A Constellation of Horrors: Analysis and Visualization of the #Cuéntalo Movement

María Soledad Bucalo
Luz Calvo
Fernando Cucchietti
David García Povedano
Artur García-Sáez
Barcelona Supercomputing Center
(BSC)
Barcelona, Spain

Juan Felipe Gómez
Camilo Arcadio González
Guillermo Marín
Patricio Reyes
Feliu Serra
Diana Fernanda Vélez
Barcelona Supercomputing Center
(BSC)
Barcelona, Spain

Irene Meta
Barcelona Supercomputing Center
(BSC)
Barcelona, Spain
Institute for Advanced Architecture
of Catalonia (IAAC)
Barcelona, Spain

ABSTRACT

In this work, we analyze content and structure of the Twitter trending topic #cuentalo with the purpose of providing a visualization of the movement. A supervised learning methodology is used to train the classifying algorithms with hand-labeled observations. The methodology allows us to classify each tweet according to its role in the movement.

KEYWORDS

visualization; machine learning; supervised learning; social networks

1 INTRODUCTION

The #cuentalo (hashtag for *cuéntalo*, which means *tell it* in Spanish) movement started on April 2018, inviting women to share on Twitter their personal experiences of sexual aggression. The movement was triggered in Spain with the news about the court decision in the so called *wolf pack* case [1]. In a few days, it generated more than 2.5 million tweets and retweets with narrations told by their protagonists.

The tweets go from the uncomfortable to the unbearable, stories in first person occasionally mixed with a woman speaking in the name of another one because she doesn't have a computer, or because she doesn't dare to tell her story, or because she was murdered.

Most cases of sexual aggression go unreported [6]. Many of the tweets and users involved in this viral phenomenon mentioned that this was the first time in their lives they talked about it. A related phenomenon was the #MeToo movement [11], although with the subtle difference that in Spanish, #MeToo induces a sense of appealing to be similar to the celebrities that started the movement, while #cuentalo is a more anonymous call asking women to end the silence and tell what happened to them, implicitly assuming that most have experienced something similar. In order to bring attention to the topic, to support the debate around it, and to help create and maintain a new collective memory, we conducted an analysis of the data produced by the movement.

In this research work we propose a methodology to analyze the content and structure of the Twitter trending topic #cuentalo which serves as the basis for a visualization tool that uncovers information related to the problem of sexual aggression. The analysis results in new perspectives and information about the frequency, types, and other details of sexual aggression.

2 RELATED WORK

Due to the similar nature of the movement, special attention deserves the literature about #MeToo. In [11], authors study common words, semantic relationships, sentiment analysis from content found on both Twitter and Reddit. Their results focus on the differences in content regarding these two platforms.

Concerning data visualization, most of the references found are data-visualization projects which analyze #MeToo and display the results on a website [5, 7]. Prior to the #MeToo movement, Jain [10] presents descriptive statistics about rape in India, with data from the justice system.

Regarding #cuentalo, with a similar dataset, in *Contando cómo se difundió el #cuéntalo* [4] the author performs a social network analysis to identify the topology of the network defined by the relationships between users (follower) or tweets (re-tweets).

3 METHODOLOGY

3.1 Dataset description

The raw dataset corresponds to the tweets published between April 27 and May 12, 2018 (except 4 days) with the hashtag #cuentalo.

This dataset contained 2.1 million tweets. It was collected by members of the *Asociación de Archiveros-Gestores de Documentos de Cataluña (AAC-DG)*. The collection was missing the days: May 29, June 4, June 6, and June 11. For these days, we were able to recover the tweets with original content (no retweets), bringing the final dataset to 2.75 million tweets in total. Although less than 3% of tweets had proper geolocation information, by reading and parsing the *location* field we were able to locate city or country information for roughly half of all the tweets in the dataset. There were over 160 thousand tweets with content written by users, while the rest (2.6 million) were retweets. Although retweets were crucial for the virality of the movement, here we focus only on those with content, that we denote as *original*.

