Sharing Emotions at Scale: The Vent Dataset

Nikolaos Lykousas,[1] **Constantinos Patsakis,**[1]
Andreas Kaltenbrunner,[2] **Vicenç Gómez**[2]

[1]University of Piraeus, Greece
[2]Universitat Pompeu Fabra. Barcelona, Spain

Abstract

The continuous and increasing use of social media has enabled the expression of human thoughts, opinions, and everyday actions publicly at an unprecedented scale. We present the Vent dataset, the largest annotated dataset of text, emotions, and social connections to date. It comprises more than 33 millions of posts by nearly a million users together with their social connections. Each post has an associated emotion. There are 705 different emotions, organized in 63 "emotion categories", forming a two-level taxonomy of affects. Our initial statistical analysis describes the global patterns of activity in the Vent platform, revealing large heterogeneities and certainly remarkable regularities regarding the use of the different emotions. We focus on the aggregated use of emotions, the temporal activity, and the social network of users, and outline possible methods to infer emotion networks based on the user activity. We also analyze the text and describe the *affective landscape* of Vent, finding agreements with existing (small scale) annotated corpus in terms of emotion categories and positive/negative valences. Finally, we discuss possible research questions that can be addressed from this unique dataset.

Introduction

Experiencing emotions is an integral part of human life, which plays significant role in the effective communication of people (Barrett 2006). In fact, sometimes, emotional intelligence is considered more important than cognitive intelligence for successful interaction (Pantic et al. 2005). Naturally, humans have an innate need to share the emotions and feelings they experience, a phenomenon known as *"social sharing of emotions"* (Rime et al. 1991). In the era of social networking, people extensively share what they feel via social media content to express their emotional experiences, reduce dissonance, deepen social connections, and convey their evaluations on a given topic (Berger and Milkman 2012). Given their importance, a lot of research in the field of affective computing has focused on efficiently recognising emotions in online user behavior (Politou, Alepis, and Patsakis 2017). To facilitate research in this field, we collected the data shared in Vent, a social network where users can *"express and share their feelings with people who*
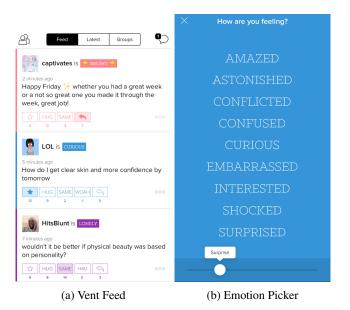
(a) Vent Feed   (b) Emotion Picker

Figure 1: Screenshots of the interface of Vent app.

care"[1]. In this work, we discuss several features and insights of this **complete** dataset (Lykousas et al. 2019).

Vent is a semi-anonymous social networking app that lets users share their feelings and frustrations without the fear of a negative backlash. It encourages users (also referred to as *venters*) to voice their opinion to a supportive community without the worry of being insulted, de-friended or upsetting people they know. Although users must register to use the app, verification is not required, and there is no option for users to connect other social accounts, thus leaving room for anonymity.

Each post (also referred to as *vent*) in the app is associated with a specific emotion by the user who submits it (Figure 1b). "*Emotions*" in the Vent platform are a fuzzy concept, probably better described as "affects", since they include a broad range of feelings, emotions and moods. Apart from posting, users can browse through the feeds of other users' vents (see Figure 1a), and interact with them by commenting or reacting to their vents via a set of emotion-

---

[1]https://www.vent.co/

specific reactions (e.g. hug, same, h4u - here for you). More-over, users can follow others and get updates on their vents, create and join groups, and exchange direct messages with the users they follow.

Maybe the most important feature of our dataset is the existence of self-reported "ground truth" affect annotations associated with each text. The unstructured and noisy nature of user-generated content on the Internet (Cambria et al. 2013), along with the scarcity of ground truth labels regarding affects, consist a major challenge in the field of affective computing, particularly for tasks such as emotion detection and analysis. As highlighted in Mohammad et al. 2018:

It is challenging to obtain consistent annotations for affect due to a number of reasons, including: the subtle ways in which people can express affect, fuzzy boundaries of affect categories, and differences in human experience that impact how they perceive emotion in text.

In the realm of social media, the closely related problem of sentiment analysis has received significant attention, with the bulk of the literature focusing around social networks such as Twitter (Pak and Paroubek 2010; Kouloumpis, Wilson, and Moore 2011; Agarwal et al. 2011) and Facebook (Ortigosa, Martín, and Carro 2014; Wang et al. 2012; Ahkter and Soria 2010; Troussas et al. 2013), mainly due to their market share and wide availability of data.

