Do all birds tweet the same?: characterizing twitter around the world

Conference Paper · October 2011

DOI: 10.1145/2063576.2063724 · Source: DBLP

4 authors, including:

Barbara Poblete
University of Chile
68 PUBLICATIONS **4,531** CITATIONS
SEE PROFILE

Marcelo Mendoza
Universidad Técnica Federico Santa María
114 PUBLICATIONS **5,331** CITATIONS
SEE PROFILE

Alejandro Jaimes
Yahoo
140 PUBLICATIONS **5,292** CITATIONS
SEE PROFILE

Do All Birds Tweet the Same?
Characterizing Twitter Around the World

Barbara Poblete[1,2] Ruth Garcia[3,4]
Marcelo Mendoza[5,2] Alejandro Jaimes[3]

{bpoblete,ruthgavi,mendozam,ajaimes}@yahoo-inc.com
[1]Department of Computer Science, University of Chile, Chile
[2]Yahoo! Research Latin-America, Chile
[3]Yahoo! Research Barcelona, Spain
[4]Universitat Pompeu Fabra, Spain
[5]Universidad Técnica Federico Santa María, Chile

ABSTRACT

Social media services have spread throughout the world in just a few years. They have become not only a new source of information, but also new mechanisms for societies world-wide to organize themselves and communicate. Therefore, social media has a very strong impact in many aspects – at personal level, in business, and in politics, among many others. In spite of its fast adoption, little is known about social media usage in different countries, and whether patterns of behavior remain the same or not. To provide deep understanding of differences between countries can be useful in many ways, e.g.: to improve the design of social media systems (which features work best for which country?), and influence marketing and political campaigns. Moreover, this type of analysis can provide relevant insight into how societies might differ. In this paper we present a summary of a large-scale analysis of Twitter for an extended period of time. We analyze in detail various aspects of social media for the ten countries we identified as most active. We collected one year's worth of data and report differences and similarities in terms of activity, sentiment, use of languages, and network structure. To the best of our knowledge, this is the first on-line social network study of such characteristics.

Categories and Subject Descriptors

H.3.1 [**Information Storage and Retrieval**]: Content Analysis and Indexing

General Terms

Measurement

Keywords

Social Media Analytics, Social Networks, Twitter

CIKM'11, October 24–28, 2011, Glasgow, Scotland, UK.

1. INTRODUCTION

The use of social media has grown tremendously all over the world in recent years, and the impact of such growth has expanded in unexpected ways. Twitter, in particular, has become the most widely used microblogging service, and messages posted on it, in many ways, reflected real life events– from the revolutions in Tunisia and Egypt, to natural disasters such as the Chilean and Japanese earthquakes. Twitter users, however, post and share all kinds of information, ranging from personal opinions on important political issues, to mundane statements that may have little interest to most, except for their closest friends.

The range and scope of the service, and the fact that most user profiles and tweets are public, creates a huge opportunity for researchers. Using Twitter data, they can gain insights not just into how that particular service is used, but also into questions that are relevant in a social system for a particular point in time. This includes how news propagates, how people communicate, and maybe how they influence each other. Given this context, two key questions in the study of social media are how its use differs across cultures and countries, and whether any patterns revealed reflect behavioral differences and similarities between different groups. In spite of a long tradition and a lot of research in cultural anthropology, sociology, and other fields that address cultural differences, very little work has been carried out which takes into account large data sets that specifically examine differences across countries[1].

In this paper we present a summary of some of our findings when analyzing a large data set from Twitter. We perform this analysis in order to examine possible differences and similarities in several aspects of the use of the service. In particular, we focus on examining a year's worth of Twitter data for a large number of "active" users in the ten countries which tweet the most. We report on differences in terms of level of activity (number of tweets per user), languages used per country, the happiness levels of tweets, the content of tweets in terms of re-tweets, mentions, URLs, and the use of hashtags. Additionally, we report differences and similarities in terms of the network structure. Our main contribution is to provide a series of insights on how tweeting behavior varies across countries, and on possible explanations for such differences. To the best of our knowledge, this is the largest study done to date

[1] It is out of the scope of this paper to provide an in-depth review, but here we refer specifically to analyzing differences in social media *across different countries*, as there has been of course a lot of work on social network analysis of large data sets.

on microblogging data, and the first one that specifically examines differences across different countries.

