic.

To study how internet users access news sources we have chosen a subject that has generated a lot of controversy in the past days, the tensions in Crimea following the 18-23 Ukrainian revolution ¹. We have chosen this subject because it is likely to be treated differently by the media in different countries, depending on each country's stance on, and interests in, Crimea and the Ukrainian revolution.

2.1 Data acquisition

We have monitored the tweets which mention the keyword "crimea", using the Twitter Streaming API, over a time frame of 16 hours, between 12 and 13 March 2014, for a total of 50.000 tweets on the subject. We

¹http://en.wikipedia.org/wiki/2014_Ukrainian_revolution

needed to be able to associate each tweet to the country in which the user is located. Only a few of the tweets are tagged with an exact location. Where that field was not available, we used the user's location, when that was specified. An issue is that the user location field is not very exact - the user can specify any string of characters to describe his location: "Portland, OR, United States" (which points to an actual location) or "Around you" (as ambiguous as it can get). We used the Google Places text search API to extract the location from this text (if it was possible). Although this approach is not the most accurate, we think it suffices for the purposes of this work. We were able to extract the country for 20.786 tweets, which we used afterwards in our analysis.

For each tweet we were interested by the following information: the tweet's text(to be used for sentiment analysis), the urls to which the tweet refers (to analyse the news sources) and the user's location.

From Twitter, for each tweet, we received a lot of unneeded information:

```
{
  \\info about tweet
  "text": "Vice News has some of the most
compelling footage from Ukraine
and Crimea. Just fascinating.
https://t.co/Qp5nNb6q0I",
  "id": ...,
  "favorite_count": 0,
  "entities": {
    \\info about other entities
    "urls": [
      {
        \\info about urls
        "expanded_url": "...",
      }
    ]
  },
  "user": {
    "location": "Olympia",
    \\information about the user
  },
  ...
  "place": null
}
```

For each place, we got from the Google Places API the following information:

```
{
  "Batam, Indonesia": {
    "geo_location": {
      "lat": 1.0456264,
```

```

        "lng": 104.0304535
    },
    "address": "Batam, Riau Islands,
                Indonesia"
}
}

```

From this, we have extracted the country from the last part of the address.

In order to assess the country for a news source, we have used the country top level domain list². The rest of the websites, with non-national domain names (.com, .org, .info etc) were classified by hand (only the most popular ones).

Integrating this data for each tweet results in the following data representation, that fits the purposes of our analysis, and in which we have the location of the tweeter and the location of each url referred in the tweet:

```

{
  "text": "Russia's 25,000-troop...",
  "urls": [
    {
      "url": "http://rt.com/news/r...",
      "country_code": "RU"
    }
  ],
  "country_code": "SK"
}

```

2.2 Analysis of the data

- tweets by country

- tweets by country (first 5)

It is worthy to note that except for USA (which has a large population), and Ukraine (directly involved in the issue), among the top tweeting countries we have Canada, United Kingdom, and Spain.

- percentage of local media focus

- main parties involved - which are the media sources

Limitations - the use of different country sources may reflect immigrants, especially when speaking less "popular" languages, probably better to delineate between news agencies, international news agencies etc

3. REFERENCES

- [1] M. Gentzkow and J. Shapiro. Media bias and reputation. Technical report, National Bureau of Economic Research, 2005.

²http://en.wikipedia.org/wiki/List_of_Internet_top-level_domains

- [2] M. E. McCombs and D. L. Shaw. The agenda-setting function of mass media. *Public opinion quarterly*, 36(2):176–187, 1972.
- [3] W. Wanta, G. Golan, and C. Lee. Agenda setting and international news: Media influence on public perceptions of foreign nations. *Journalism & Mass Communication Quarterly*, 81(2):364–377, 2004.