However emotions are much more expressive than sentiments, and the current approaches for emotion analysis have a long way to go before matching the success and ubiquity of sentiment analysis (Wang and Pal 2015). Nonetheless, the amount of useful information which can be gained by moving past the negative and positive sentiments and towards identifying discrete emotions can help improve many applications.

To this end, a variety of affective lexica have been proposed to offer information about affect expressed in text at a fine level of granularity (Liu and Zhang 2012). Lexica-based approaches, however, have some limitations in the domain of emotion detection, e.g., the lack of coverage, especially in social media/micro-blogging context, and the inherent incapability of recognising sentences without keywords (Kao et al. 2009; Mudinas, Zhang, and Levene 2012).

Other works take different approaches to analyze emotions online. For example, Garcia et al. (2016) use principles of dynamical systems to study the emotional states of a small group of users during their participation in online discussions. Bazarova et al. (2015) performed an experiment in which participants labelled the emotions of their interaction on Facebook. Xu et al. (2017) developed a chatbot for customer service, and their content analysis revealed that more than 40% of the user requests were emotional.

Based on the above, we consider that the Vent dataset can provide a baseline corpus for emotion analysis of the user-generated text. To the best of our knowledge, this is the largest annotated dataset of texts with affects. Moreover, the labelling has been made by the authors of the texts who can classify their texts more accurately according to what they felt when they wrote it, as in (Bazarova et al. 2015). Table 1 shows an illustrative sample of vents and their as-

Table 1: Some illustrative sample vents.

Emotion	Vent text
Sad	I hate fuckin every single person on this fuckin planet. Someone kill me pls
Happy	Best day I have had in a long time :)
Frustrated	i wish it was as easy to forget someone as it is to get attached to them
Stressed	really can't deal with school again today
Anxious	constantly worried that my boyfriend will fall out of love with me
Supportive	Hey guys, just a lil note that it's okay to want attention, it's super natural and there's nothing wrong with it
Affectionate	boy I like called me princess He's so precious
Disappointed	the Highschool life isn't really fun.
Curious	Do really hairy people use shampoo on their body or shower gel? Does it need conditioner?

sociated emotions. We argue that this may overcome many biases, give more insight on how feelings can be expressed in written speech.

Data Collection

The Vent platform is offered exclusively as a mobile application for iOS and Android. To the best of our knowledge, there is no open-source client available at the time of writing, hence, we follow a similar method as in (Siekkinen, Masala, and Kämäräinen 2016; Lykousas, Patsakis, and Gómez 2018) to analyze the network traffic between the app and the backend service. To this end, we employ an SSL-capable man-in-the-middle proxy between a mobile device with the Vent app installed and the Vent service, that acts as a transparent proxy.

The proxy intercepts the HTTPS requests sent by the mobile device and pretends to be the server to the client and the client to the server, enabling us to examine and log the requests and responses between the client app and the server. Based on them, we identified a set of APIs allowing us to collect data about emotions and emotion categories, vents, and user relationships. Some of the APIs had limitations regarding the amount of returned data. For example, the vent feed API returned results up to one month ago. To overcome this limitation, we focused our efforts on obtaining a complete list of usernames and gathering the vents of each user individually.

The collection of the usernames was made thanks to the searching mechanism of Vent. More precisely, when querying the username search API, we observed that the input query was matched from the starting character of each username. Moreover, the API provided unlimited pagination until no more usernames were beginning with the input query. Therefore, we provided a list of all the valid characters for usernames. More precisely, all characters of the English al-

phabet (both lower/upper case), and numbers 0-9.

This procedure resulted in a set of $1,161,265$ distinct usernames, $50,217$ of which belonged to users with private profiles. For each of the remaining $1,111,048$ public profiles, we collected the full set of posted vents and their directed social links (lists of followed/following users). Social links between private profile users were not accessible.

The obtained dataset is *complete*, in a sense it contains every vent posted and every social link (in any direction) of a user with a public profile since the genesis of Vent (oldest vent created at 17/12/2013) until 02/12/2018, the date at which we initiated the crawling procedure. Note that the crawling lasted approximately two weeks, meaning that the dataset also includes vents created in the meantime since the collection of data was carried out in batches.

Ethical considerations

The used methodology, if efficiently implemented, may collect a wide range of information, part of which could be sensitive. Nevertheless, by using the service, as stated in terms of service, one grants all users to view the shared content for their personal, non-commercial purposes. In this regard, all content for users of Vent is "public".