The rest of the paper is organized as follows. In Section 2 we give a brief overview of related work and in Section 3 we describe our data set. In Section 4 we describe the distribution of languages used in each of the ten most active countries in our data set, and the main findings of our analysis on the level of happiness in each country. Section 5 focuses on the content of the tweets and network structure, and we conclude by summarizing our main findings in sections 6 and 7.

2. RELATED WORK

The structure of social networks has been studied extensively because structure is strongly related to the detection of communities and to how information propagates. Mislove et al. [10], for example, studied basic characteristics of the structure of Flickr, Orkut, LiveJournal, and YouTube, and found Power-law, small-world, and scale free properties. The authors argue that the findings are useful in informing the design of social network-based systems. Kwak et al. [9], examined the Twitter network aiming to determine it's basic characteristics. One of their main findings is that Twitter does not properly exhibit a "traditional" social network structure since it lacks reciprocity (only 22% of all connections on Twitter were found to be reciprocal), so it behaves more like news media, facilitating quick propagation of news. Java et al. [8], on the other hand, studied the topological and geographical properties of Twitter's social network and observed that there is high reciprocity and the tendency for users to participate in communities of common interest, and to share personal information. Onnela et al. [11], present a study on a large-scale network of mobile calls and text messages. They found no relationship between topological centrality and physical centrality of nodes in the network, and examined differences among big and small communities.

One of the key questions relating to communities and network structure is influence. De Choudhury et al. [4] examine Twitter data and study how different sampling methods can influence the level of diffusion of information. They found that sampling techniques incorporating context (activity or location) and topology have better diffusion than if only context or topology are considered. They also observed the presence of homophily, showing that users get together with "similar" users, but that the diffusion of tweets also depends on topics. Cha et al. [3] studied the in-degree and out-degree of the Twitter network and observed that influence is in fact not related to the number of followers, but that having active followers who retweet or mention the user is more important.

3. DATASET DESCRIPTION

Twitter is a platform which allows users to choose between keeping their profiles and activity (*tweets*) public or private. Users with private profiles make their information available only to a selected group of friends. For obvious reasons, we limit our research only to information provided by users with public profiles[2].

The focus of our research is mostly on characterizing large online social networks, based on user geographical location, for which Twitter provides limited information. Therefore, we perform an initial filter of users based on activity and profile information. First, we choose users which we determine to be *active*. For this we examine a 10-day continuous time window of user activity, selecting day-1 randomly from the year 2010. Then, we consider *only* as active, users which generated tweets during this time frame. Secondly, we filtered the resulting users to keep only active users which had also entered a *valid location* into their profiles during this same time period. We considered as a valid location any text which could be parsed correctly into latitude and longitude (using the Yahoo! public PlaceMaker API[3]). It should be noted, that we performed basically a static analysis, so we did not consider user mobility during this period. We did not process location information which was automatically generated for the user with a GPS device on their client application, based on the fact that location changes continuously with the user. Since in this work we are interested more in characterizing geographical communities of users, we decided to use the location which reflects more accurately the user's *home country*.

Using this criteria, we obtained a set of $6,263,457$ active users with valid location information, which were divided into 246 different countries. For the rest of our analysis, we selected the Top-10 countries with more activity and gathered all of the tweets generated by these users for the entire duration of 2010. In total our working dataset consisted of $4,736,629$ users (76% of the initial 10-day user sample), and $5,270,609,213$ tweets. Figure 1 shows the distribution of the users in our dataset into the top-10 countries, and the activity that they generated for 2010. Note that the amount of activity registered for each country is not necessarily proportional to the number of users. This is explicitly shown in Figure 2, which displays the tweet/user ratio for each country. This ratio is independent of the number of users in each local network.