3.2 Classification techniques

3.2.1 Training database. The dataset [12] used for training purposes deserves a special mention. It consists of around 10632 randomly chosen original tweets categorized by volunteers specifically for this project. This corresponds to 6.6% of the original tweets. Only tweets in Spanish were categorized, which helped standardize the training dataset, and was the language of the vast majority of the tweets.

The categorization process was as follows: Each volunteer received a table, one row per tweet and one column per category or label, where each volunteer marked *yes/no* if the tweet content corresponds to the labels in **what** (see description below), except for the category **who**, where the volunteer chose between 5 possibilities. For each tweet, the volunteer had access to (1) the tweet id, (2) the tweet user name and (3) the content of the tweet.

The categories and labels were the following:

- **who**: (mutually exclusive choices: *1–5*) the writer of the tweet either:
  1. tells something about himself/herself.
  2. tells something about someone else.
  3. tells something supporting the movement.
  4. tells something not related to the movement.
  5. tells something against the movement.
- **what**:
  - **murder**: (*yes/no*) the tweet describes a murder
  - **rape**: (*yes/no*) the tweet states a rape or attempted rape
  - **sexual assault**: (*yes/no*) the tweet describes a sexual assault (but not situations that belongs to the category **rape**)
  - **abuse**: (*yes/no*) the tweet states abuse.
  - **harassment**: (*yes/no*) the tweet discusses about a non-physical harassment situation
  - **fear**: (*yes/no*) the tweet explicitly describes fear
  - **disgust/sadness/anger**: (*yes/no*) the tweet explicitly describes disgust/sadness/anger, only if **who** is different than 4.

Among the 10.632 categorized tweets, 31% correspond to testimonies written in first person (i.e. the answer to **who** is 1), 8.9% in third person or on behalf of someone else (**who** is 2), 40.2% are supportive tweets (**who** is 3), 3.1% correspond to tweets against the movement (**who** is 5), and 16.7% are tweets not related to the movement (unclassifiable) (**who** is 4). Percentages are depicted in Figure 1 and Figure 2.
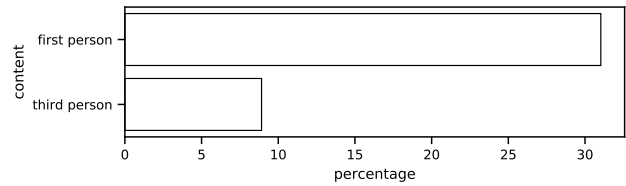

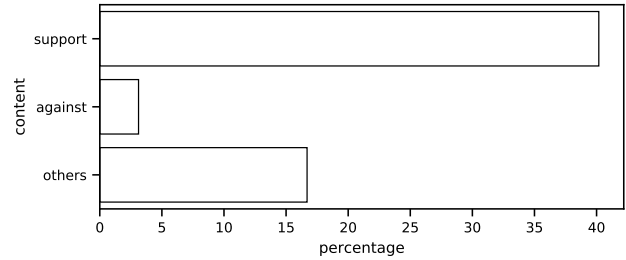**Figure 1: Type of content in original tweets according to the writer**

Figure 2: Type of content in the original tweets according to the support to the movement

Among the tweets with testimonies (written in either 1st or 3rd person), 3.9% talk about a murder, 5.6% about rape, 11.2% about a sexual assault, 6.3% about abuse, 14.2% harassment, 11.8% mention fear, and 19% disgust/sadness/anger (See Figure 3).
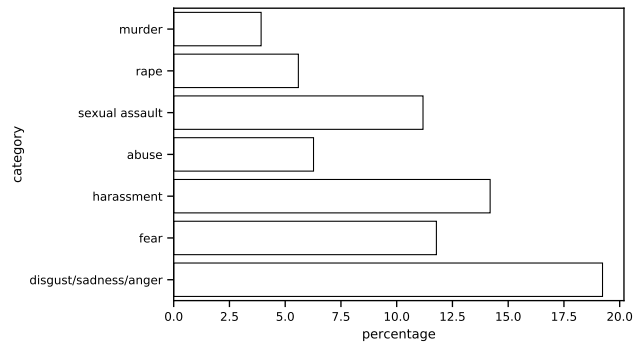