Despite the public nature of the shared data, we consider that data about individuals must be published only in an anonymized and/or aggregated form. Therefore, all direct identifiers of users have been removed, as well as shared URLs and usernames. In addition, all unique identifiers have been masked to prevent user linking. Finally, rather than publishing all the collected information, since this is a complete dataset, we have opted to remove the text from the public dataset to further guarantee user privacy as in (Garimella and Tyson 2018). Researchers who want to access the texts for research and non-commercial use are welcome to contact us.

Structure of the dataset

The crawled dataset contains $33,623,414$ vents in total, posted by $934,095$ users. These vents are annotated with a total of 705 emotions organized in 63 "emotion categories", concretely forming a two-level taxonomy of affects.

The provided dataset is structured in files each containing a different entity (i.e. emotion categories, emotions, vents and social links). Entities external to each file are cross-referenced via the anonymized universally unique identifiers (UUIDs).

Our full dataset consists of the following files and data:
- **emotion_categories.csv**:
  - id (string): a UUID associated with a specific emotion category.
  - name (string): The name of the emotion category.
- **emotions.csv**:
  - id (string): a UUID associated with a specific emotion.
  - emotion_category_id (string): a UUID associated with the corresponding emotion category.
  - name (string): The name of the emotion.
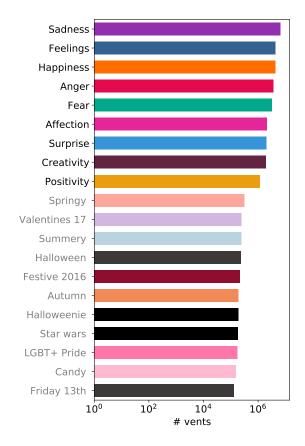 - enabled (boolean): Whether the emotion was enabled at the time of crawling.

Figure 2: The twenty most used emotion categories (labels in grey indicates non-permanent categories). Colors for each category correspond to the ones provided in Vent app.

- **vents.csv**:
  - emotion_id (string): a UUID associated with a specific emotion (cross-reference to emotions.csv).
  - user_id (string): a UUID associated with a specific user.
  - created_at (string): Date when the vent was posted, in UTC. Provided as a string in the format of "YYYY-MM-DD hh:mm:ss.sss".
  - reactions (integer): Total number of reactions to a vent.
  - text (string): The raw textual content of each vent. To preserve the anonymity of our dataset and at the same time reduce noise, we replace user mentions and URLs found in vents, following an approach similar to the one described in (Joshi and Deshpande 2018). We used a set of regular expressions to replace all URLs with the _URL_ token, and references to usernames with the _USER_REFERENCE_ token. No other text processing tasks were performed. This file is available only upon request as a restricted-access dataset [2]. Instead, in the publicly available dataset (Lykousas et al. 2019), we include a file named *vents_metadata.csv* which contains all the fields of vents.csv, except *text*.

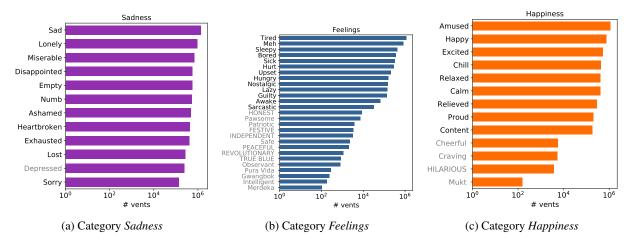[2]http://doi.org/10.5281/zenodo.2537982

Figure 3: Histograms of vents per emotion distributed per emotion categories. Disabled emotions are in gray.

- **vent.edges**: A snapshot of the social graph of Vent, at the time of crawling. It contains the directed friendship links between users. The UUIDs of the users (nodes) are the same as in the file **vents.csv**.
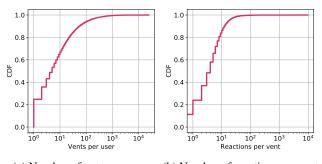
Data Exploration

We now present an exploratory analysis of the Vent dataset. We first look at the usage of the different emotions at the level of vent, emotions, and users. We then briefly look at the temporal evolution of vents, the social network and outline a possible way to construct networks of emotions. Finally, we analyze the text of the vents and describe the affective landscape of Vent and relate it with an existing emotion lexicon.
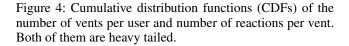
Emotion Categories and Their Usage

The main emotion categories in Vent are *Fear*, *Surprise*, *Feelings*, *Sadness*, *Anger*, *Creativity*, *Affection*, *Happiness*, and *Positivity*, some of which match Plutchik's primary emotions (Plutchik 1991). These emotion categories are always available. In contrast, there is an additional set of emotion categories that are not always enabled (non-permanent ones) but can become active as in-app purchases, unlocked during a specific season (e.g. Spring, Autumn) or event/festivity (e.g. Women's Day, Hanukkah). Other emotion categories can be disabled/deprecated. We also identified several (normally disabled) emotions with the same name per category, that can only be differentiated through the provided unique identifiers.