4. LANGUAGES AND SENTIMENT

Languages. To analyze the language in which tweets are written, we classify each tweet using proprietary software. As a result, 99.05% of the tweets were classified into 69 languages. The 10 most popular languages are shown in Figure 3. English is the most popular language, and it corresponds to nearly 53% of the tweets. Additionally, Figure 4 shows the three most common languages for each of the top-10 countries, as well as the percentage of tweets which correspond to these languages. It is worth noting that English is one of the three most frequently used languages for these countries, and for the Netherlands, Indonesia, and Mexico more than 10% of tweets are in English, while for Brazil it is 9%. Additionally, a special consideration should be taken for the languages of Italian and Catalan, which appear in Figure 3 and Figure 4. This is strange finding given the fact that Italy is not considered in the top-10 countries of our study and the number of people who speak Catalan world-wide is very small. By sampling the tweets for Catalan and Italian we find that many of them correspond to false positives given by our classifier, since they actually correspond to Portuguese and Spanish. The high resemblance of these languages, in addition to the common use of slang, along with misspellings, makes automatic language identification particularly challenging.

Sentiment Analysis. We also analyzed the sentiment component of tweets, for this we use the measure of *happiness* as coined by Dodds et al. [5], which is also more commonly referred to as *valence*. This value represents the psychological reaction which humans have to a specific word, according to a scale which ranges from "happy" to "unhappy". In particular, we analyze the happiness levels for each of the top-10 countries, considering only tweets classified as English and Spanish. To achieve this, we used the 1999 Affective Norms for English Words (ANEW) list by Bradley and Lang [1] for English tweets, and for Spanish, we used its adaptation by Redondo et al. [12]. The ANEW list contains 1,034 words

[2] All processing was anonymous and aggregated. No personal user information was used

[3] http://developer.yahoo.com/geo/placemaker/

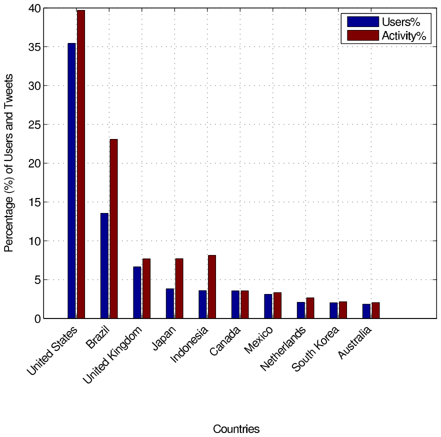


Figure 1: Distribution of users (%) in the dataset for each Top-10 country and their activity (%).

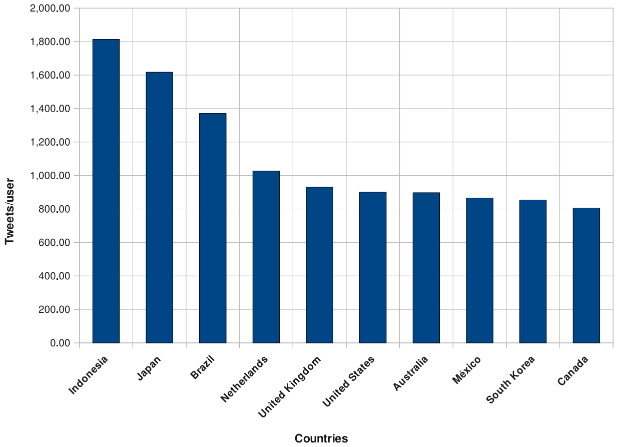


Figure 2: tweet/user ratio for each Top-10 country.

and each word has a score in a 1 to 9 range, which indicates its level of happiness. We computed the "weighted average happiness level", based on the algorithms of Dodds et al. [5], as follows:

$$happiness(C_l) = \frac{\sum_{i=1}^{N_l} w_i f_{i,C_l}}{\sum_{i=1}^{N_l} f_{i,C_l}} = \sum_{i=1}^{N_l} w_i p_{i,C_l} \qquad (1)$$

where $happiness(C_l)$ represents the weighted average happiness level for a country C, based on all of its tweets in language l (English or Spanish), during 2010. Therefore C_l represents all of the tweets registered for the country C which are expressed in the language l. Additionally, N_l represents the number of words in the ANEW list for the language l, while w_i is the score for the i-th word in the ANEW list for l, and f_{i,C_l} corresponds to the frequency of this word in the collection C_l. Finally, we denote p_{i,C_l} as the normalized frequency of each sentiment scored word in C_l.

