Influence and Sentiment Homophily on Twitter

Hugo Lopes
INESC-ID Lisbon
Instituto Superior Técnico (IST)
Lisbon, Portugal
hugomalopes@tecnico.pt

Abstract

Web-based social relations mirror several known phenomenons identified by Social Sciences, such as Influence and Homophily. Social circles are inferable from those relations and there are already solutions to find the underlying sentiment of social interactions. We present an empirical study that combines existing Graph Clustering and Sentiment Analysis techniques for reasoning about Sentiment dynamics at cluster level and analyzing the role of Social Influence on Sentiment contagion, based on a large dataset extracted from Twitter during the 2014 FIFA World Cup. Exploiting WebGraph and LAW frameworks to extract clusters, and SentiStrength to analyze sentiment, we propose a strategy for finding moments of Sentiment Homophily in clusters. We found that clusters tend to be neutral for long ranges of time, but denote volatile bursts of sentiment polarity locally over time. In those moments of polarized sentiment homogeneity, there is evidence of an increased, but not strong, chance of one sharing the same overall sentiment that prevails on the cluster to which he belongs.

1 Introduction

Twitter is a highly dynamic social environment where 316 million monthly active users generate a stream of 500 million tweets per day [1]. It not only allows millions of users to interact among each other, but it is also a window for those interactions. Since it is an accessible and prolific source of social data, Twitter and other web-based social networks are widely used in the literature for different Social-related Analysis [2], such as Network Dynamics [3, 4], Community Detection [5, 6], Event Detection and Prediction [7, 8, 9, 10], Information Flow [11, 12], Influence and Homophily Analysis [13, 14, 15, 16], Sentiment Analysis [17, 18]. Some of these study the interdependencies and possible correlations among the different topics, however we found that there is not an extensive study about sentiment prevalence on clusters and whether this sentiment can be propagated by influence into a state of sentiment homophily inside those clusters. Understanding how sentiment behaves at a cluster level can be useful for mining the overall mood of communities, and it may also be useful for improving sentiment classification techniques using enriched information about surrounding emotions. The hypotheses that motivate our work are:

- *H1:* The sentiment expressiveness inside clusters is highly dynamic over time.
- *H2:* Clusters show moments of sentiment prevalence.
- *H3:* During moments of sentiment homogeneity in a cluster, there is an increased chance that a user is influenced by the surrounding emotion and shows a similar sentiment to the one prevailing at that moment.

Regarding some specific terms related with Twitter, a tweet is a message with a maximum size of 140 characters that can include photos and videos. By retweeting a tweet, a user is forwarding that tweet to his own followers. A mention is an explicit reference to a user using the tag "@" followed

by the unique username. For instance, typing "@maria" is a mention to the user "maria". A reply is a particular case of a mention, in which the mention is located at the bottom of the tweet. Replies are used to comment or answer something that the mentioned user has tweeted.

Using existing clustering and sentiment classification techniques, we propose to measure the overall sentiment of clusters based on the frequency of tweets for each possible sentiment value, regarding their sentiment classification. We found that the neutral value is the most frequent classification during the clusters' time-life, however different sentiment values appear, usually in spikes and with different polarities over time, confirming the highly dynamic nature of clusters' sentiment (*H1*). We also observed moments of sentiment homophily (*H2*), for instance in chains of retweets or topic-related discussions and we describe a systematic strategy for finding those moments. Finally, we used dubious sentiment classifications for testing the role of influence in the origin of those moments of sentiment homophily by comparing the extrapolation of the clusters' overall sentiment with human-coders' evaluations. With this strategy we found a tendency for ambiguous classifications being correctly relabeled with the prevalent sentiment of its cluster (*H3*).

2 Related work

Fowler and Christakis [19] conducted a study about the spread of happiness within social networks, using data from the Framingham Heart Study [1], collected between 1983 and 2003. From this data, they extracted a network of $5,124$ individuals and $53,228$ respective social ties. Each person was weekly asked how often they experienced certain feelings during the previous week: "I felt hopeful about the future", "I was happy", "I enjoyed life", "I felt that I was just as good as other people". They used this information to measure the state of happiness of individuals throughout a period of time. According to their results there is happiness homophily in clusters with up to three degrees of separation between nodes. They also had information about people's address, which allowed them to find that geographic proximity among connected people increases the probability of sharing the same state of happiness. This study not only found evidence of sentiment propagation through influence, it also suggests that it may cause sentiment homophily at a cluster level.

Thelwall [20] searched for homophily in social network sites using data extracted from MySpace, concluding that there was a highly significant evidence of homophily for several characteristics such as ethnicity, age, religion, sexual orientation, country, and marital status. Then, he conducted another study on emotion homophily [21], based on the same type of data. Using an initial version of SentiStrength [22] for sentiment classification, two different methods were tested to seek emotion homphily between pairs of friends: a direct method and an indirect method. The direct method compares only the sentiment of the conversational comments between each pair of friends. The indirect method compares the average emotion classification of comments directed to each node, independently, in each pair of friends. Weak but statistically significant levels of homophily were found for both methods. However, the direct method can only give insight of the average homophily at a maximum distance of 1, while the indirect method covers a maximum distance of 3. Both methods do not take into account cluster configurations in the network and the covered range of time considered in the analysis is not specified.