Figure 2 shows the distribution of vents across the different emotion categories. For clarity, we show the twenty most used categories from the total of sixty-three categories included in the dataset. The top ones correspond to the main ones, being *Sadness* at the top with more than 3M vents, followed by *Feelings* and *Happiness*, with approximately 1.5*M* of vents. The non-permanent ones (with grey labels) were used less frequently and are grouped at the end of the list. Despite being less popular, some of the non-permanent ones were extensively used in many vents, e.g., *Springy* or *Valentines17*, which were associated to more than 100*K* vents.

Figure 3 shows in more detail the distribution of vents across the different emotions for the top three categories. We observe a similar pattern regarding enabled/disabled emotions like the one with permanent/non-permanent categories. The category *Feelings*, despite being the one containing the largest number of different types of emotions, received many millions of vents less than *Sadness*.
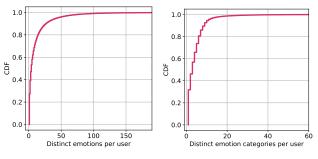

(a) Number of vents per user   (b) Number of reactions per vent

Figure 4: Cumulative distribution functions (CDFs) of the number of vents per user and number of reactions per vent. Both of them are heavy tailed.

We now examine different aspects of user behavior in the Vent social platform. Figure 4 shows the cumulative distribution function (CDF) of the number of vents per user and the number of reactions per vent. Both of them are governed by heavy-tailed distributions that span four orders of magnitude, indicating that a vast majority of 60% of the users posted less than 10 vents while a small group of few users posted more than 10^4 vents. A similar pattern holds for the number of reactions per vent.

To determine the extent to which venters use the wide range of emotions offered by the Vent platform, we plot the cumulative distribution functions (CDFs) in Figure 5, for the number of distinct emotions and number of distinct emotion categories associated with the vents of each user. We observe that the majority of users post vents within a limited

(a) Distinct emotions per user

(b) Distinct emotion categories per user

Figure 5: Cumulative distribution functions (CDFs) of the emotions and emotion categories per user. Users typically focus on a set of few emotions and emotion categories, but a few users can use a large number of them.

set of emotions and emotion categories, with 50% of users only using up to five different emotions and three emotion categories. A few users, however, can use up to 60 emotion categories and more than 200 emotions.

Temporal activity

Figure 6 shows the number of vents posted per month, according to the field `created_at` in each vent. The activity shows an increase in the number of vents from December 2013 (when the Vent app was launched) until a peak of activity was reached around April 2015. That maximum of activity comprised more than one millions of vents during that month. Since then, the activity has generally been sustained, but slowly decreasing during the last two years.
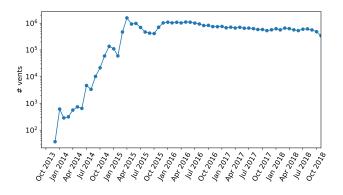

Figure 6: Aggregated monthly activity shows an increase that reached nearly a million of vents .

The social network of Vent

Vent users can form social links to other users in the platform. The resulting social network of Vent users has approximately a million of nodes (users) and contains around 13.5 millions of edges (directional links between them). Table 2 shows some additional global indicators of the network.

Table 2: Main network indicators of the Vent social graph: number of nodes N, number of edges E, average degree $\langle k \rangle$, density D, and reciprocity ρ.

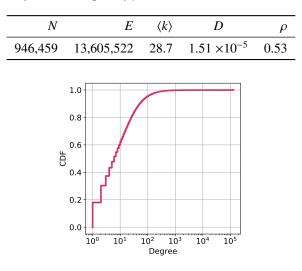| N	E	$\langle k \rangle$	D	ρ
946,459	13,605,522	28.7	1.51×10^{-5}	0.53

Figure 7: Degree distribution of the social network in Vent.

Figure 7 shows the degree distribution of the social network. We observe the typical heavy tail behavior, with a vast majority of venters linked to a small number of other users, and a small minority of venters having more than $10K$ links.

Text properties of vents

We proceed to examine some of the textual properties of the vents. Table 1 shows an illustrative sample of vents and their associated emotions. Figure 8 shows the distribution of the vents length, which have a well defined typical length, suggesting a log-normal behavior. On average, vents are 163.27 characters long (with a large standard deviation of 388.81).

