


RESEARCH

Open Access



# Sentiment and structure in word co-occurrence networks on Twitter

Mikaela Irene Fudolig<sup>1\*</sup> , Thayer Alshaabi<sup>1,2</sup>, Michael V. Arnold<sup>1</sup>, Christopher M. Danforth<sup>1,3</sup> and Peter Sheridan Dodds<sup>1,4</sup>

\*Correspondence:

mikaela.fudolig@uvm.edu

<sup>1</sup> Computational Story Lab, Vermont Complex Systems Center, MassMutual Center of Excellence for Complex Systems and Data Science, University of Vermont, Burlington, VT, USA

Full list of author information is available at the end of the article

## Abstract

We explore the relationship between context and happiness scores in political tweets using word co-occurrence networks, where nodes in the network are the words, and the weight of an edge is the number of tweets in the corpus for which the two connected words co-occur. In particular, we consider tweets with hashtags #imwithher and #crookedhillary, both relating to Hillary Clinton's presidential bid in 2016. We then analyze the network properties in conjunction with the word scores by comparing with null models to separate the effects of the network structure and the score distribution. Neutral words are found to be dominant and most words, regardless of polarity, tend to co-occur with neutral words. We do not observe any score homophily among positive and negative words. However, when we perform network backboning, community detection results in word groupings with meaningful narratives, and the happiness scores of the words in each group correspond to its respective theme. Thus, although we observe no clear relationship between happiness scores and co-occurrence at the node or edge level, a community-centric approach can isolate themes of competing sentiments in a corpus.

**Keywords:** Sentiment analysis, Word co-occurrence, Co-occurrence networks

## Introduction

Large-scale analysis of user-generated text has been instrumental in understanding recent political campaigns and movements, from the Arab Spring (Howard et al. 2011; Wolfsfeld et al. 2013) to the Black Lives Matter protests (Wu et al. 2021). Complementing traditional survey based approaches, social media has also shown promise as a medium for gauging public sentiment. Unlike surveys, however, where questions are carefully constructed so that the answers can be directly interpreted, extracting sentiment from large-scale unstructured text requires automated interpretation. The field of sentiment analysis in natural language processing (NLP) developed in response to this need, not just in politics but in other areas as well. Retailers have used sentiment analysis techniques to analyze product reviews and provide a better experience for its users (Fang and Zhan 2015; Shivaprasad and Shetty 2017). Sentiment analysis has also been applied to financial markets, where the opinion of market participants plays an important role in future prices (Smailovic et al. 2013; Pagolu et al. 2016; Mishev et al. 2020).

Understanding public opinion also matters in public health, as in the case of addressing perceptions of vaccines (Raghupathi et al. 2020; Klimiuk et al. 2021), and predicting signals of depression in self-expressed social media posts (Wang et al. 2013; Coppersmith et al. 2014; Reece et al. 2017; Stupinski et al. 2021).

Studies on sentiment in social media can be broadly divided into two groups: one where the emphasis is on identifying the sentiment of a particular statement, and another that focuses on obtaining a collective measure of sentiment. In this work, we deal with the latter, where we are interested in the general sentiment associated with a particular topic rather than the emotional state expressed by a particular individual. This paradigm of studying collective sentiment has been used to monitor general sentiment on Twitter, and has been validated by cross-referencing changes in measured sentiment with high-profile events, such as election results, terrorist activities, holidays and even birthdays of popular artists (Dodds et al. 2011). For example, holidays such as Christmas, New Year, Fourth of July and Thanksgiving correspond to spikes in happiness, while mass shootings in the United States, the fire at Notre Dame cathedral, the murder of George Floyd and the storming of the US Capitol coincide with a marked decrease in happiness on Twitter. In contrast to this approach, which looks at social media posts as a bag of words, we look at it as a collection of tweets that are in turn related by their use of specific words. In particular, do certain words go together, and if so, what is the sentiment profile of these groups of words, and what do they say about the sentiment profile of the tweets themselves? We want to know if this approach will provide a more nuanced picture of collective sentiment that is missed by a simple aggregation.

One way to infer collective sentiment or meaning from text is through lexicons, where a subset of words is compiled, either through expert curation or by frequency of use, to provide a representative sample of the corpus. Sentiment scores are then assigned to each word, usually by human annotators. One such lexicon is the labMT dataset used in the Hedonometer (Dodds et al. 2011, 2015), which contains happiness scores from human annotators for more than 10,000 words. The word list was later expanded (Alshaabi et al. 2022) to extrapolate happiness scores for any unscored word using a word embedding model. Aside from happiness scores, semantic differentials have also been used to quantitatively capture meanings of words. The valence-arousal-dominance (VAD) semantic differentials have been commonly used since its inception, and a number of studies have focused on creating large-scale VAD lexicons (Bradley and Lang 1999; Warriner et al. 2013; Mohammad 2018). Other examples of sentiment lexicons include SentiWordNet (Baccianella et al. 2010) and LIWC (Pennebaker et al. 2015).

In contrast to lexicon scores, co-occurrence models infer the meaning of a word using the surrounding context, i.e., the adjacent words that occur proximate to the anchor word in real-world corpora. These models output word embeddings, which are high-dimensional vectors that represent a word's meaning. Examples of such models are Word2Vec (Mikolov et al. 2013), GloVe (Pennington et al. 2014), and FastText (Bojanowski et al. 2017; Joulin et al. 2017). Although these word embeddings perform well in complex tasks such as machine translation and language detection, the meaning is abstracted in several dimensions and is thus less interpretable than the scores provided by lexicons. A study by Hollis and Westbury (2016) examined whether the 300 dimensions of Word2Vec corresponds to any of the well-known semantic and lexical variables, including

valence, arousal and dominance. However, they found that no single dimension corresponded to a particular semantic or lexical variable, although the principal components that contributed the most to the variance were found to be correlated to some semantic and lexical variables. This is particularly interesting as Word2Vec was not developed using these semantic and lexical concepts, but nevertheless captures aspects of them in the embedding.

Conversely, context is not used to obtain lexicon scores; human annotators are presented with words and are asked to score them in the desired scale. The interplay between sentiment scores and co-occurrence patterns in words has mostly been viewed from a predictive standpoint, i.e., whether sentiment scores can be predicted using word co-occurrence. Early work predicted sentiment polarity (positive or negative) (Turney and Littman 2002, 2003), while later models predicted valence, arousal, and dominance using latent semantic analysis (Bestgen and Vincze 2012), pointwise mutual information (Recchia and Louwerse 2015) and co-occurrence models such as HiDEx (Shaoul and Westbury 2006; Westbury et al. 2015) and Word2Vec (Hollis et al. 2017). Recently, Alshaabi et al. (2022) used FastText word embeddings and the transformer model DistilBERT (Sanh et al. 2019), both pre-trained on large-scale, general corpora, to predict happiness scores for out-of-vocabulary words. While prediction success implies the existence of a relationship between word co-occurrence and lexicon scores, we are not aware of any study that delves deeper into the details of this relationship. In particular, given a corpus, how are the sentiment scores of co-occurring words related? And can the co-occurrence structure of words be used to uncover mixed sentiments in a corpus that would be missed by an aggregate measure, such as the average happiness score?