Gruzd et al. [23] followed the study of Fowler and Christakis with web-based social network data, focusing on the potential propagation factors for sentiment contagious instead of searching for evidence of sentiment homophily. They performed a topic-oriented data extraction from Twitter in order to minimize possible bias caused by the occurrence of multiple events that generate multiple unrelated discussions, and they found on the 2010 Winter Olympics a well covered and very popular event on Twitter, from which they got strong emotional content. Using SentiStrength for tweets' sentiment classification, they found that a tweet is more likely to be retweeted through a network of follow relations if its tone and content are both positive. Fan et al. [24] decomposed sentiment into four emotions: angry, joyful, sad and disgusting. They used a bayesian classifier to infer these emotions based on emoticon occurrence in interactions extracted from Weibo. Considering pairs of direct friends in a follow-relation network, they only found evidence of emotion homophily regarding anger and joy, observing that anger was the most influential emotion and the chance of contagion was higher in stronger ties. Using a follow-relation network extracted from Twitter, Bollen et al. [25] also found sentiment homophily but regarding sentiment polarity, which they called subjective well-

[1] Medical study about cardiovascular disease – `https://www.framinghamheartstudy.org/`

	All Tweets	*Simple Tweets*	*Retweets*	*Replies*
Total	97,403,564	37,222,855	53,818,351	6,362,358
Rate	100%	38.2%	55.3%	6.5%

Table 1: Tweet type distribution in the knock-out stage subset.

being assortativity. They observed that pairs of friends connected by strong ties are more assortative, however they did not identify whether this phenomenon was caused by selection or social influence. None of these studies analyzed sentiment dynamics over time nor looked into an overall sentiment at community level.

Following these findings, we propose to look for signs of sentiment homophily at a cluster level and understand whether prevalent sentiment in social circles can be used for estimating individuals' sentiment.

3 Dataset overview

To find social circles and analyze their behavior over time, a large amount of data needs to be extracted during a period of several weeks. We extracted the dataset using Twitter Public Streaming API, through the endpoint `https://stream.twitter.com/1.1/statuses/filter.json` that retrieves the data filtered according to a requested list of keywords. Our data was filtered using a list of words related to 2014 FIFA World Cup, and was stored in zipped files of at most $1,000,000$ messages. Extraction started on March 13th of 2014 and it ended on July 15th of 2014, covering the entire event that took place from June 12th to July 13th of 2014. It resulted in 166 GB of compressed data, distributed in 419 files, containing a collection of $339,702,345$ tweets.

Due to the large amount of countries participating in the World Cup, we only considered a subset of the entire dataset for our analysis. This subset covers the knock-out stage of the event, from June 27th until July 15th, which represents 28.7% of the entire data. We did this to minimize the sparsity of the information, since only 16, from the initial 32 participating countries, were still in competition. English is the most spoken language in the subset, representing 45.8% of the tweets, followed by Spanish with 24.2%, and Portuguese with 10.2%. Table 1 shows the distribution of each type of tweets in this subset.

We found that 64.7% of all tweets have at least one mention, which makes it the most frequent type of strong relation in the dataset, followed by retweets and then replies. However, the set of mentions contains the set of replies and also intersects the set of retweets. Therefore, for conducting an independent analysis for each type of interaction, it was more valuable to consider only retweets and replies, since they are mutually exclusive.

4 Approach

Our approach is divided into four stages: User Clustering; Tweet Clustering; Sentiment Analysis; and Influence and Homophily Analysis in time series, as it is outlined in Figure 1.

The first three stages integrate existing solutions for clustering and sentiment analysis with several scripts for data transformation. They were used to process the extracted dataset into time-series of sentiment information about social circles. With preprocessed data obtained from these three stages, we propose a set metrics to evaluate the extent of sentiment homophily. Then, we propose a strategy to ascertain a possible relation between influence and sentiment, which can eventually improve the sentiment classification of tweets in clusters that denote sentiment homophily.

4.1 User clustering

Before finding the social circles, we needed to find the social network that comprises them. We decided to build the network's graph considering only strong ties, which the literature states to be found in retweets and mentions [26, 14, 4]. However, we chose to use only replies, because retweets

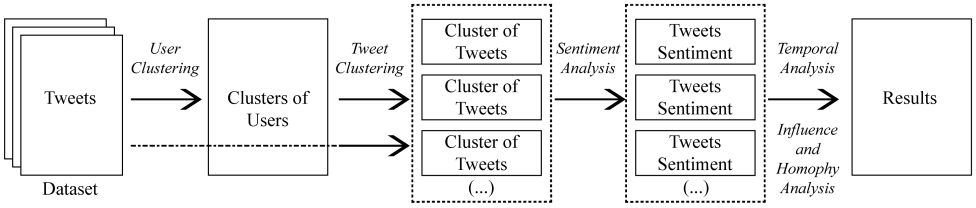

Figure 1: High-level view of the designed workflow.