In this preliminary analysis, we use an off-the-shelf language identification tool (Lui and Baldwin 2012) to infer the language of each vent. As highlighted by the authors, the short size and language novelties of text produced in the context of this type of data have a considerable impact on the performance of language recognition. Nonetheless, their model identified 93% of the vents as English.

We also generated word clouds from the vents associated with two contrasting groups of emotions: one with emotions belonging to *negative* categories such as *Sadness*, *Anger* and *Fear*, and one with *positive* emotions from *Happiness*, *Affection* and *Positivity* categories. Figure 9 shows these two word clouds generated using `word_cloud`[3].

Interestingly, one of the predominant words in positive vents appears to be the term **NSFW**. An explanation from this can be found in the community guidelines of the Vent app[4], stating that:

"*Vent posts that contain sexually explicit content must be flagged with the text "NSFW" in the body of the*

---

[3]https://github.com/amueller/word_cloud
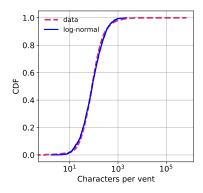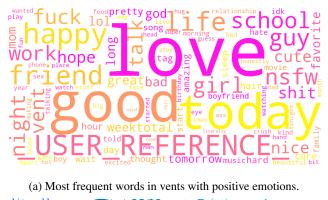[4]https://www.vent.co/cg/

Figure 8: The Vent length (in characters) probability distribution is well approximated by a lognormal ($\mu = 85.57$, $\sigma = 1.06$).

(a) Most frequent words in vents with positive emotions.

(b) Most frequent words in vents with negative emotions.

Figure 9: Word clouds built from the text of vents belonging to different (positive and negative) categories.

Vent. NSFW stands for Not Safe For Work and is a common way of describing content that is sexually explicit in nature.".

In total, we identified 508,545 vents in our dataset flagged as explicit.

The network of emotions

Our dataset can also be used to analyze the relations between emotions. For example, one can build a network using the following procedure: let $v(u, e_k)$ the (normalized per user) number of events of user annotated with emotion e_k. We define a pair of emotions e_i and e_j to be related with respect
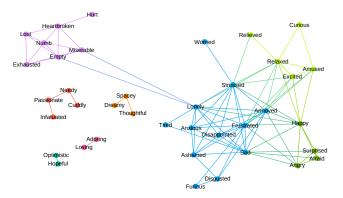

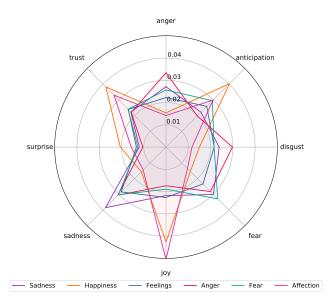Figure 10: A example emotion network ($T_1 = 0.1$ and $T_2 = 0.05$). See text for details.

to their common appearance in a particular user's vents, if $v(u, e_i)$ and $v(u, e_j)$ exceed the value of an arbitrary threshold T_1 (criterion 1). Moreover, we can assign a weight to an edge $e_i \sim e_j$ consisting of the number of users satisfying criterion 1 for both emotions, and normalized as the Jaccard similarity between the sets of users satisfying criterion 1 for each emotion. Additionally, for the sake of visualization, we can control the density of such a graph by filtering out edges with weight below a second threshold T_2. This way the behavior of vent users and, by extension, the affective mental states they express in Vent platform can contribute to the formation of links between different emotions.
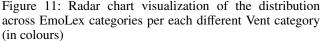
Figure 10 shows the emotion network resulting from applying the steps described above. Different node colors denote the communities returned by applying the standard Louvain method (Blondel et al. 2008). It is noteworthy that similar emotions are connected, forming distinct clusters, without necessarily belonging to the same emotion category, e.g., *Heartbroken* and *Hurt*. Moreover, we observe the existence of small connected components, meaning that there exist users that mainly vent about specific moods (such as *Thoughtful*, *Needy*, etc.), outside from the typical spectrum of affect.

From this analysis, one can get some initial insight into the similarity and co-occurrence between different emotions. Our proposed approach can be extended and enriched with the provided temporal information to shed light on complex and largely unexplored affective mechanisms such as the emotion transitions (Thornton and Tamir 2017).

The "affective landscape" of Vent

We now analyze the text of the vents and explore the "affective landscape" of the different emotion categories. To this end, we use the NRC Emotion Lexicon "EmoLex" (Mohammad, Kiritchenko, and Zhu 2013). The EmoLex dictionary contains 14,182 words crowd-labelled according to the eight Plutchik's primary emotions (Plutchik 1991): "sadness", "joy", "disgust", "anger", "fear", "surprise", "trust", and "anticipation". Additionally, it includes positive and

Figure 11: Radar chart visualization of the distribution across EmoLex categories per each different Vent category (in colours)

negative valence categorizations for every included term.