We explore these questions in the context of tweets, which because of character count limitations (140 before 2018, and 280 thereafter) generally contain words that are related by a single theme. Rather than using techniques resulting in word embeddings that abstract words into non-interpretable vectors, we retain the nature of the word itself and instead use tools from network science to analyze the word co-occurrence structure in the corpus. Although NLP has leaned more towards probabilistic models and neural networks, network science has nevertheless found a wide range of applications in computational linguistics (Ferrer i Cancho et al. 2004; Liu and Cong 2013; Cong and Liu 2014; Al Rozz et al. 2017; Wang et al. 2017; Chen et al. 2018; Jiang et al. 2019). In fact, some techniques commonly used in NLP, such as latent Dirichlet allocation, have been shown to parallel network science techniques (Gerlach et al. 2018). By constructing word co-occurrence networks, where the words are nodes in a network, the sentiment scores are node properties, and the edges in the network represent the co-occurrence structure, we can see more directly how co-occurrence and sentiment are related.

Specifically, each node in the network represents a word found in a tweet. As a tweet contains words that are related to each other, each tweet corresponds to a connected subgraph, where each word appearing in the tweet is connected to every other word appearing in the same tweet. We superpose the subgraphs for each tweet in the corpus to form the full word co-occurrence network, with the edge weights corresponding to the number of tweets in which two words co-occur together. The degree and strength of a node tell us how often a word co-occurs with others, while the weight of an edge

indicates how often any two words co-occur. This network generation method has also been used in prior work that studied word co-occurrence in tweets (Garg and Kumar 2018).

We then analyze the context-dependent network structure and its relation to the context-independent word lexicon scores, in particular the happiness scores from both the labMT dataset and the predictive model in Alshaabi et al. (2022). We look at the network from three perspectives: node-centric, edge-centric, and community-centric. In the node-centric perspective, we examine if there is a relationship between node characteristics, such as degree and node strength, and the happiness scores. In the edge-centric perspective, we consider pairs of words and their respective scores. Lastly, in the community-centric approach, we use network backboning and community detection techniques to obtain groups of related words within the network, and explore the relationship of scores within each group to scores across groups.

As we are interested in whether we can uncover mixed sentiments in corpora, we focus on political tweets, where polarization in stance is common. Stance is distinct from sentiment, in that it reflects the opinion (“favor”, “against”, or “neither”) towards a target (Küçük and Can 2020). Regardless, sentiment has been shown to be important, though not sufficient, to predict stance (Mohammad et al. 2016). Thus, by combining tweets using hashtags that are from opposite stances, we are likely to get tweets with opposing sentiments. Further, within each stance, we will also likely encounter a minority of tweets with a sentiment opposite to that expected, e.g., an “against” stance tweet may have a positive sentiment towards the target’s opponent. We also examine whether these cases will be picked up by our analysis. As the political situations associated with these tweets are likely to be reported in the news, we can use domain knowledge to check the validity of our results.

We detail the selection of tweets and the method of our analysis in the “[Methodology](#)” section. In the “[Results](#)” section, we present our findings regarding the relationship between the co-occurrence network structure and happiness scores at the node-centric, edge-centric, and community-centric levels. We summarize our results in the “[Discussion and summary](#)” section and present avenues for future work.

## **Methodology**

### **The tweets**

For our analysis, we consider a collection of tweets relating to Hillary Clinton’s presidential bid, in particular, tweets with any of the following anchors, #imwithher or #crookedhillary (case-insensitive). Both anchors were mostly used in 2016 during the US presidential election season, where #imwithher is associated with being in favor of Hillary Clinton and #crookedhillary associated with being against her. We use Twitter’s Decahose API, which provides a random 10% of all the tweets for a particular period. From this sample, we get tweets matching the above hashtags during the 1-week period of election season when the anchors were most popular, respectively. These periods were November 1-8 for #imwithher and October 7-14 for #crookedhillary, both in 2016 (Alshaabi et al. 2021). We acknowledge that although these stance association assumptions are general, they are not universal, and each hashtag may

be used for the opposite stance. We also investigate whether such cases can be uncovered using our analysis.

We then sampled the tweets so that we get balanced classes, with each hashtag getting 7222 unique original tweets (no retweets, quote tweets, or duplicates). We also note that we found an obvious case of systematic hashtag hijacking with content unrelated to Hillary Clinton and removed the corresponding tweets before sampling (see the Additional file 1 for details). The tweets containing #imwithher are labeled as “favor” tweets, while the ones containing #crookedhillary are labeled as “against” tweets. These labels correspond to the general stance associated with each anchor hashtag. The aggregated set of tweets is denoted by the label “all”.

### Happiness scores

For the word ratings, we consider the Language Assessment by Mechanical Turk (labMT) dataset that contains scores for happiness (Dodds et al. 2011, 2015) for more than 10,000 words. The list was compiled from the most frequently appearing words in more contemporary sources such as Twitter, Google Books, the New York Times and music lyrics. Human annotators were asked to score a word using a scale of 1–9, with 1 being the least happy and 9 being the happiest, and 5 being neutral. The means of the measurements as well as their standard deviations are reported in the dataset.

To get the happiness scores for words not present in the labMT dataset, we use deep learning models trained on the labMT dataset to extrapolate happiness scores for words and *n*-grams (Alshaabi et al. 2022). The first proposed model, namely the token model, breaks up the input into character-level *n*-grams and uses subword embeddings to estimate their happiness, while the second model uses dictionary definitions to gauge happiness scores. The predictive scoring models have been found to be reliable, making predictions well within the range of scores that are obtained from different human annotators. Although the dictionary model provides better performance for words in the dictionary, the token model performs almost as well with lower computation cost and gives more reliable results for words that are not in dictionaries but are made up of component subwords, such as hashtags. We use the token model to infer the happiness scores for the words in the network that are not in the original labMT dataset, allowing us to compare the effect of the choice of words to evaluate the network structure. To further improve accuracy, we use the expanded form of the most commonly used acronyms in the corpora examined (Table 1) as an input to the predictive scoring model to obtain their respective scores.

**Table 1** Acronyms and their full versions

Acronym in tweet	Model input
maga	Makeamericagreatagain
msm	Mainstreammedia
tcot	Topconservativesontwitter
potus	Presidentoftheunitedstates

Acronyms in the corpus with high word counts are scored using their full versions in the predictive scoring model using tokens Alshaabi et al. (2022). The model breaks the input token into character-level *n*-grams, so there is no need to add whitespaces here

### Generating the co-occurrence network

We parse each tweet as follows. As the writing style on Twitter involves a liberal use of hashtags, we disregard the hashtag symbol in each tweet. Each tweet is then converted into lower case and parsed by splitting each string into a set of words separated by whitespaces, while preserving contractions, punctuation, handles, hashtags, dates and links (see Alshaabi et al. 2021 for more information). For simplicity, we further exclude punctuation marks, and undo contractions, as shown in Table 2. Twitter handles (1-grams beginning with a “@”), 1-grams with numbers, and URL links were also removed, as well as the anchor terms #crookedhillary and #imwithher in order to eliminate community structure associated with the choice of these hashtags. In our preliminary analysis, we find that the names of Hillary Clinton and her opponent, Donald Trump, also effectively served as anchor words, masking finer structure in the network. We thus also remove the terms “hillary”, “clinton”, “hillaryclinton”, “trump”, “donald”, and “donaldtrump”.