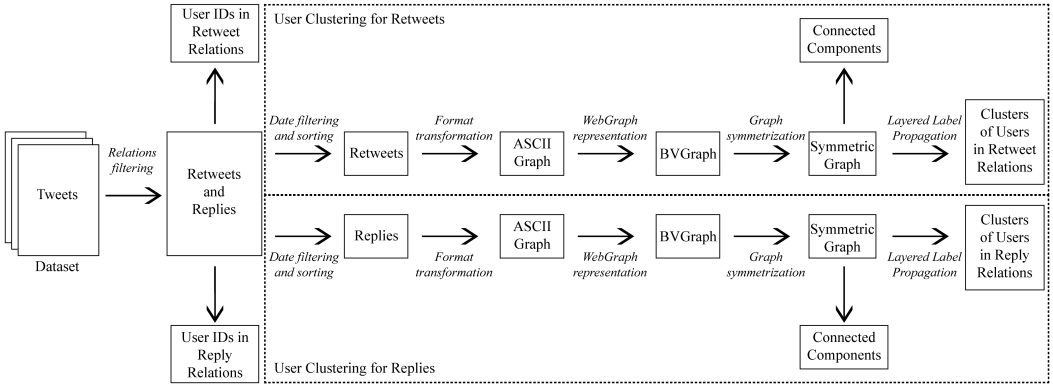

Figure 2: User clustering process.

and replies are mutually exclusive and replies represent direct conversations, which may not be necessarily true with mentions.

We started by filtering all retweets and replies from the dataset, converting them from JSON to a condensed format *"type tweetID userID receiverID timestamp"*. To analyze the clusters in different periods of time, we filtered and sorted the set of retweets and replies by their timestamp values, according to the desired time interval. We also separated retweets and replies from each other, for independent analysis.

Once we were dealing with networks with millions of nodes and edges, we chose to use WebGraph [2] [27] to build and analyze their graphs, and used LAW software library [3] for clustering them. Besides compressing the ASCIIGraph to the WebGraph's format BVGraph, we had to symmetrize it to an undirected and loop-less graph to be used by the LAW implementation of the Layered Label Propagation algorithm, to do user clustering. The symmetric graph was also used to calculate the connected components of the network.

The Layered Label Propagation algorithm [28] is an iterative strategy that reorders the graph such that nodes with the same label are close to one another. This node reordering is useful for graph compression, however, for our purposes we only require the node labeling assignment produced by the label propagation algorithm that returns a clustering configuration of the graph. The clustering result is mappable with the sorted list of user IDs, and all these steps are outlined in Figure 2.

4.2 Tweet clustering

At the end of the User Clustering stage, we get a list of cluster labels that is mappable with the list of user IDs. With these two lists we are able to know the cluster that each user belongs to. Our

[2] `http://webgraph.di.unimi.it/`
[3] `http://law.di.unimi.it/software.php`

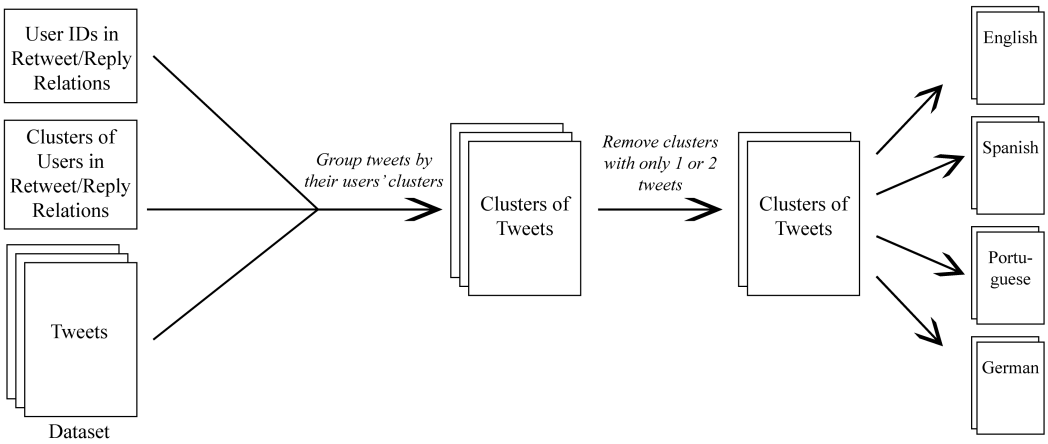

Figure 3: Tweet clustering process.

strategy to classify the sentiment of a cluster is getting the tweets that the users in that cluster tweeted during the lifetime of the cluster, and then classify each one, independently, to sum up an overall result. For that, first, we extracted from the dataset all the tweets created in the same period of time used to cluster the users, then we converted them to the shorter format *"userID tweetID language epochTimestamp hashtagCounter URLCounter mentionCounter tweetText"*. All the clusters with only one or two tweets were removed. Each cluster of tweets was filtered and divided by its prevalent language, in order to perform the sentiment classification without mixed languages.