We sampled uniformly at random a set of 3M vents (approximately 10% of the dataset) and associated each vent to a set of scores per each EmoLex category in the following way: we count the number of words belonging to each EmoLex category, normalize them by the total number of words in the vent, and consider the category-wise mean value. Notably, there are 26% of the sampled vents for which none of their words was found in EmoLex, suggesting a potential opportunity to extend existing affective lexica such as EmoLex. This experiment was repeated multiple times with different random samples of vents and the results that we describe next were consistent across the runs.

Figure 11 shows how the top six Vent's emotion categories distribute across each of the EmoLex's emotion categories. We observe that vents belonging to the categories of *Happiness* and *Affection* exhibit high levels of "joy", "trust" and "anticipation" while vents of all other categories are dominated by mostly negative EmoLex emotions ("anger", "disgust", "fear", sadness"). Interestingly, EmoLex words from the categories "anger", "fear" and "sadness" are over-represented in their corresponding Vent categories, as shown in the radar chart visualization. This agreement confirms the alignment and coverage to a large extent of the Vent categories with the ones of the EmoLex annotated corpus.

Finally, we also analyze the distribution of valence scores (according to EmoLex) for different Vent categories. We found that the "positive" categories of *Happiness* and *Affection* are clearly differentiated from the rest of the categories, which mostly contain negative or neutral emotions. Figures 12a and 12b show the CDFs of positive/negative valence scores for the same Vent categories aggregated in two groups and including only nonzero values for readability.

These results agree with the previous findings, suggesting that there exist substantial differences in emotions expressed in vents across different categories of our dataset. Also, they highlight certain regularities, despite the large heterogeneity found in the dataset.

Recommendations for Future Work

In the above section, we have presented a preliminary exploration of our dataset of vents voicing the emotions of approximately 1M users. Given the volume and the content diversity of our dataset, in place of a conclusion, we outline some directions for potential future research.

Emotion analysis: Emotion analysis represents a natural evolution of sentiment analysis. Modeling emotions expressed in text beyond the basic polarity classifications/scales used in sentiment analysis is a challenging task since emotions not only depend on the semantics of a language but are also inherently subjective and ambiguous (Sudhof et al. 2014). Many researchers argue that accounting for affects is crucial in approximating real-world true natural language understanding, especially in areas involving human-computer interactions (Park, Xu, and Fung 2018; Fung 2015).

Recent work has demonstrated that artificial neural networks have great potential in tasks such as emotion recognition (Baziotis et al. 2018; Baziotis, Pelekis, and Doulkeridis 2017). This can be attributed to their ability to learn features directly from data in addition to using hand-crafted features where necessary, thus outperforming conventional approaches which require extensive feature engineering from experts. Such methods, due to their dependence on emotion lexica and hand-crafted features, cannot keep up with rapid language evolution (Mudinas, Zhang, and Levene 2012), especially in social media/micro-blogging context. To this end, we hope the Vent dataset will contribute towards the advancement of emotion analysis in text, by enabling the development and evaluation of novel, neural-network-based models, capable of naturally exploiting its volume and diversity.

Interplay between network structure and emotional behavior: Another line of research exploiting the social graph data of Vent could be the study of social relationships of users with respect to their emotional profiles, and vice-versa. Previous studies in a variety of social media contexts such as Wikipedia (Iosub et al. 2014), chats (Singla and Richardson 2008) and blogs (Thelwall 2010), have shown evidence that users tend to interact and associate with others expressing similar emotions, referred as the phenomenon of "emotional homophily". It would be interesting to investigate if relationships among Vent users with respect to the emotions they express adhere to the principle "birds of a feather flock together".

Furthermore, it has been observed that emotions can be diffused through social networks, across the links connecting individuals. Although the phenomenon of emotional contagion is well established in laboratory experiments and real-world social networks (Fowler and Christakis 2008), in the context of online social networks, it is largely unexplored. Nonetheless, a highly controversial Facebook study
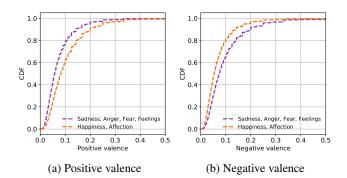
(a) Positive valence   (b) Negative valence

Figure 12: Cumulative distribution function of the valences grouped by Vent emotion category. The positive valences have more probability mass for Vent's categories *Happiness* and *Affection*, coinciding with the "positive" Vent emotions (CDF is clearly on the right side). The opposite holds if we consider negative valences.