Figure 3: Type of content in original tweets according to the situation and sentiment

3.2.2 Classification process. The cost of classifying tweets manually is very large and can be imprecise because of human biases. Instead of cross-checking results between classifiers, we set some strict rules to help homogenize the classification criteria, which in our case turned out to be the most cost-effective approach. For example, the tweets containing images, text screen-captures, or written in a language other than Spanish were categorized as not related to the movement (**who**=4).

We did not cross-check the classification from different volunteers. However, all of them reached very similar percentages of the

different categories, hence we assume that the classification was precise enough.

Notice that our classification criteria might be seen as not sufficiently specific to describe correctly or in sufficient detail such a dramatic topic, but follows a critical discussion trying to balance important factors like frequency or duration of the aggression against technical limitations.

3.2.3 Supervised learning. Due to the properties of the dataset, we decided to use a supervised learning methodology in order to train the algorithms. We settled on a deep neural network multi-output classifier built using the Keras API to Tensorflow [3]. The version of the Keras package used is the 2.1.5.

3.2.4 Aggregation of outputs and cross validation. The complexity of the natural language used in tweets, such as spelling mistakes or occasional mixed languages, makes the multi-output classification even hard for humans. In order to improve the accuracy we had to narrow down the multi-output problem, and ended up with a simplification of the categories as follows: The answers for **who** were re-grouped as **testimony**/**support**/**others**: a tweet belongs to **testimony** if **who** $\in \{1, 2\}$, to **support** if **who**=3 or to **others** if **who** $\in \{4, 5\}$. In the same way, the answers for **what** were re-grouped as **physical**/**non-physical**/**others**: a tweet belongs to **physical** if any of the labels **murder**, **rape**, **sexual assault**, **abuse** was answered as *yes*; it belongs to **non-physical** if any of **harassment**, **fear**, **disgust/sadness/anger** was answered as *yes*; finally, it belongs to **others** if none of the labels in **what** were affirmative. By aggregating groups in this way we made the learning task easier for the learning algorithm, improving from an accuracy of 56% on the 5 output **who** classification to 75% on the 3 output classification problem. The re-grouping of **what** improved the accuracy from 56% on the 7 multi-output classification to a 68% accuracy on the grouped 3 categories (See section 3.2.5).

For the cross-validation we used a 60%/20%/20% split for the training, validation and testing sets respectively.

3.2.5 Neural network architecture. We tested different architectures, and here we report two of them. The first architecture A1, has the pretrained GloVe [14] word embedding, then a *dense* layer, an *activation* layer, and finally a *flatten* layer and a *dense* layer. The second architecture A2, was used to perform the final classification. It starts with the GloVe word embedding. After a dropout layer to regularize the embedding signal, it uses a convolutional layer with its respective maxpooling layer to reduce the dimensionality of the data. Next, long-short term memory layer, with a dropout layer, dense layer and a batch normalization layer in order to avoid over fitting. Finally, the dimensionality was reduced using another dense layer with a final softmax output layer for the multioutput classification task.

Table 1 shows a classification summary of the test set made using sklearn package's classification report method [8, 13]. It shows the precision for each category predicted, the recalls or true positive rate and the support or number of observations that have been hand classified as belonging to those classes. Table 2 shows the accuracy per each architecture (A1, A2) and category (who,what) considered.

Table 1: Out of sample classification summary

		precision		recall		support
		A1	A2	A1	A2	
who	testimony	0.80	0.70	0.88	0.92	772
	support	0.79	0.84	0.76	0.65	836
	others	0.64	0.73	0.56	0.65	392
	avg/total	0.73	0.76	0.71	0.75	2,000
what	physical	0.62	0.77	0.61	0.63	437
	non-physical	0.53	0.55	0.44	0.79	710
	others	0.65	0.82	0.75	0.60	853
	avg/total	0.60	0.71	0.61	0.67	2,000

Table 2: Accuracy according to the models and categories

	A1	A2
who	0.77	0.75
what	0.63	0.68

As we can see in Tables 1 and 2, there is no significant improvement between A1 and A2. This is partly due to the GloVe word embedding being an important part of the process.