4.3 Sentiment Analysis

We chose the lexicon-based SentiStrength tool [22] to perform automatic sentiment classification of the tweets, because (1) it does not require training data when working in unsupervised mode; (2) it has good performance and it is able to process more than $16,000$ tweets/second in standard machines; (3) and has good results on Twitter datasets [22, 23]. Giving a text file as input, SentiStrength outputs another file with each line of text of the input file annotated with two sentiment values: a positive integer $s_+ \in \{1, ..., 5\}$ and a negative integer $s_- \in \{-5, ..., -1\}$. The higher the absolute value, the higher the polarity strength of that value.

To classify the tweets in each cluster of tweets we filtered only the tweet text. To avoid words out of context that could be matched by SentiStrength, we removed all the mentions, retweet indicatives and URLs occurrences in the text. After running SentiStrength over the clusters of tweets we got, for each cluster, a matching file with the classified sentiment annotated for each tweet.

4.4 Influence and Sentiment Homophily Analysis over Time

The user clustering, tweet clustering and sentiment analysis stages were scripted to extract the information about the clusters in the network and their sentiment, during desired time intervals. For our analysis we performed a round-based clustering for each round of the knock-out stage subset, which includes the round of 16, quarter-finals, semi-finals and final stage of the World Cup.

Since we were seeking an overall sentiment, we chose to condensate the two sentiment values in one unique value, calculating the Absolute Sentiment value,

$$|s| = s_+ + s_-, \in \{-4, ..., 0, ..., 4\} \tag{1}$$

This way, a tweet is positive with a strength between 1 and 4, neutral when 0, or negative with a strength between -1 and -4. This approach promotes clearly polarized sentiment results and penalizes balanced strength results. This way, the results $(5, -5)$, $(4, -4)$, $(3, -3)$, $(2, -2)$, which

we consider ambiguous results, have the same absolute sentiment of 0 as the SentiStrength neutral result $(1, -1)$.

We focused on polarity changes over time and we calculated the distribution of the absolute sentiment values per hour, in each cluster, by counting the number of tweets for each absolute sentiment result. By analyzing these distributions over time we were able to observe sentiment dynamics and detect sentiment homophily, when existing.

To systematically find periods of polarity homophily, assuming that sentiment homophily is found locally in time, we defined a time window t, a minimum number of tweets m needed to consider a sentiment prevalence in t, and minimum rate of polarity prevalence p in t, as metric for sentiment homogeneity. Let $\Delta t(x_1, x_2)$ be the time interval between two tweets, and $pol(x_1, ..., x_n)$ be the rate of the prevalent polarity in a sequence of tweets, there is sentiment homophily for a sequence of tweets $x_1, x_2, ..., x_n$ when,

$$n \geq m \wedge pol(x_1, ..., x_n) \geq p \wedge \forall \{x_i, x_{i+1}, ..., x_{i+m}\} \in \{x_1, x_2, ..., x_n\}, \Delta t(x_i, x_{i+m}) \leq t. \quad (2)$$

However, finding time intervals that satisfy this metric does not show if there is an increased chance of any user in that cluster of sharing the same befitting sentiment with the overall sentiment that surrounds him, i.e., being influenced by his peers' mood. Our approach to evaluate whether moments of sentiment homophily are caused by influence is to look for ambiguous tweets in moments of prevalent polarized sentiment in the cluster, to which we assign that same prevalent polarization, and then we compare this updated sentiment classification with human coders classifications. To evaluate the extent of sentiment homogeneity in those periods we used K-fold Cross Validation [29].

Lets assume the pairs $(1, -1), (2, -2), (3, -3), (4, -4), (5, -5)$ as ambiguous results in polarized clusters. The reason for this assumption regarding $(2, -2), (3, -3), (4, -4), (5, -5)$ is that they reveal sentiment strength but not a decided polarization, even in a polarized environment. We also include $(1, -1)$ because SentiStrength outputs this value both for neutral sentences and for sentences that do not match any word in the lexicon, which gives a dubious meaning to this value. This way, we trust more in polarized pairs.

After identifying ambiguous results, we search for an ambiguity a that has a number of surrounding tweets equal or greater than m, with a prevalence of a certain polarity equal or greater than p during a period of time t that includes a. For each ambiguity a, found in a context with these characteristics, we set its polarity to be the same as the prevalent polarity of the tweets surrounding it. We proposed two algorithms, that only differ in the position that the ambiguity occupies in the context configuration.

The first algorithm searches for ambiguities that have a central position in the polarized context, being fixed at the center of the time window. For a set of ambiguities A found in a sequence of tweets $T = \{x_1, ..., x_n\}$, when $x_a \in A \wedge x_a \in T$, and

$$\exists x_b, x_e \in T, (b \leq a < e \vee b < a \leq e) \wedge \Delta t(x_b, x_a) \leq \frac{t}{2} \wedge \Delta t(x_a, x_e) \leq \frac{t}{2} \wedge e - b \geq m \wedge pol(x_b, x_e) \geq p, \quad (3)$$

The sentiment polarity of x_a is relabeled with the prevalent sentiment polarity in $x_b, ..., x_e$.