(Kramer, Guillory, and Hancock 2014) has provided evidence that emotional states can be transferred to others via emotional contagion, leading people to experience the same emotions without their awareness. Additionally, research in Flickr (Yang et al. 2016) has demonstrated that factors such as the social role of individuals are significant with respect to the extent they influence their social connections. We believe the release of the Vent dataset, an example of both network structure (the social network) and diffusion (the vent activity) will stimulate research leading to better understanding the underlying mechanism of emotional contagion in online social networks.

Acknowledgements

References

Agarwal, A.; Xie, B.; Vovsha, I.; Rambow, O.; and Passonneau, R. 2011. Sentiment analysis of twitter data. In *Proceedings of the workshop on languages in social media*, 30–38. Association for Computational Linguistics.

Ahkter, J. K., and Soria, S. 2010. Sentiment analysis: Facebook status messages. *Unpublished master's thesis, Stanford, CA*.

Barrett, L. F. 2006. Solving the emotion paradox: Categorization and the experience of emotion. *Personality and social psychology review* 10(1):20–46.

Bazarova, N. N.; Choi, Y. H.; Schwanda Sosik, V.; Cosley, D.; and Whitlock, J. 2015. Social sharing of emotions on facebook: Channel differences, satisfaction, and replies. In *Proceedings of the 18th ACM conference on computer supported cooperative work & social computing*, 154–164. ACM.

Baziotis, C.; Nikolaos, A.; Chronopoulou, A.; Kolovou, A.; Paraskevopoulos, G.; Ellinas, N.; Narayanan, S.; and Potamianos, A. 2018. Ntua-slp at semeval-2018 task 1: Predicting affective content in tweets with deep attentive rnns and transfer learning. In *Proceedings of The 12th International Workshop on Semantic Evaluation*, 245–255. Association for Computational Linguistics.

Baziotis, C.; Pelekis, N.; and Doulkeridis, C. 2017. Datastories at semeval-2017 task 4: Deep lstm with attention for message-level and topic-based sentiment analysis. In *Proceedings of the 11th International Workshop on Semantic Evaluation (SemEval-2017)*, 747–754.

Berger, J., and Milkman, K. L. 2012. What makes online content viral? *Journal of marketing research* 49(2):192–205.

Blondel, V. D.; Guillaume, J.-L.; Lambiotte, R.; and Lefebvre, E. 2008. Fast unfolding of communities in large networks. *Journal of statistical mechanics: theory and experiment* 2008(10):P10008.

Cambria, E.; Schuller, B.; Xia, Y.; and Havasi, C. 2013. New avenues in opinion mining and sentiment analysis. *IEEE Intelligent Systems* 28(2):15–21.

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

Fung, P. 2015. Robots with heart. *Scientific American* 313(5):60–63.

Garcia, D.; Kappas, A.; Küster, D.; and Schweitzer, F. 2016. The dynamics of emotions in online interaction. *Royal Society open science* 3(8):160059.

Garimella, V. R. K., and Tyson, G. 2018. Whatsapp, doc? a first look at whatsapp public group data. In *ICWSM*.

Iosub, D.; Laniado, D.; Castillo, C.; Morell, M. F.; and Kaltenbrunner, A. 2014. Emotions under discussion: Gender, status and communication in online collaboration. *PloS one* 9(8):e104880.

Joshi, S., and Deshpande, D. 2018. Twitter sentiment analysis system. *International Journal of Computer Applications* 180(47):35–39.

Kao, E. C.-C.; Liu, C.-C.; Yang, T.-H.; Hsieh, C.-T.; and Soo, V.-W. 2009. Towards text-based emotion detection a survey and possible improvements. In *Information Management and Engineering, 2009. ICIME'09. International Conference on*, 70–74. IEEE.

Kouloumpis, E.; Wilson, T.; and Moore, J. D. 2011. Twitter sentiment analysis: The good the bad and the omg! *Icwsm* 11(538-541):164.

Kramer, A. D.; Guillory, J. E.; and Hancock, J. T. 2014. Experimental evidence of massive-scale emotional contagion through social networks. *Proceedings of the National Academy of Sciences* 201320040.

Liu, B., and Zhang, L. 2012. A survey of opinion mining and sentiment analysis. In *Mining text data*. Springer. 415–463.

Lui, M., and Baldwin, T. 2012. langid. py: An off-the-shelf language identification tool. In *Proceedings of the ACL 2012 system demonstrations*, 25–30. Association for Computational Linguistics.