The results of this sentiment analysis for (a) English and (b) Spanish, are shown in Figure 5. These results agree with those reported by Dodds et al. [5]: the values are between 5 and 7 for both languages and there is also a general increase in happiness towards the end of the year. It is interesting to note that Brazil has the highest values almost every month, even though we are not particularly considering Portuguese. Nevertheless, after August happiness levels in Brazil decreases until November. Also, in December all countries show an increase in their happiness level.

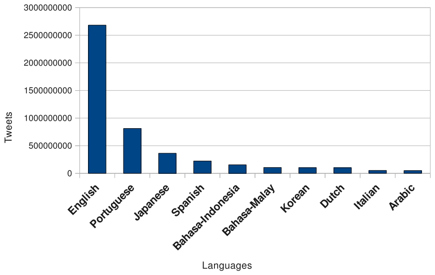


Figure 3: Most commonly used languages in each top-10 country.

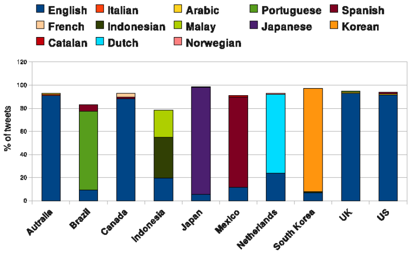


Figure 4: Three most popular languages for tweets in each top-10 country.

Some differences can be appreciated in the results for Spanish tweets, Figure 5.b. The number of tweets in Spanish is disproportional as 7 countries account for less than 1% of the tweets, while Mexico, USA and Brazil together account for almost 98% of the total. Nevertheless, USA and Mexico have happiness patterns that are similar to most countries. Only Brazil and Indonesia results which differ from the rest: there is a strong increase in happiness from June to July for Brazil and Indonesia. Interesting drops in levels happen in Indonesia during the months of May and August. Brazil has clearly the highest values for all months, but it also presents higher ups and downs.

5. CONTENT AND NETWORK STRUCTURE

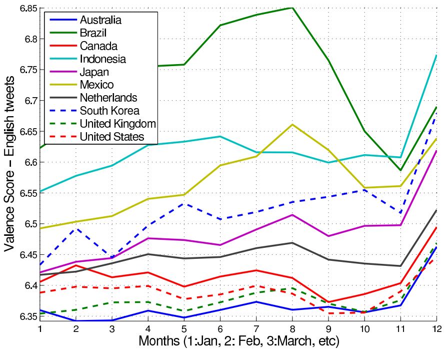
Tweet Contents. In this part of our study, we analyze briefly certain tweet features for each top-10 country. These features have also been used in prior work, such as [2]:

- *#*: indicates if a tweet contains a "#" symbol, which denotes a tweet with a particular topic.
- *RT*: indicates if a tweet has a "RT", which indicates a *re-tweet* or re-post of a message of another user.
- *@*: indicates whether a tweet contains an "@" symbol, used preceding a user name and which indicates a user mention.
- *URL*: Denotes whether a tweet contains a URL or not.

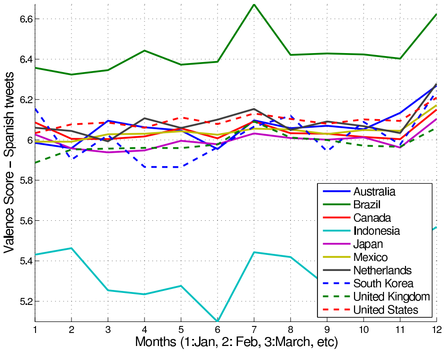
We computed the average per user for each country as follows:

$$AVG(symbol) = \frac{\sum_{i=1}^{N} \frac{T(symbol)_{u_i}}{T_{U_i}}}{\sum_{i=1}^{N} U_i} \qquad (2)$$

Where $AVG(symbol)$ is the average number of tweets per user of a particular country containing a feature denoted by *symbol* (e.g. #, RT, URL, @). Also, N is the total number of users for a particular country and $T(symbol)_{U_i}$ is the total number of tweets containing that feature for user U_i



(a)



(b)

Figure 5: Average *happiness* level (a) English and (b) Spanish.