The second algorithm considers any ambiguity that belongs to a sliding time window t that fulfills those restrictions, independently of its position towards the context. For a set of ambiguities A found in a sequence of tweets $T = \{x_1, ..., x_n\}$, when $x_a \in A \wedge x_a \in T$, and

$$\exists x_b, x_e \in T, (b \leq a < e \vee b < a \leq e) \wedge \Delta t(x_b, x_e) \leq t \wedge e - b \geq m \wedge pol(x_b, x_e) \geq p, \quad (4)$$

The sentiment polarity of x_a is relabeled with the prevalent sentiment polarity in $x_b, ..., x_e$.

5 Results and discussion

We used the Modularity coefficient Q, that measures the division of the nodes in a graph into different clusters and the strength of their connections [30], to evaluate the quality of the clusters obtained with Layered Label Propagation algorithm. For clusters obtained from retweet-relation graphs we got an average of $Q = 0.620$, while for reply-relation graphs this value increased for $Q = 0.800$.

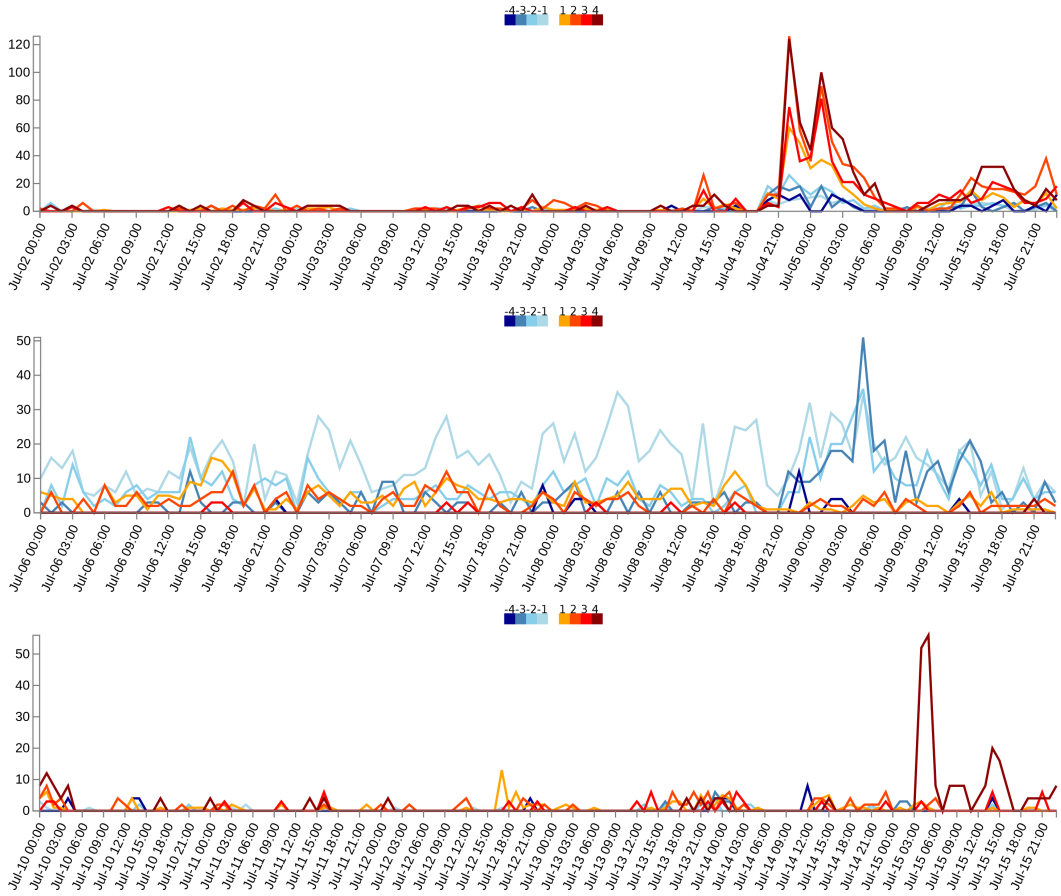

Figure 4: Time-line of tweets' frequency of absolute sentiment for each accumulation of 3 hours. Cluster "413547" from the Spanish-speaking set of reply-based clusters over the quarter-finals stage, cluster "1000883" from the English-speaking set of reply-based clusters over the semi-finals, and cluster "2049176" from the Spanish-speaking set of retweet-based clusters over the final stage.

This denotes that reply-relations are more restrict than retweets and generate smaller but denser clusters. The size distribution of all sets of clusters followed a power-law, regardless the round, language, or type of relation of the graphs.

Considering hypothesis *H1* and *H2* we can observe in Figure 4 that sentiment is highly dynamic, especially for reply-based clusters. With periods of sentiment neutrality interleaved with periods of sentiment polarity, there are moments in which a certain polarity prevails, even though they appear to be quite ephemeral. It is in these moments that we find periods of local sentiment homophily.

Regarding *H3*, we gathered 24 human-coders, in which 23 of them are Portuguese native-speakers and the remaining one is a Spanish native-speaker. All of them are able to read and interpret English, and 18 are also able to read and interpret Spanish. We shuffled them in 8 groups of 3, and each group evaluated two sets of 100 ambiguous tweets each. This way, each ambiguity was classified by three different human-coders.