We then obtain counts for the words used in the tweets (the document-term matrix) using the Python package scikit-learn (Pedregosa et al. 2011). In case a word appears multiple times inside a single tweet, it is only counted once in generating the network; however, the original word counts are also recorded. We finally obtain the adjacency matrix that is used to generate the undirected word co-occurrence network by taking the dot product of the document-term matrix (where each word is only counted once in a tweet) and its transpose, setting the diagonal elements to zero. In the resulting network, each node is a word occurring in the set of tweets. An edge between two nodes indicates that the two nodes occur together in some tweets, and the edge weights tells us how many tweets the two connected nodes occur together. We construct a separate network for each set of tweets (e.g., for the Hillary Clinton tweet corpus, we create a separate network for the “favor”, “against”, and “all” tweets).

### Network characterization

We look at the basic properties of the network, such as the node strength, degree and edge weight distribution, and examine how these are related to the scores of the node. To determine if the results we observe are due to the happiness scores or to the network structure, we compare our results with a number of null models, where we modify either the network structure or the scores of the words. These include:

**Table 2** Contractions and equivalents

Contractions	Converted to	Remarks
n't	<space>not	
's	<space>	either “is” or possessive marker
'm	<space>	“am”; for consistency, since “s” is converted to a space
'd	<space>	either “had” or “would”
're	<space>	“are”; for consistency, since “s” is converted to a space
've	<space>have	
'll	<space>will	

Some contractions in tweets were converted to equivalent strings. In case of ambiguity, the contraction is converted to a blank space



*Configuration model* The network is randomly rewired using the node strength sequence to roughly preserve the node strength distribution; self-loops are discarded, and parallel unweighted edges connecting two nodes are combined to form a weighted edge (Serrano 2005).

*Erdos-Renyi model* An Erdos-Renyi network is constructed with the probability  $p$  of an edge occurring set to be  $E_{\text{obs}}/E_{\text{max}}$ , where  $E_{\text{obs}}$  is the number of observed (unweighted) edges and  $E_{\text{max}}$  is the total number of possible (unweighted) edges. The edge weights are assigned by drawing with replacement from the distribution of edge weights in the original graph.

*Shuffled score model* The network is kept as is, but the scores are reshuffled among the nodes.

*Uniform score model* The network is kept as is, but the scores of the nodes are sampled from a uniform distribution from 1 to 9 (the prescribed range of happiness scores in labMT).

### Community detection and network backboning

To explore whether clusters in the co-occurrence network are also related to sentiment scores, we perform community detection on the co-occurrence network. However, because the resulting co-occurrence networks have words that occur very frequently (such as the words “is” and “the”) and also many words that are used together in only a few tweets, we find that performing community detection on the full set of words resulted in communities that were dominated by function words and did not correspond to any discernible theme. We thus need to trim down the network to expose a more robust underlying structure. To do this, we use *network backboning* techniques, which remove edges that are likely to occur at random. In particular, we apply the disparity filter (Serrano et al. 2009) and the noise-corrected model (Coscia and Neffke 2017) using the implementation given by Coscia and Neffke (2017). Compared to other recent network backboning methods (Grady et al. 2012; Slater 2009), the disparity filter and the noise-corrected model allow for tunable parameters and provide a theoretical foundation to address noisy edge weights. The disparity filter looks at each node and the normalized weights of the  $k$  edges connecting it to its neighbors. These weights are compared to a null model, where the interval  $[0, 1]$  is subdivided by  $k - 1$  points uniformly distributed along it. The lengths of each segment would represent the expected values for the  $k$  normalized weights, and the probability of each edge having a weight compatible with this null model is calculated, corresponding to a p-value. Those edges compatible with the null model at a preset statistical significance level  $\alpha$  are removed from the backbone. The noise-corrected model is similar, but looks at a node pair rather than a node in constructing a null model. In the noise-corrected model, the edge weight distribution is assumed to be binomial, and a preset threshold analogous to the p-value decides which edges to keep in the backbone.

In most real-world networks, hubs are considered crucial to the structure and dynamics; transportation and friendship networks are notable examples where this holds. Similarly, edges that have high weights are also considered important and will not be eliminated by network backboning algorithms. However, in our word co-occurrence networks, this is not necessarily so. For example, consider the words “is” and “the”, two

of the network nodes with the highest degrees, which are also connected by an edge with one of the highest weights. As both are important components in constructing English sentences, this prominence is not unexpected. However, in terms of characterizing the meaning of a corpus, neither of these words is useful. Backboning of the word co-occurrence network requires removing these hubs as well. A tricky complication is that not all hubs are function words. For example, for the Hillary Clinton corpus, there are also words such as “vote” or “maga” that are relevant to the meaning of the corpus, but are similar to the function words “is” and “the” in that they are high-degree nodes that are connected by a high-weight edge.

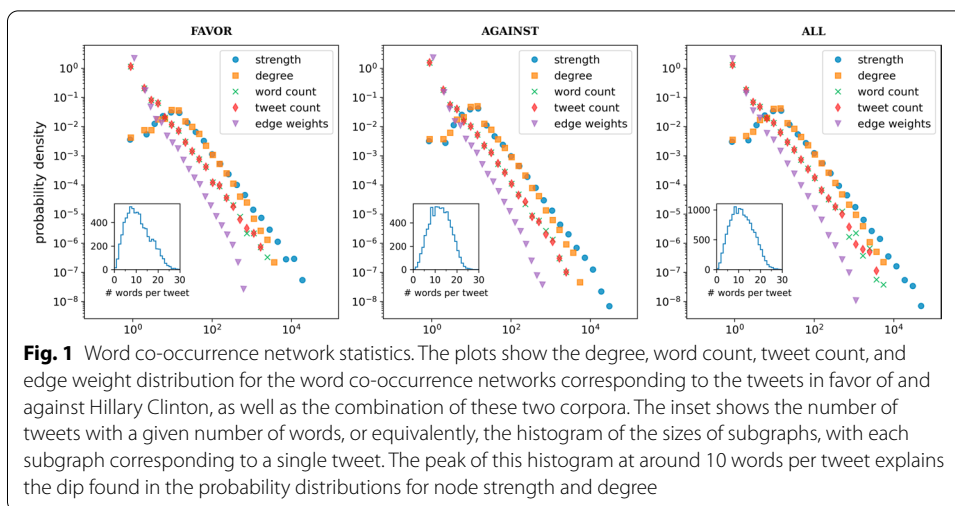
To differentiate between these two types of high-degree nodes, we create a list of words that are frequently used among several corpora. One option is to use the list of stop words included in the Python natural language toolkit (NLTK) (Loper and Bird 2002), which includes pronouns, prepositions, articles, and other function words. This gives us 180 words (including the preposition “us” which we added, as we saw it remained a prominent pronoun in the corpus after our analysis), some of which include the list of contractions we removed in parsing. Another option is to use the most commonly seen words on Twitter, which are less likely to have relevant meaning in reference to the anchored set.