Lykousas, N.; Patsakis, C.; Kaltenbrunner, A.; and Gómez, V. 2019. Dataset for paper "Sharing emotions at scale: The Vent dataset". https://doi.org/10.5281/zenodo.2537838.

Lykousas, N.; Patsakis, C.; and Gómez, V. 2018. Adult content in social live streaming services: Characterizing deviant users and relationships. In *2018 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM)*, 375–382. IEEE.

Mohammad, S.; Bravo-Marquez, F.; Salameh, M.; and Kiritchenko, S. 2018. Semeval-2018 task 1: Affect in tweets. In *Proceedings of The 12th International Workshop on Semantic Evaluation*, 1–17.

Mohammad, S. M.; Kiritchenko, S.; and Zhu, X. 2013. Nrc-canada: Building the state-of-the-art in sentiment analysis of tweets. *arXiv preprint arXiv:1308.6242*.

Mudinas, A.; Zhang, D.; and Levene, M. 2012. Combining lexicon and learning based approaches for concept-level sentiment analysis. In *Proceedings of the first international workshop on issues of sentiment discovery and opinion mining*, 5. ACM.

Ortigosa, A.; Martín, J. M.; and Carro, R. M. 2014. Sentiment analysis in facebook and its application to e-learning. *Computers in human behavior* 31:527–541.

Pak, A., and Paroubek, P. 2010. Twitter as a corpus for sentiment analysis and opinion mining. In *European Language Resources Association*, volume 10, 1320–1326.

Pantic, M.; Sebe, N.; Cohn, J. F.; and Huang, T. 2005. Affective multimodal human-computer interaction. In *Proceedings of the 13th annual ACM international conference on Multimedia*, 669–676. ACM.

Park, J. H.; Xu, P.; and Fung, P. 2018. Plusemo2vec at semeval-2018 task 1: Exploiting emotion knowledge from emoji and #hashtags. In *Proceedings of The 12th International Workshop on Semantic Evaluation*, 264–272. Association for Computational Linguistics.

Plutchik, R. 1991. *The emotions*. University Press of America.

Politou, E.; Alepis, E.; and Patsakis, C. 2017. A survey on mobile affective computing. *Computer Science Review* 25:79–100.

Rime, B.; Mesquita, B.; Boca, S.; and Philippot, P. 1991. Beyond the emotional event: Six studies on the social sharing of emotion. *Cognition & Emotion* 5(5-6):435–465.

Siekkinen, M.; Masala, E.; and Kämäräinen, T. 2016. A First Look at Quality of Mobile Live Streaming Experience. In *Proceedings of the 2016 ACM on Internet Measurement Conference*, 477–483. ACM.

Singla, P., and Richardson, M. 2008. Yes, there is a correlation:-from social networks to personal behavior on the web. In *Proceedings of the 17th international conference on World Wide Web*, 655–664. ACM.

Sudhof, M.; Goméz Emilsson, A.; Maas, A. L.; and Potts, C. 2014. Sentiment expression conditioned by affective transitions and social forces. In *Proceedings of the 20th ACM SIGKDD international conference on Knowledge discovery and data mining*, 1136–1145. ACM.

Thelwall, M. 2010. Emotion homophily in social network site messages. *First Monday* 15(4).

Thornton, M. A., and Tamir, D. I. 2017. Mental models accurately predict emotion transitions. *Proceedings of the National Academy of Sciences* 201616056.

Troussas, C.; Virvou, M.; Espinosa, K. J.; Llaguno, K.; and Caro, J. 2013. Sentiment analysis of facebook statuses using naive bayes classifier for language learning. In *Information, Intelligence, Systems and Applications (IISA), 2013 Fourth International Conference on*, 1–6. IEEE.

Wang, Y., and Pal, A. 2015. Detecting emotions in social media: A constrained optimization approach. In *IJCAI*, 996–1002.

Wang, H.; Can, D.; Kazemzadeh, A.; Bar, F.; and Narayanan, S. 2012. A system for real-time twitter sentiment analysis of 2012 us presidential election cycle. In *Proceedings of the ACL 2012 System Demonstrations*, 115–120. Association for Computational Linguistics.

Xu, A.; Liu, Z.; Guo, Y.; Sinha, V.; and Akkiraju, R. 2017. A new chatbot for customer service on social media. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*, CHI '17, 3506–3510. New York, NY, USA: ACM.

Yang, Y.; Jia, J.; Wu, B.; and Tang, J. 2016. Social role-aware emotion contagion in image social networks. In *AAAI*, 65–71.