Country	$\frac{Tweets}{Users}$	$(URL)\%$	$(\#)\%$	$(@)\%$	$(RT)\%$
Indonesia	1813.53	14.95	7.63	58.24	9.71
Japan	1617.35	16.30	6.81	39.14	5.65
Brazil	1370.27	19.23	13.41	45.57	12.80
Netherlands	1026.44	24.40	18.24	42.33	9.12
UK	930.58	27.11	13.03	45.61	11.65
US	900.79	32.64	14.32	40.03	11.78
Australia	897.41	31.37	14.89	43.27	11.73
Mexico	865.7	17.49	12.38	49.79	12.61
S. Korea	853.92	19.67	5.83	58.02	9.02
Canada	806	31.09	14.68	42.50	12.50

Table 1: Average usage of features per user for each country

Table 1 shows the average per country as well as the ratio $\frac{tweets}{user}$. Countries are ordered according to the ratio $\frac{Tweets}{User}$. Results show that Indonesia ranks first in tweets per user, followed by Japan and Brazil. It is interesting also to see that Indonesia and South Korea have the highest percentage of mentions in contrast to Japan that has the lowest, and it seems also to be the country with the fewest re-tweets in our data set. This indicates a higher use of Twitter for conversation than in other countries. The Netherlands is the country with most hashtags per user, while the US seems to be the country with most mentions of URLs per user. At first glance, this could indicate that the US uses Twitter more for formal news dissemination, citing constantly external sources.

Network. Twitter also provides a social network structure for its users. This is, users connect to each other through directed links, therefore relationships are not necessarily reciprocal, as in Facebook[4]. Users can choose to *follow* other users, by subscribing to their updates. These connections between users can be viewed as a large directed graph.

Country	Users	Cov.(%)	Links	Cov.(%)	Recip.(%)
US	1,616,702	12.47	11,310,538	12.46	18.91
Brazil	688,427	5.31	4,248,259	4.68	13.49
UK	286,520	2.21	1,370,699	1.51	17.22
Japan	133,536	1.03	408,486	0.45	**32.01**
Canada	132,240	1.02	553,726	0.61	**26.11**
Indonesia	130,943	1.01	199,704	0.22	**26.97**
Mexico	112,793	0.87	399,409	0.44	17.27
Netherlands	86,863	0.67	354,021	0.39	22.11
South Korea	80,381	0.62	499,261	0.55	**28.14**
Australia	67,416	0.52	299,556	0.33	23.51

Table 2: General summary of network statistics per country.

Country	Avg. δ	Density	Avg. Clus. Coef.	Strongly CC
US	8.95	0.56 E-04	0.0645	9,667
Brazil	7.55	1.09 E-04	0.0711	4,813
Indonesia	2.12	1.62 E-04	0.0618	7,942
United Kingdom	6.05	2.11 E-04	0.0933	14,818
Japan	4.36	3.26 E-04	0.0603	6,052
Mexico	4.44	3.91 E-04	0.0826	6,885
Canada	5.73	4.33 E-04	0.1001	6,630
South Korea	8.61	10.67 E-04	0.0879	3,864
Netherlands	5.39	6.16 E-04	0.1017	4,626
Australia	5.83	8.52 E-04	0.0959	3,423

Table 3: Summary of network density statistics per country.

In this section we focus on the analysis of the Twitter social network graph for each top-10 country and its active users (as defined in Section 3). In order to obtain this graph, we extracted user relationships using the public Twitter API (4J), collecting the list of followers/followees for each user. In this particular graph, connections between users are highly dynamic, so we worked with a snapshot of the graph, which was crawled between November 25 to December 2, 2010. This crawl resulted in $12,964,735$ users and $90,774,786$ edges. We cleaned this dataset to keep only edges and users which corresponded to our *active user set*. Prior work [10] has shown that analysis of partial crawls of social networks can underestimate certain measures, such as degree distribution, but continue to preserve accuracy for other metrics, such as density, reciprocity and connectivity. Therefore, by preserving the active component of the graph we are analyzing the most relevant part of the social structure.