The testing samples were randomly collected from the set of ambiguous tweets found with the sliding window algorithm, using the fixed parameters $t = 6$, $m = 10$, and $p = 0.7$. These samples sum a total of $1,600$ ambiguous tweets, divided into 800 for English, 600 for Spanish, and 200 for Portuguese. Half of the sets for each language was extracted from retweet-based clusters, and the other half from reply-based clusters.

Set of Ambiguities		*Human-coders Agreement*			*Cluster Sentiment Polarity Mismatch*			*Neutral Sentiment*			*Cluster Sentiment Polarity Match*		
		≥ 2	*Unanimity*	*Total Disagreement*	≥ 2	*Unanimity*	*Random*	≥ 2	*Unanimity*	*Random*	≥ 2	*Unanimity*	*Random*
	RT	92.50%	36.25%	7.50%	30.00%	28.28%	29.00%	24.86%	20.69%	37.75%	45.14%	**51.03%**	33.25%
en	*RE*	93.75%	39.00%	6.25%	16.80%	20.51%	27.00%	36.53%	33.97%	34.75%	**46.67%**	45.51%	38.25%
	Total	93.13%	37.63%	6.88%	23.36%	24.25%	28.00%	30.74%	27.57%	36.25%	45.91%	**48.17%**	35.75%
	RT	86.33%	32.00%	13.67%	32.43%	32.29%	36.33%	15.83%	7.29%	30.67%	51.74%	**60.42%**	33.00%
es	*RE*	90.33%	37.67%	9.67%	34.32%	33.63%	36.33%	17.71%	7.96%	33.00%	47.97%	**58.41%**	30.67%
	Total	88.33%	34.83%	11.67%	33.40%	33.01%	36.33%	16.79%	7.66%	31.83%	49.81%	**59.33%**	31.83%
	RT	87.00%	29.00%	13.00%	32.18%	27.59%	33.00%	29.89%	24.14%	32.00%	37.93%	**48.28%**	35.00%
pt	*RE*	95.00%	42.00%	5.00%	36.84%	42.86%	35.00%	20.00%	9.52%	37.00%	43.16%	**47.62%**	28.00%
	Total	91.00%	35.50%	9.00%	34.62%	36.62%	34.00%	24.73%	15.49%	34.50%	40.66%	**47.89%**	31.50%
	RT	89.50%	33.75%	10.50%	31.15%	29.63%	32.25%	22.21%	16.30%	34.38%	46.65%	**54.07%**	33.38%
Global	*RE*	92.63%	38.88%	7.38%	25.78%	28.30%	31.50%	27.53%	21.22%	34.38%	46.69%	**50.48%**	34.13%
	Total	91.06%	36.31%	8.94%	28.41%	28.92%	31.88%	24.91%	18.93%	34.38%	46.67%	**52.15%**	33.75%

Table 2: Manual evaluation results regarding the approach implemented in the sliding window algorithm, and comparison with a random approach.

Each person was asked to classify the sentiment expressed in the tweet message, as *positive*, *neutral*, or *negative*. We chose to only ask for the polarity and not the sentiment strength to simplify the classification process. We included the *neutral* option assuming that there are indeed some tweets that do not express any kind of polarization.

The results in Table 2 suggest a tendency for the real sentiment of ambiguous tweets to match the overall sentiment of their clusters, over having a neutral or mismatching sentiment polarity, and this value is clearly higher than it would be assigned by chance. However, this matching rate is not sufficient to claim that when there is a period of sentiment homophily there is a strong chance of a user in that cluster sharing a tweet with an equivalent polarity.

We evaluated the reliability of the human coder classifications in terms of agreement using the Krippendorff's alpha-coefficient [31], which varied between 0.24703 and 0.53167, i.e, they are statistically reliable but with a certain level of disagreement, unveiling the subjective nature of this task.

With K-Fold Cross Validation we tested the error rate of both fixed and sliding window algorithms, which gives the extent of homogeneity in periods of sentiment homophily. The results in Table 3 show that homogeneity is stronger in retweet-based clusters and decreases when the considered time window increases.

6 Conclusion and future work

With this work we observed that sentiment reveals a highly dynamic behavior at a cluster level, having ephemeral spikes of polarity usually lasting for a few hours. We were able to find those spikes of sentiment homogeneity by setting a time window t, a minimum number of tweets m needed to consider a sentiment prevalence in t, and minimum rate of polarity prevalence p in t. For understanding if an existing overall sentiment in a cluster may influence the sentiment of its individuals, we relabeled the sentiment of ambiguous classifications surrounded by a context of sentiment homophily with the prevalent sentiment of that cluster during t and we evaluated this extrapolation with human coders. The matching rate between the human-coders classification and the clusters' sentiment polarity extrapolation always shows higher and more stable expressiveness over mismatching and neutral rates. However, with the best matching result around 60%, we can only say we found a weak but statistically significant tendency of a user sharing a befitting sentiment in a cluster during a period of sentiment homogeneity. The K-fold Cross Validation unveiled that this homogeneity is usually stronger in retweet-based clusters.

Given the level of disagreement between human coders it would be desirable to use an higher odd number of coders for each evaluation set. In the future it would be interesting to separate neutral sentiment classifications from undecidable sentiment classifications, which have the same value $(1, -1)$ when classified by SentiStrength, and see what would happen to the rate of neutral classifications among the human coder classifications.