Using Storywrangler (Alshaabi et al. 2021), we identify the top 400 case-sensitive 1-grams for 100 randomly chosen days from 2010/01/01 to 2015/12/31 (prior to the period corresponding to our corpus tweets) and take the intersection of these 100 lists. The resulting words frequently occur on Twitter, indicating that they are not likely to hold much distinctive meaning. After removing punctuation marks, digits and symbols, and case-insensitive duplicates from this list, we get 194 commonly used words on Twitter. We then take the intersection of this list with the top 200 words in the network with the highest degrees. The final list, which depends on the corpus, is the resulting list of stop words. Both the NLTK and the top Twitter 1-gram method give similar results. For brevity, we present the results obtained using the Twitter 1-gram method. A table containing the words removed using this method is included in the Additional file 1.

We perform network backboning in two passes. We first eliminate the words that are likely to hold less distinctive meaning using the top ranked words on general Twitter. This is equivalent to disregarding stop words in parsing, where the stop words are based on general Twitter and are only removed if they serve as hubs in the network. We then remove irrelevant edges using the disparity filter or the noise-corrected model. With these steps, both the most irrelevant and the most connected words are removed to obtain the network backbone, which is in turn used for further analysis.

We then perform community detection on the network backbone. Although an overlapping community detection algorithm is the natural choice for this task, we found that using hierarchical link clustering (Ahn et al. 2010) gives clusters with overlapping themes. We present instead the results obtained using the Louvain algorithm (Blondel et al. 2008), a widely used greedy optimization method that assigns each node to a community and returns the partitioning that maximizes the graph’s modularity. Surprisingly, despite the limitation that the communities must be disjoint, it gives the word co-occurrence network an interpretable community structure. Each community is then analyzed in terms of the words it contains and the corresponding word scores. With a





**Table 3** Number of tweets, nodes, node strengths, degree and the size of connected components for tweets

Anchors	$N_{tweets}$	$N_{nodes}$	Total edge weight	Number of edges	Component sizes
#imwithher	7222	6085	408,303	186,688	(6080, 1 × 5)
#crookedhillary	7222	9912	489,294	264,226	(9898, 3, 3, 1 × 8)
#imwithher OR #crookedhillary	14,444	12,023	897,597	405,448	(12,005, 3, 3, 1 × 12)

non-overlapping community detection algorithm, since a word only appears once in every community, the score contributions of a particular word to the community and to the unpartitioned network can also be easily interpreted.

**Results**

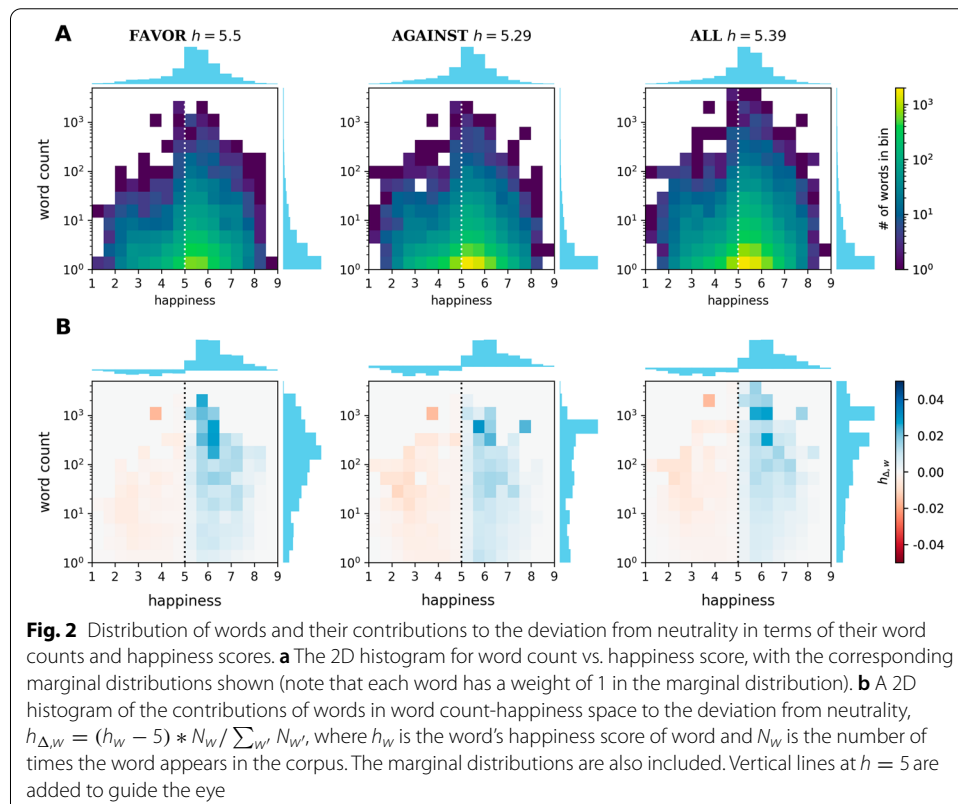
We begin by presenting the basic characteristics of the corpora, such as the degree, count, and score distributions. We then look at how node-centric properties, such as the node strength and degree, relate with word scores, and how the scores of the words connected by each edge are related.

**Network characteristics**

The number of nodes, edges, and tweets for each set of anchored tweets are given in Table 3. The distributions for the node strengths, node degrees and edge weights, as well as the word counts (the number of times a word appears in the corpus) and the tweet counts (the number of tweets a word is in), are shown in Fig. 1. The tweet and word counts display a heavy-tailed distribution, as expected (Ryland Williams et al. 2015; Williams et al. 2015), while the tail of the degree and node strength distributions follow a truncated power law with similar exponents (Clauset et al. 2009; Alstott et al. 2014). Because the tweets are short and words are not often repeated in each tweet, the construction of the co-occurrence network creates a high correlation between the degree of a node and the corresponding word count. However, even if many words are usually mentioned in very few tweets, they will co-occur with other words and will have degrees

**Table 4** Assortativity coefficients

Anchors	Strength	Degree	Happiness	
			Weighted	Unweighted
#imwithher	− 0.189	− 0.278	0.022	0.031
#crookedhillary	− 0.175	− 0.254	0.006	0.008
#imwithher OR #crook-edhillary	− 0.166	− 0.261	0.022	0.024



higher than 1. This is consistent with the distribution of the number of nodes extracted per tweet (insets of Fig. 1) and explains the dip in the probability density for the degree distribution below the value of 10, while the probability densities for the word and tweet counts are monotonic. We also note that for the vast majority of nodes, the degree is the same as the node strength. The network also has mild degree disassortativity and low node strength disassortativity (Table 4).