Table 2 shows a summary of each countries' statistics. For each local network analysis, we consider only connections between users in the same country. The second and third columns in Table 2 show the node and edge coverage of each country in relation to the complete graph. We also show the percent of reciprocity, which is the fraction of ties between users which are symmetric. Overall, the top-10 most active countries cover 25.73% of the total of active users in the social graph. Additionally, these countries cover 21.64% of the total number of edges in the global network. Table 2 shows that for some countries reciprocity is very significant in particular for Japan, South Korea, Indonesia and Canada. The symmetric nature of social ties affects network structure, increasing connectivity and reducing the diameter, as we show in the remaining of this work.

Table 3 shows a summary of graph density statistics, such as average degree (δ), density and average clustering coefficient. The US and South Korea are the countries with the highest averaged degree per node, meaning that users tend to concentrate more followers and followees than in other countries. Indonesia, on the other hand, presents a very low degree (only 2.12 edges per node on average)

[4] http://www.facebook.com

Country	Modularity	Number of communities
US	0.418	2,954
Brazil	0.462	2,896
Indonesia	0.537	3,358
United Kingdom	0.397	2,486
Japan	0.458	1,998
Mexico	0.358	1,406
Canada	0.568	1,269
South Korea	0.312	756
Netherlands	0.412	936
Australia	0.452	634

Table 4: Summary of graph modularity statistics per country

Country	Diameter	Avg. Path length	Shortest paths
US	18	6.49	5,746,903,535
Brazil	16	6.37	2,147,483,647
Indonesia	35	9.69	33,940,227
United Kingdom	15	5.19	402,698,573
Japan	18	5.26	86,348,633
Mexico	16	5.27	50,512,898
Canada	16	4.71	77,645,673
South Korea	11	4.02	33,517,802
Netherlands	17	4.81	33,628,133
Australia	14	4.52	22,271,542

Table 5: Summary of graph distance measures per country

in spite of being a very active community. The second column in Table 3 shows each local network's density values. Density is computed as $\frac{m}{n(n-1)}$, where n is the number of nodes and m is the number of edges. The density is 0 for a graph without edges and 1 for a fully connected graph. In our study, South Korea displays the highest density of all countries. Additionally, density increases as the network becomes smaller, i.e. the US has the lowest density, and the highest values correspond to South Korea, Netherlands and Australia. Therefore, smaller communities are in general more well connected, within their own country.

The third column in Table 3 shows the average clustering coefficient. We can observe that communities with high clustering coefficient and less reciprocity may indicate more hierarchical-type relationships between users (i.e., two users who share a reciprocal tie follow a same third user who does not reciprocate). The fourth column of Table 3 shows the number of strongly connected components that exist in each country.

Table 4 shows the the modularity of each social network graph, we use this coefficient as defined by Girvan and Newman [6], which evaluates how well a graph can be partitioned. A value of 0.4 or greater is generally considered meaningful. In our analysis we can appreciate that Indonesia and Canada display high modularity, which indicates that the communities found in these countries are more compact and closed than in other countries. On the other hand, Mexico, South Korea and United Kingdom indicate less separation between their communities. In addition, in Table 5 we summarize some general network distance measures per country. Table 5 shows that Indonesia presents the highest diameter, indicating that this network is very partitioned, which agrees with its high modularity coefficient. Several countries register diameter values in the range of 16-18. The lowest diameter is found South Korea. We can also see that average path lengths are proportional to diameter values. In general, the number of shortest paths is proportional to the number of edges in the graph. Notice that, for example, the three graphs with the highest edge coverage values (US, Brazil and United Kingdom) are also the three countries with the highest shortest path values (see Table 2).

We analyze the existence of a direct relationship between average path length and diameter with reciprocity. Intuitively, we would expect that shorter paths and diameters would result from networks with high reciprocity. Nevertheless, experimentally, we do not observe any apparent relationship. On the contrary, several countries show significant reciprocity and at the same time large diameters. The most noticeable case is Indonesia, which shows the largest diameter and also high reciprocity. This suggests that graph structure strongly influences the relationship between reciprocity and diameter. Given our previous observation, which was that Indonesia had high modularity (see Table 4), this supports the idea that this country has very compact and isolated communities of users. On the other hand, Canada also shows a very significant modularity value but its diameter and average path length values are very similar to countries that do not show a community structure. The main difference between Indonesia and Canada from our observation is that the first has a much lower clustering coefficient and density than the second. This might indicate that Indonesia has more users than Canada which do not participate in large communities.