We tried other variations of the architecture, such as using bidirectional LSTMs or stacking multiples LSTMs with their respective dropout layers, but the learning algorithm tended to overfit when there were too many parameters to tune considering the small amount of observations of the training dataset.

4 RESULTS

The results (See Tables 1 and 2) of the methodology above allows us to classify each tweet in two ways. First, into three mutually exclusive categories: Testimony, support and others. Second, the three classes regarding what the tweet was about: Physical content, non-physical content or others (See section 3.2.4). Each tweet was given the probability of belonging to each one of these categories. We show in Figure 4 the resulting classification using a Ternary plot [9]. The plot shows each classified tweet as a dot where the probability assigned to each of the three categories is represented as one minus the distance to one of the vertices of the triangle. Then, an almost perfect classification would show all dots near the vertices. Figure 4 shows relatively few dots near the center, which would indicate confusion.

5 VISUALIZATION

After the tweets were classified, a visualization that uncovers information related to the problem of sexual aggression was designed with the aim of informing the general audience about the movement, making the content easily available. After several discussions and prototypes [2], we decided that the resulting visualization must satisfy the following statements:

- All the tweets must be represented individually.
- Every tweet must have the same weight.
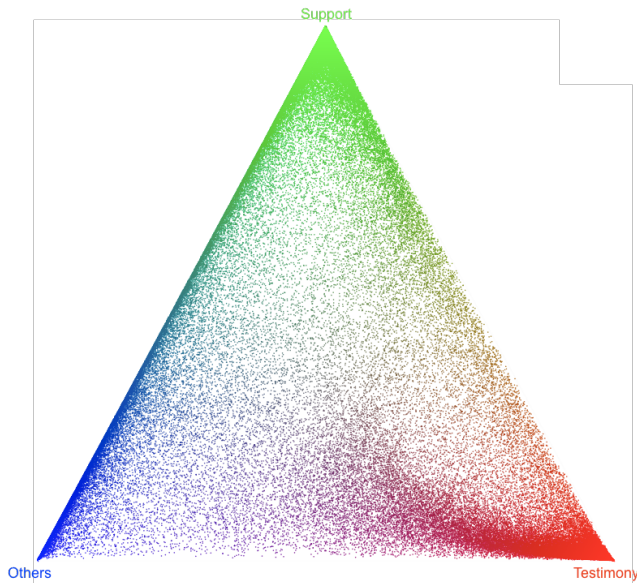- The visualization must be flexible enough to receive new tweets.

Figure 4: Ternary plot for the classification of the original tweets into three main categories.

Taking into account the statements above, and the classification depicted in Figure 4, we decided to classify the tweets with respect to their role in the campaign.

In the resulting visualization (See Figure 5) all the tweets are depicted in a radial distribution, where the angle corresponds to the time when the tweet was published. The radius, which is the distance with respect to the center, is related to the role of the tweet in the movement. The inner, red ring corresponds to testimonial tweets, surrounded by the outer pink ring, which corresponds to supporters. Out of the outer pink ring there are the tweets against or not related to the movement. The online visualization can be seen in the website of the project https://www.bsc.es/viz/cuentalo/.
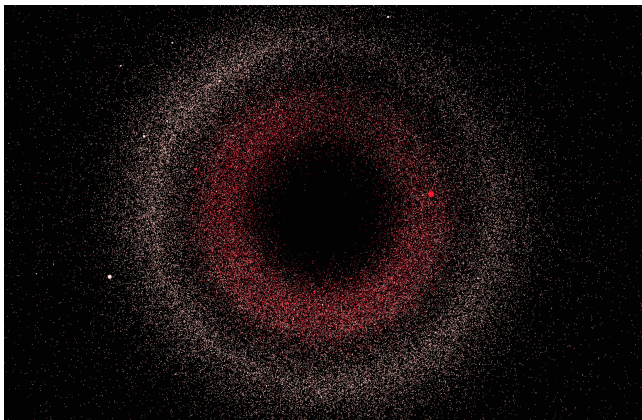

Figure 5: Snapshot of the final visualization.