Stage	*Language*	*Type*	*Fixed time window*				*Sliding time window*			
			$t = 1$	$t = 3$	$t = 6$	$t = 12$	$t = 1$	$t = 3$	$t = 6$	$t = 12$
Round of 16	*en*	*RT*	15.47%	16.64%	17.46%	17.62%	16.60%	17.64%	17.99%	18.41%
		RE	20.55%	22.04%	22.38%	22.73%	22.55%	22.61%	23.06%	22.92%
	es	*RT*	14.63%	17.93%	18.78%	18.61%	17.82%	18.97%	19.23%	19.01%
		RE	21.55%	20.87%	20.57%	19.47%	21.18%	20.04%	19.80%	18.55%
	pt	*RT*	10.00%	13.36%	13.90%	16.51%	14.20%	15.71%	17.32%	17.17%
		RE	30.00%	25.00%	24.00%	21.39%	22.50%	22.50%	24.17%	23.82%
Quarter-finals	*en*	*RT*	14.98%	15.23%	15.48%	16.52%	15.68%	16.08%	17.11%	17.67%
		RE	19.69%	20.33%	20.18%	21.76%	20.36%	21.34%	22.65%	22.20%
	es	*RT*	13.81%	14.97%	14.34%	15.72%	13.42%	15.35%	17.14%	17.65%
		RE	22.24%	21.59%	22.48%	22.93%	21.64%	22.98%	22.68%	23.62%
	pt	*RT*	14.88%	14.69%	11.97%	15.20%	17.59%	14.01%	15.78%	15.45%
		RE	18.75%	22.74%	22.35%	21.38%	20.38%	20.95%	22.09%	21.36%
Semi-finals	*en*	*RT*	14.75%	16.23%	16.56%	17.12%	16.35%	16.98%	17.40%	17.83%
		RE	19.54%	19.55%	20.13%	20.62%	20.50%	20.82%	21.34%	21.30%
	es	*RT*	15.15%	17.15%	16.64%	17.67%	16.82%	18.06%	18.40%	18.66%
		RE	20.68%	23.27%	21.98%	21.80%	22.84%	23.28%	23.29%	23.06%
	pt	*RT*	16.83%	14.61%	15.85%	16.70%	17.14%	14.94%	17.57%	18.31%
		RE	18.13%	16.61%	22.06%	22.80%	16.95%	21.88%	22.84%	24.62%
Final	*en*	*RT*	13.78%	14.48%	15.00%	16.09%	14.81%	15.50%	16.44%	17.57%
		RE	17.72%	19.91%	20.06%	21.22%	20.04%	21.28%	21.48%	21.87%
	es	*RT*	19.09%	14.14%	16.96%	16.10%	16.75%	18.17%	17.60%	18.24%
		RE	22.79%	22.67%	23.69%	23.73%	24.22%	22.47%	23.76%	23.83%
	pt	*RT*	11.03%	15.42%	14.10%	15.21%	18.25%	15.77%	15.24%	15.76%
		RE	18.38%	22.78%	25.75%	25.69%	18.10%	25.98%	26.26%	24.63%
%			**17.69%**	**18.43%**	**18.86%**	**19.36%**	**18.61%**	**19.30%**	**20.03%**	**20.15%**

Table 3: Error rate E average of K-Fold Cross Validation, for $k = 10$, over sets of tweets in periods of prevalence of a certain sentiment polarity.

References

[1] Twitter. About twitter - twitter.com, 2015. [Online at `https://about.twitter.com/company`; accessed 2015-September-19].

[2] David Easley and Jon Kleinberg. *Networks, Crowds, and Markets: Reasoning About a Highly Connected World*. Cambridge University Press, New York, NY, USA, 2010.

[3] Jure Leskovec, Daniel Huttenlocher, and Jon Kleinberg. Signed networks in social media. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '10, pages 1361–1370, New York, NY, USA, 2010. ACM.

[4] Bernardo Huberman, Daniel Romero, and Fang Wu. Social networks that matter: Twitter under the microscope. *First Monday*, 14(1), 2008.

[5] Julian Mcauley and Jure Leskovec. Discovering social circles in ego networks. *ACM Trans. Knowl. Discov. Data*, 8(1):4:1–4:28, Feb 2014.

[6] VA Traag and Jeroen Bruggeman. Community detection in networks with positive and negative links. *Physical Review E*, 80(3):036115, 2009.

[7] Takeshi Sakaki, Makoto Okazaki, and Yutaka Matsuo. Earthquake shakes twitter users: Real-time event detection by social sensors. In *Proceedings of the 19th International Conference on World Wide Web*, WWW '10, pages 851–860, New York, NY, USA, 2010. ACM.

[8] Cynthia Chew and Gunther Eysenbach. Pandemics in the age of twitter: Content analysis of tweets during the 2009 h1n1 outbreak. *PLoS ONE*, 5(11):e14118, 11 2010.

[9] Justin Cheng, Lada Adamic, P. Alex Dow, Jon Michael Kleinberg, and Jure Leskovec. Can cascades be predicted? In *Proceedings of the 23rd International Conference on World Wide Web*, WWW '14, pages 925–936, New York, NY, USA, 2014. ACM.