**Happiness scores and word counts**

To differentiate the contributions of frequent and rare words as well as positive, neutral and negative words, we construct a 2D histogram for the word counts vs. the happiness scores of the words in each corpus (Fig. 2a). Words close to the prescribed neutral score  $h = 5$  are dominant in each corpus, especially among frequently used words. The average happiness score per corpus, computed by weighting the score of each word by

its word count, is given in the annotations above the plots. To see which score ranges influence the average score the most, we construct a 2D histogram where each word is weighted by its contribution to the average happiness score. In Fig. 2b, each word  $w$  is given the weight

$$h_{\Delta,w} = (h_w - 5) * N_w / \sum_{w'} N_{w'} \quad (1)$$

where  $h_w$  is the happiness score of word  $w$  and  $N_w$  is the number of times the word appears in the corpus. Although neutral words are not expected to contribute much to this deviation, there are a few words with happiness scores  $h \in (5, 6)$  that raise the average happiness significantly, consistent with the observed positivity bias in language (Dodds et al. 2015; Aithal and Tan 2021). The average happiness scores in these corpora, however, are lower than those generally obtained with the entire Twitter Decahose dataset, which is also observed in other subsets containing political tweets (Cody et al. 2016). We also note that while low-count words do not individually contribute much to the average happiness score, their ubiquity in the corpus increases their combined influence. As expected, those with higher word counts have higher values of  $h_{\Delta,w}$ .

### Network structure and happiness scores

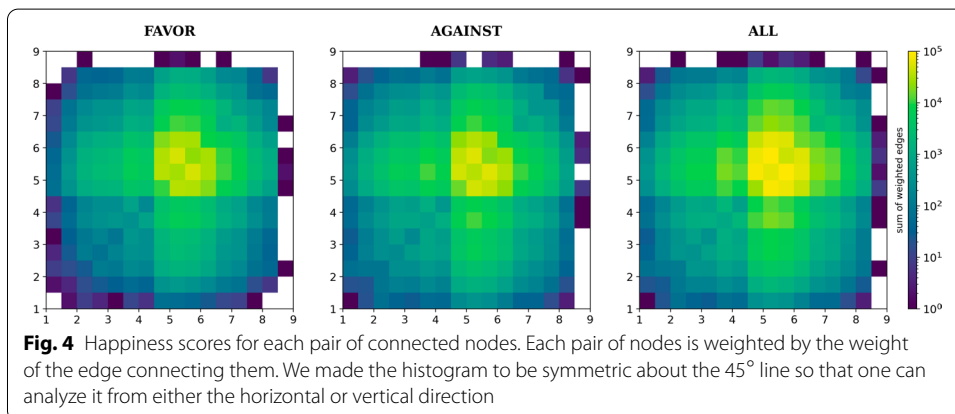
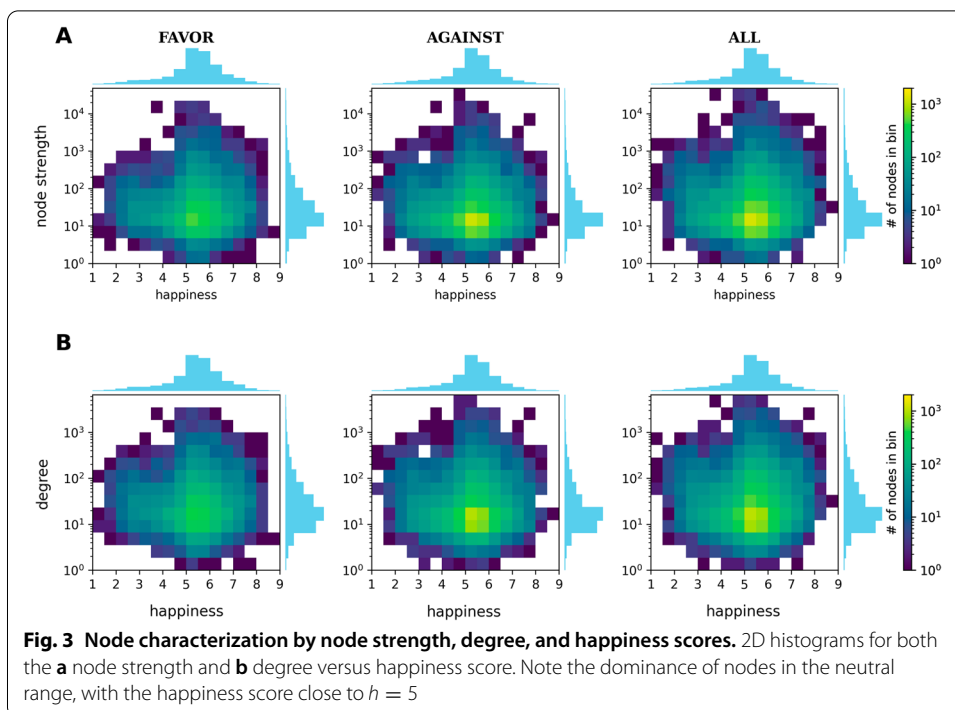
We now look at how the happiness scores and the network structure are related. We begin by looking at the relationship between a node's degree and node strength to its score, and then that of an edge and the scores of the nodes it connects.

#### *Node degree, node strength and happiness scores*

A 2D histogram showing the distribution of words in node strength-score space is given in Fig. 3. The dominance of neutral words for all values of node strengths and degrees is evident. Comparing these results to the null models (Additional file 1: Figs. S5–S8), we find that neither shuffling the scores nor changing the network structure results in a different profile, while keeping the network and using a uniform distribution for the scores changes the result substantially. Thus, this relationship between the node's score and its degree and strength is mostly due to the dominance of neutral words in the score distribution and not the structure of co-occurrence of words in the corpus.

#### *Happiness scores of connected words*

We now look at the property of each edge, in particular the scores of connected nodes and their edge weights. Figure 4 shows the profile of the scores of connected nodes in the network. We see that words of any score tend to be connected to words with happiness scores  $h \in [4.5, 6.5)$ . Whereas words in this region tend to co-occur with words of similar scores, words with more extreme scores do not exhibit such homophily. This is consistent with the low assortativity coefficients obtained for the happiness scores (Table 4) computed using both the weighted and unweighted versions of the network. Comparing these results with null models (Additional file 1: Figs. S9–S12), we find that rewiring the network and shuffling the scores give similar profiles, while keeping the network but using a uniform distribution changes the profile drastically. Thus, this homophily among



neutral words is mainly due to the abundance of neutral words in the corpus and not the co-occurrence network structure.