6 CONCLUSIONS AND FUTURE WORK

We have proposed a methodology which classifies tweets with respect to their role in the #cuentalo movement. A visualization was delivered as a mean to present the general audience with information about sexual aggression in a more appealing manner.

As a byproduct, we have provided a dataset [12] of 10 thousand #cuentalo tweets classified by volunteers which can be useful for future studies of the movement.

As a future work, we are interested in an in-depth data analysis to determine the profile of the protagonists of the tweets. Moreover, we would like to compare the tweets between #cuentalo and the #MeToo movement.

7 ACKNOWLEDGMENTS

REFERENCES

[1] James Badcock. 2018. Fury over Rape Case Sparks Spanish #MeToo. (May 2018). https://www.bbc.com/news/world-europe-43960647
[2] bscvizteam. 2018. #Cuéntalo – La Visualización – The BSC Viz Corner. http://www.bsc.es/viz/corner/?p=210&lang=es#comments
[3] François Chollet et al. 2015. Keras. (2015).
[4] Mariluz Congosto. 2018. Contando Cómo Se Difundió El #cuéntalo. http://www.barriblog.com/2018/05/contando-se-difundio-cuentalo/?lang=en
[5] Valentina D'Efilippo. 2018. The Anatomy of a Hashtag — a Visual Analysis of the MeToo Movement. https://medium.muz.li/the-anatomy-of-a-hashtag-a-visual-analysis-of-the-metoo-movement-ba4ecf9b130b.
[6] Bonnie S. Fisher, Leah E. Daigle, Francis T. Cullen, and Michael G. Turner. 2003. Reporting Sexual Victimization To The Police And Others: Results From a National-Level Study of College Women. *Criminal Justice and Behavior* 30, 1 (Feb. 2003), 6–38. https://doi.org/10.1177/0093854802239161
[7] Lazaro Gamio and Haley Britzky. 2018. What to Know about All of the Men Facing Sexual Misconduct Allegations. https://www.axios.com/sexual-misconduct-allegations-since-weinstein-6da885ce-83a5-40bc-9fae-2b1b65f9c68e.html.
[8] Shantanu Godbole and Sunita Sarawagi. 2004. Discriminative Methods for Multi-labeled Classification. In *Advances in Knowledge Discovery and Data Mining*, Honghua Dai, Ramakrishnan Srikant, and Chengqi Zhang (Eds.). Springer Berlin Heidelberg, Berlin, Heidelberg, 22–30.
[9] Marc Harper et al. 2015. Python-Ternary: Ternary Plots in Python. (Dec. 2015). https://doi.org/10.5281/zenodo.34938
[10] Aditya Jain. 2018. Rape in India. https://adityajain15.github.io/Rape_In_India/.
[11] Lydia Manikonda, Ghazaleh Beigi, Subbarao Kambhampati, and Huan Liu. 2018. #metoo Through the Lens of Social Media. In *Social, Cultural, and Behavioral Modeling (Lecture Notes in Computer Science)*, Robert Thomson, Christopher Dancy, Ayaz Hyder, and Halil Bisgin (Eds.). Springer International Publishing, 104–110.
[12] pareyesv. 2019. BSCCNS/cuentalo-dataset. https://doi.org/10.5281/zenodo.2585526
[13] F. Pedregosa et al. 2011. Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research* 12 (2011), 2825–2830.
[14] Jeffrey Pennington, Richard Socher, and Christopher D. Manning. 2014. GloVe: Global Vectors for Word Representation. In *Empirical Methods in Natural Language Processing (EMNLP)*. 1532–1543.