We also examine the graph structure of each network by considering node degree distribution. Degree distributions of many social networks have been shown power laws behaviors. This kind of networks are networks where the probability that a node has degree k is proportional to k^{γ}, where γ is known as the power law coefficient. Table 6 shows a summary of the graph degrees and assortativity coefficient.

Country	In-degree	Out-degree	Assortativity
	Power Law	Power Law	
US	9.51	13.62	-0.19
Brazil	7.56	12.42	-0.17
Indonesia	6.21	9.48	-0.06
United Kingdom	7.31	10.89	-0.18
Japan	7.48	9.68	-0.07
Mexico	5.91	9.26	-0.21
Canada	8.31	9.30	-0.11
South Korea	6.36	8.21	-0.27
Netherlands	6.67	8.64	-0.17
Australia	7.72	8.12	-0.11

Table 6: Summary of graph degrees and assortativity statistics per country

Finally, we analyze the number of in-links and out-links from one community to another. These results are shown in Figure 6. In this illustration it is interesting to observe that all countries direct the majority of their external out-links to the US. Nevertheless, several countries concentrate their most significant amount of links towards themselves, with the exceptions of Canada, Australia and UK, which connect to the US almost as much as to themselves.

6. SUMMARY AND DISCUSSION

Network reciprocity tells us about the degree of cohesion, trust and social capital in sociology [7]. In this context, the equilibrium tendency in some human societies is to have reciprocal connections. Nevertheless, Twitter networks seem to work towards an equilibrium which is not reciprocal and more hierarchical. Therefore, it tends more to follow a model in which we have authorities which receive many ties but do not reciprocate.

On the other hand, we observe that countries which have high reciprocity tend to have a higher tweets/user ratio. Additionally, smaller networks also have a tendency to have higher reciprocity. Indicating more local communities. Additionally, we see that communities which tend to be less hierarchical and more reciprocal,

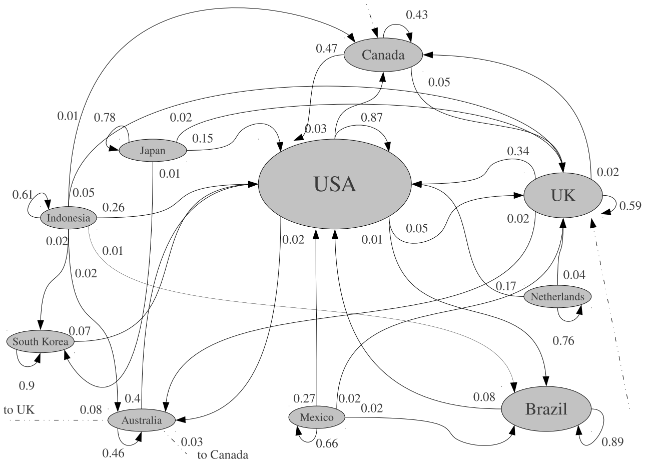


Figure 6: Ties between countries, sizes are proportional the size of the community. Weights represent fraction of ties estimated over the total number of ties of a given country.

also displays *happier* language in their content updates. In this sense countries with high conversation levels (@), shown in Table 1, display higher levels of *happiness* too. This is reasonable, if we think that higher conversational levels can imply that users privilege more informal communication, as opposed to formal news dissemination. Following this reasoning, we can hypothesize that these users, use Twitter more as a conversation channel than formal information source. Therefore, their interaction is highly conversational with friends, as opposed to being more formal. which is the case of the US. High reciprocity does not imply that these countries are more well connected overall, in cases such as Indonesia, reciprocity is contained within very compact and closed groups. This increases the overall diameter of the network, as seen in Table 4.

By analyzing Table 3, we can observe that smaller communities, tend to present more reciprocal ties between users in general, along with high density and average clustering coefficient (not always, but this is the tendency). Pointing towards more reciprocal relationships. Additionally, high reciprocity creates high activity in communities such as Indonesia and Japan. On the other hand, we can see in Table 3 that countries with high density and high clustering coefficient, such as South Korea, Netherlands and Australia, contain users which participate in small and compact communities. Other countries, such as the US, have low clustering coefficient and density, which indicates that many users do not participate in a small compact community.