### Network backboning

We saw in “[Network structure and happiness scores](#)” section that the network itself plays a smaller role than the score distribution, both from a node-centric perspective (score, degree, and node strength), and a edge-centric perspective (score difference and edge weights). In particular, the dominance of neutral words in the corpus makes it more likely that any two connected words are close to neutral and thus have similar scores. Previous work has observed this dominance of neutrality in text corpora which tends to reduce the signal in differences in sentiment across time (Dodds et al. 2011). This issue was previously addressed by using a manually tuned filter to remove neutral words while

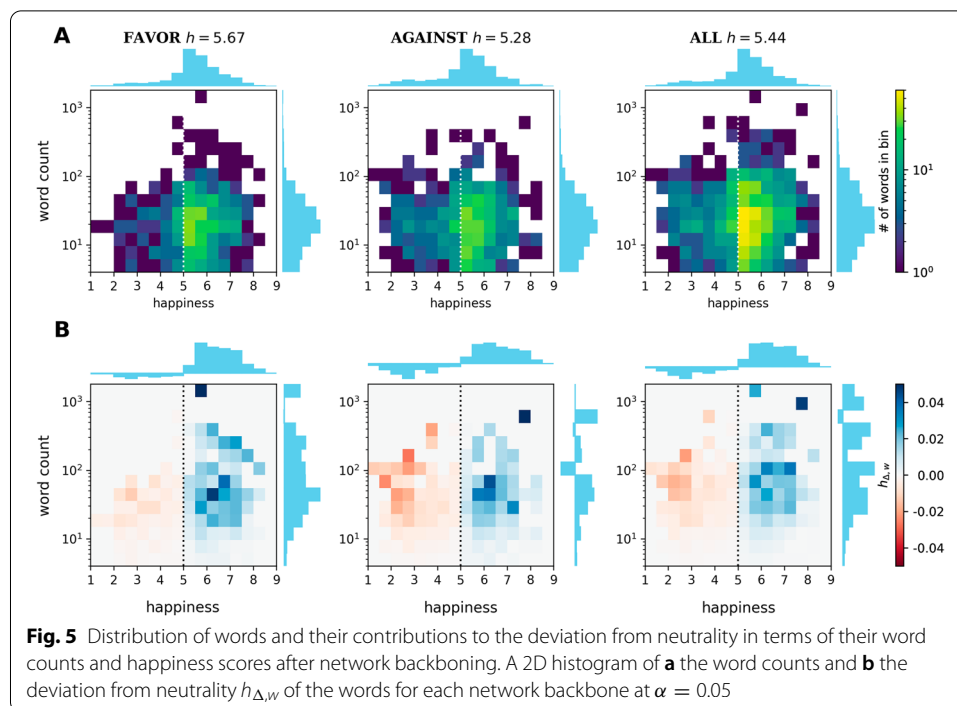
still capturing time series correlations (Dodds et al. 2011); two commonly used filters exclude words with scores between 4.5 and 5.5 or 4 and 6 before taking the average of the happiness scores (weighted by word count) of the remaining words in the corpus.

Although the removal of neutral words may be sufficient in obtaining an average happiness score for a given corpus, this may not be the best approach when the aim is to explore the relationship between network structure and scores, as removing neutral words in performing network analysis begs the question by presupposing the existence of this relationship. Further, as neutral words form a big chunk of the network regardless of node degree, node strength or edge weight, disregarding them will remove possibly influential nodes that may help us understand hidden substructures in the network, such as opposing sentiments within the corpus, that would not be visible if we look at just the weighted average of the happiness scores. Some context-relevant words (e.g., “vote”) may also be neutral, and the removal of these words may make it more difficult to interpret the co-occurrence network structure. We want to see if we can obtain useful word sentiment information from the network structure without resorting to score cutoffs.

Network backboning uncovers relevant structure in a network by removing edges that are likely to be spurious connections. If the pruning of these edges results in the isolation of nodes, one can consider these isolated nodes to be less relevant to the network structure than the ones that remain in the backbone. As discussed in “[Community detection and network backboning](#)” section, we perform network backboning in two steps. To remove the nodes which are not likely to hold much meaning, we remove the most common words on Twitter from the network if they are also frequently occurring in the corpus, which is equivalent to parsing the original tweets with these words as stop words. We then use the disparity filter or the noise-corrected model to remove low-weight edges. We find that the noise-corrected model removes few edges from the network despite setting the significance level to be very low; in contrast, the disparity filter removes both edges and nodes at a wide range of thresholds. This is unsurprising, as the noise-corrected model gives a higher importance to edges that connect low-degree nodes compared to the disparity filter and is known to be the less restrictive algorithm of the two. In the following discussion, we only consider results obtained using the disparity filter. The disparity filter only begins to take out edges at threshold values  $\alpha < 0.5$ , with a sharp drop in the number of nodes and edges beginning at  $\alpha = 0.3$ . Although similar communities are found for  $\alpha = 0.3$  up to the most restrictive value  $\alpha = 0.05$ , we find that  $\alpha = 0.05$  removes all edges with a weight of 1, and consequently produces fewer irrelevant words. We thus present the results corresponding to  $\alpha = 0.05$  in the rest of the manuscript. More details regarding the differences between different values of  $\alpha$  are found in the Additional file 1.

Depending on the corpus, the edges that are removed at  $\alpha = 0.05$  are not random (Additional file 1: Fig. S17). In the “favor” corpus, the vast majority of edges connecting negative words to other negative words have been removed, as well as edges that connect extremely positive words. In contrast, the removal of edges in the “against” corpus does not have a clear dependence on the scores. In all corpora, the dominance of neutral-neutral connections still holds, albeit at a tighter range.

Figure 5 shows the histogram of the contribution to the deviation from neutrality of each word,  $h_{\Delta,w}$  for  $\alpha = 0.05$ , the most restrictive of the values tested. Compared to the



complete network (Fig. 2), the “favor” corpus lost a significant number of words with happiness scores below  $h = 5$ . On the contrary, the “against” corpus did not lose as many negative words, although many positive words remained. There is also a clearer difference in the score profiles of the “favor” and “against” backbone networks, and this was attained not through the use of score cutoffs but by using the co-occurrence network structure itself. Aside from this, there are now fewer words in the backbone that have low word counts, increasing the overall influence of non-neutral words that occur with higher frequency. The “all” corpus, being a combination of the two, has similarities to both. To see if we can extract the properties of the “favor” and “against” stances in the “all” corpus, we turn to community detection.

### Community detection and characterization

We have seen that by removing commonly used words on Twitter, a number of high-degree neutral words were removed, increasing the proportion of positive and negative words in the corpus. Using the disparity filter further removed low-weight edges. With more restrictive thresholds, edges connecting negative words to other negative words are removed in the “favor” corpus, while no distinct patterns are seen in the “against” and “all” corpora. Whereas negative words were removed from both the “favor” and “against” corpora, the “favor” corpus lost more negative words, resulting in an amplified difference between the average happiness score of the two corpora.