[10] Andranik Tumasjan, Timm Sprenger, Philipp Sandner, and Isabell Welpe. Predicting elections with twitter: What 140 characters reveal about political sentiment. 2010.

[11] Seth A. Myers and Jure Leskovec. The bursty dynamics of the twitter information network. In *Proceedings of the 23rd International Conference on World Wide Web*, WWW '14, pages 913–924, New York, NY, USA, 2014. ACM.

[12] Eytan Bakshy, Itamar Rosenn, Cameron Marlow, and Lada Adamic. The role of social networks in information diffusion. In *Proceedings of the 21st International Conference on World Wide Web*, WWW '12, pages 519–528, New York, NY, USA, 2012. ACM.

[13] Eytan Bakshy, Jake M. Hofman, Winter A. Mason, and Duncan J. Watts. Everyone's an influencer: Quantifying influence on twitter. In *Proceedings of the Fourth ACM International Conference on Web Search and Data Mining*, WSDM '11, pages 65–74, New York, NY, USA, 2011. ACM.

[14] Meeyoung Cha, Hamed Haddadi, Fabrício Benevenuto, and Krishna P. Gummadi. Measuring user influence in twitter: The million follower fallacy. In *in ICWSM '10: Proceedings of international AAAI Conference on Weblogs and Social*, 2010.

[15] Mao Ye, Xingjie Liu, and Wang-Chien Lee. Exploring social influence for recommendation: A generative model approach. In *Proceedings of the 35th International ACM SIGIR Conference on Research and Development in Information Retrieval*, SIGIR '12, pages 671–680, New York, NY, USA, 2012. ACM.

[16] Jiliang Tang, Huiji Gao, Xia Hu, and Huan Liu. Exploiting homophily effect for trust prediction. In *Proceedings of the Sixth ACM International Conference on Web Search and Data Mining*, WSDM '13, pages 53–62, New York, NY, USA, 2013. ACM.

[17] Umesh Rao Hodeghatta. Sentiment analysis of hollywood movies on twitter. In *Proceedings of the 2013 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining*, ASONAM '13, pages 1401–1404, New York, NY, USA, 2013. ACM.

[18] Amir Asiaee T., Mariano Tepper, Arindam Banerjee, and Guillermo Sapiro. If you are happy and you know it... tweet. In *Proceedings of the 21st ACM International Conference on Information and Knowledge Management*, CIKM '12, pages 1602–1606, New York, NY, USA, 2012. ACM.

[19] J.H. Fowler and N.A Christakis. Dynamic spread of happiness in a large social network: longitudinal analysis over 20 years in the framingham heart study. *British Medical Journal*, 337:a2338, 2008.

[20] Mike Thelwall. Homophily in myspace. *Journal of the American Society for Information Science and Technology*, 60(2):219–231, 2009.

[21] Mike Thelwall. Emotion homophily in social network site messages. *First Monday*, 15(4), 2010.

[22] M. Thelwall, K. Buckley, and G. Paltoglou. Sentiment strength detection for the social Web. *Journal of the American Society for Information Science and Technology*, 63(1):163–173, 2012.

[23] Anatoliy Gruzd, Sophie Doiron, and Philip Mai. Is happiness contagious online? a case of twitter and the 2010 winter olympics. In *Proceedings of the 2011 44th Hawaii International Conference on System Sciences*, HICSS '11, pages 1–9, Washington, DC, USA, 2011. IEEE Computer Society.

[24] Rui Fan, Jichang Zhao, Yan Chen, and Ke Xu. Anger is more influential than joy: Sentiment correlation in weibo. *PLoS ONE*, 9(10):e110184, 10 2014.

[25] Johan Bollen, Bruno Gonçalves, Guangchen Ruan, and Huina Mao. Happiness is assortative in online social networks. *Artif. Life*, 17(3):237–251, Aug 2011.

[26] Jiliang Tang, Yi Chang, and Huan Liu. Mining social media with social theories: A survey. *SIGKDD Explor. Newsl.*, 15(2):20–29, June 2014.

[27] P. Boldi and S. Vigna. The webgraph framework i: Compression techniques. In *Proceedings of the 13th International Conference on World Wide Web*, WWW '04, pages 595–602, New York, NY, USA, 2004. ACM.

[28] Paolo Boldi, Marco Rosa, Massimo Santini, and Sebastiano Vigna. Layered label propagation: A multiresolution coordinate-free ordering for compressing social networks. In *Proceedings of the 20th International Conference on World Wide Web*, WWW '11, pages 587–596, New York, NY, USA, 2011. ACM.

[29] A. R. Webb and K. D. Copsey. *Statistical pattern recognition*, chapter 13.1.2. Wiley and Sons Publishing, 3rd edition, 2011.

[30] Mark Newman. *Networks: An Introduction*. Oxford University Press, Inc., New York, NY, USA, 2010.

[31] Klaus Krippendorff. Computing krippendorff's alpha reliability. Technical report, University of Pennsylvania, Annenberg School for Communication, Jun 2011.