7. CONCLUSIONS

We have presented a broad study of the Twitter on-line social network. We have segmented our analysis into the top-10 countries with most activity and collected data from a representative sample of users for one year. We analyze several aspects, such as language, sentiment, content and network properties.

In particular some countries standout, based on their peculiar characteristics. Some smaller networks display high reciprocity and more conversation. This indicates a high conversational use within compact communities, as opposed to broad dissemination of news information. In this sense, there also seems to be a correlation between reciprocity and levels of happiness detected in language, examples are Indonesia and Brazil. On the other hand, users in the US give Twitter a more informative purpose, which is reflected in more globalized communities, which are more hierarchical.

In future work we expect to explore more formally correlations between network structure measurements, hierarchy and happiness levels. Also we would like to compare our findings with similar research in other social networks, such as sociological studies.

Acknowledgments. This research was partially supported by European Community's Seventh Framework Programme FP7/2007-2013 under the ARCOMEM project and by the Spanish Centre for the Development of Industrial Technology under the CENIT program, project CEN-20101037 (`www.cenitsocialmedia.es`), "Social Media". Additionally, this work has also been partially supported by FONDEF-CONICYT project D09I1185.

8. REFERENCES

[1] M. M. Bradley and P. J. Lang. Affective norms for english words (ANEW): Stimuli, instruction manual, and affective ratings. In *In Technical Report C-1, The Center for Research in Psychophysiology*, Gainesville, Florida, 1999.

[2] C. Castillo, M. Mendoza, and B. Poblete. Information credibility on twitter. In *Proceedings of the 20th international conference on World wide web*, WWW '11, pages 675–684, New York, NY, USA, 2011. ACM.

[3] M. Cha, H. Haddadi, F. Benevenuto, and K. P. Gummadi. Measuring User Influence in Twitter: The Million Follower Fallacy. In *In Proceedings of the 4th International AAAI Conference on Weblogs and Social Media (ICWSM).*

[4] M. D. Choudhury, Y.-R. Lin, H. Sundaram, K. S. Cnadan, L. Xie, and A. Kelliher. How does the data sampling strategy impact the discovery of information difussion in social media? Many May, 2010.

[5] P. S. Dodds, K. D. Harris, I. M. Kloumann, C. A. Bliss, and C. M. Danforth. Temporal patterns of happiness and information in a global social network: Hedonometrics and Twitter. *Computing Research Repository abs/1101.5120v3[physics.soc-ph]*, Feb. 2011.

[6] M. Girvan and M. E. J. Newman. Community structure in social and biological networks. *Proceedings of the National Academy of Sciences*, 99(12):7821–7826, 2002.

[7] R. Hanneman and M. Riddle. *Introduction to social network methods*. University of California Riverside, CA, 2005.

[8] A. Java, X. Song, T. Finin, and B. Tseng. Why we twitter: understanding microblogging usage and communities. In *Proceedings of the 9th WebKDD and 1st SNA-KDD 2007 workshop on Web mining and social network analysis*, WebKDD/SNA-KDD '07, pages 56–65, New York, NY, USA, 2007. ACM.

[9] H. Kwak, C. Lee, H. Park, and S. Moon. What is twitter, a social network or a news media? In *Proceedings of the 19th international conference on World wide web*, WWW '10, pages 591–600, New York, NY, USA, 2010. ACM.

[10] A. Mislove, M. Marcon, P. K. Gummadi, P. Druschel, and B. Bhattacharjee. Measurement and analysis of online social networks. In *Internet Measurement Comference*, pages 29–42, 2007.

[11] J.-P. Onnela, S. Arbesman, A.-L. Barabási, and N. A. Christakis. Geographic constraints on social network groups. Nov. 2010.

[12] J. Redondo, I. Fraga, I. Padrn, and M. Comesaa. The spanish adaptation of anew (affective norms for english words). In *Volumne 39*, number 3, pages 600–605. Psychonomic Society Publications, 2007.