We now investigate whether there is meaningful community structure in the co-occurrence network, and whether this structure can be used to uncover and isolate mixed sentiments in the corpora, particularly in the “all” corpus where we expect opposing sentiments to exist. We find that different values of the disparity filter threshold  $\alpha$  all give consistent communities that fit a given theme, with more restrictive values of



$\alpha$  keeping the communities that can be attributable to known events and reducing the number of communities with words that are less coherent. The following results were obtained using the Louvain algorithm on the backbone derived using  $\alpha = 0.05$ .

### **Characterizing communities**

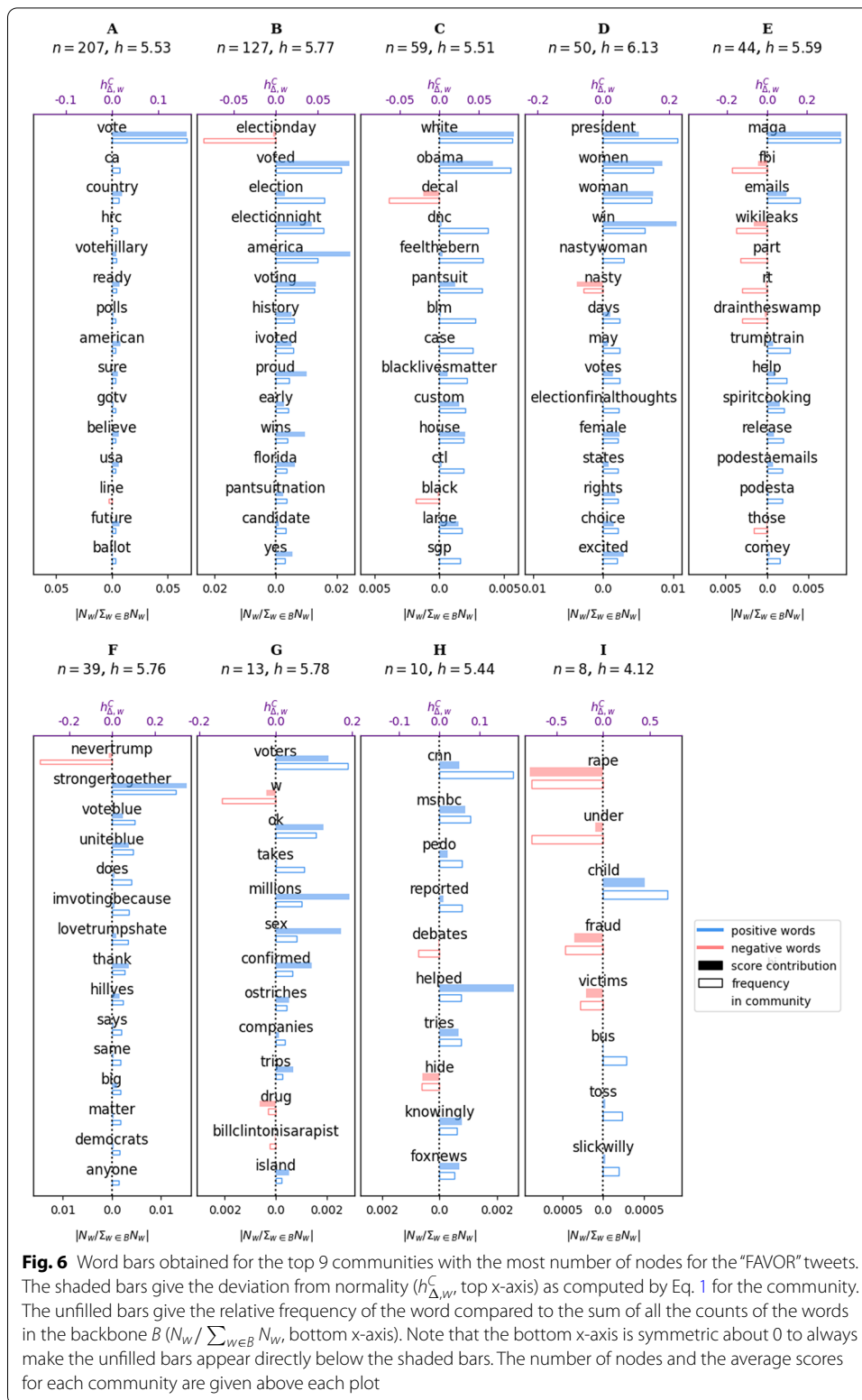
Figures 6 to 8 show the words with the highest word counts in each of the biggest communities (note that for the backbone network, the degree is no longer necessarily correlated with the word counts), their contributions to the deviation from normality of the average score of the community, and their word counts in the corpus. We see that the biggest communities consist of related words and form a consistent narrative.

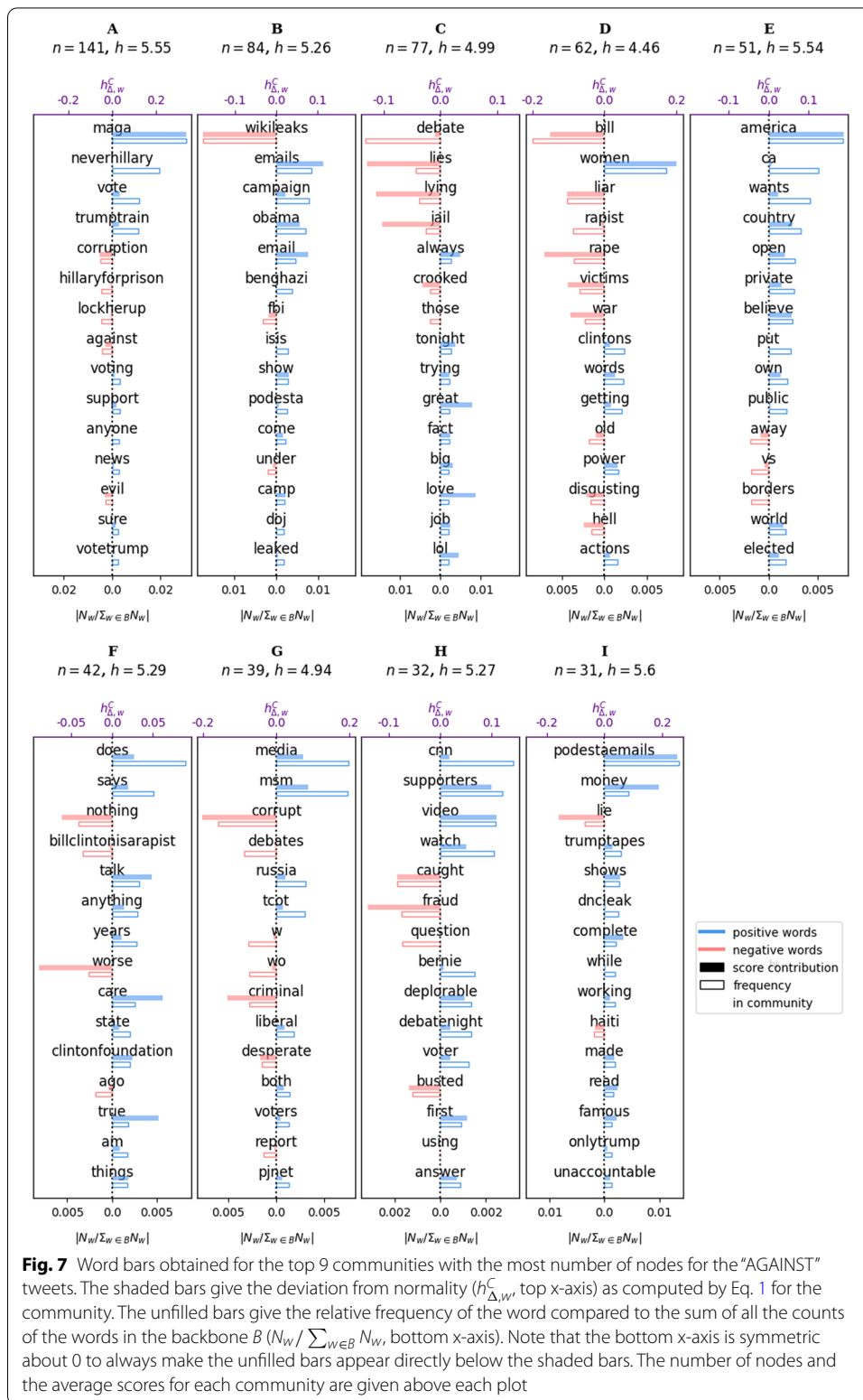
In the “favor” corpus (Fig. 6), we see that the communities A and B are about the elections; D relates to being a woman president. We note that despite (or rather, due to) “nasty woman” being a term used by Donald Trump to denigrate Hillary Clinton, the phrase has been used as a rallying cry by feminists to support Clinton (Gray 2016), so it is not a surprise to find it with words that are likely to favor her. On the other hand, Community E is a community with Donald Trump’s “MAGA” slogan (short for “Make America Great Again”). As expected, it also contains other words and hashtags that relate to scandals concerning Hillary Clinton, but the corresponding word counts are less than those in the biggest communities A and B. Community C contains words that are related to the Democratic party, with the word “white” likely often used in the context of both the “White House” and the Black Lives Matter movement.

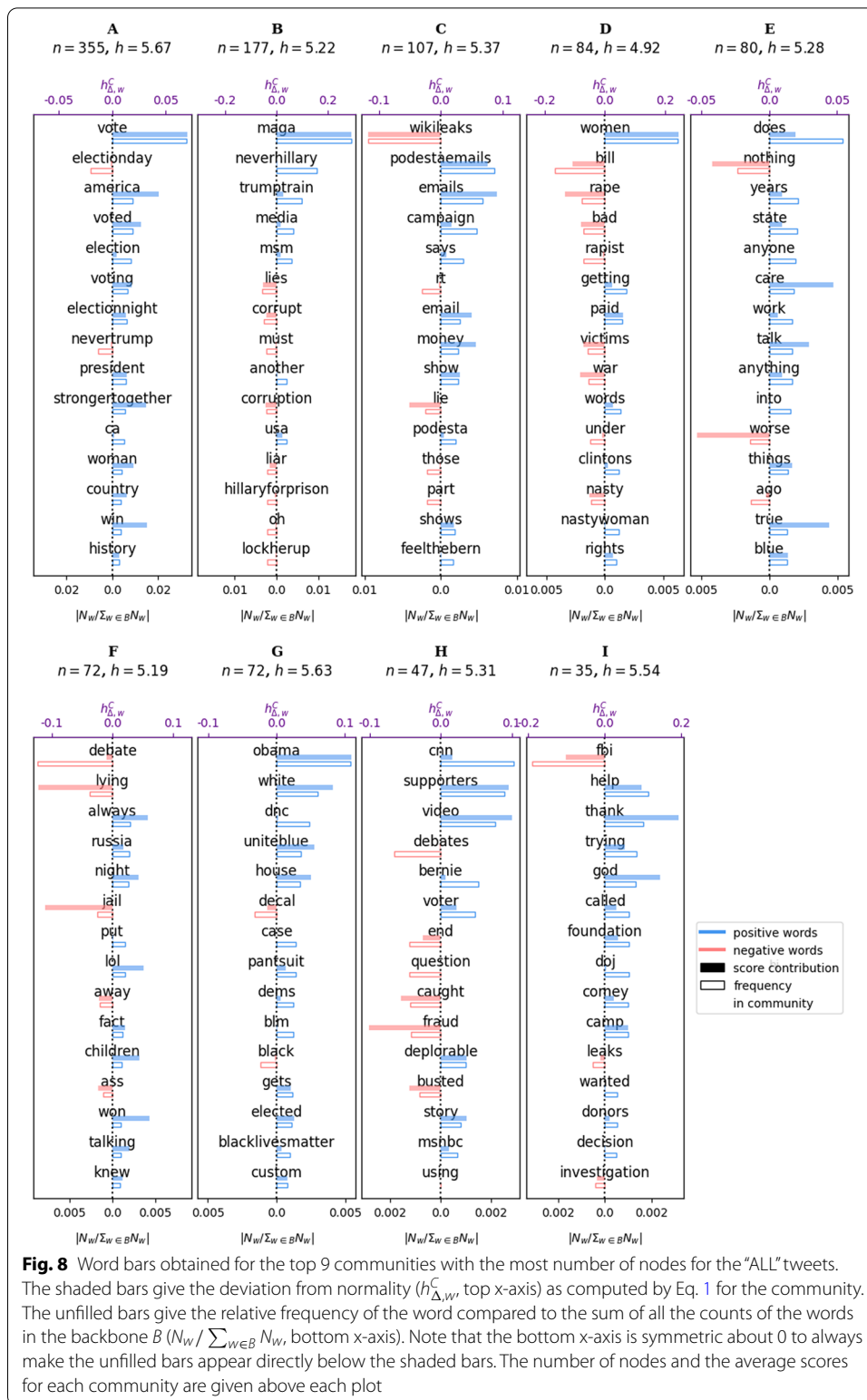
In each of the larger communities (except for community E), the most frequently occurring words are the words that we expect to be associated with a favorable stance towards Hillary Clinton. The smaller communities contain words that do not necessarily conform to a theme, although we note that there are a number of isolated communities that contain infrequently used words. Examples of these are: “louis” and “ck” (Louis C.K. is a comedian); “west” and “coast” (West Coast is used to refer to the west coast of the United States); and “lady” and “gaga” (Lady Gaga is a singer). Overall, the consistency of our results with the expectations from our domain knowledge supports the validity of the backboning and community detection algorithms used.

Similar patterns can be found in the “against” corpus (Fig. 7). The biggest community, A, is characterized by words that are for voting Trump rather than Clinton (“maga”, “vote”, “trumptrain”) and also include hashtags that are against Hillary such as “never-hillary”, “hillaryforprison”, and “lockherup”). Community B contains words relating to scandals against Hillary Clinton (“wikileaks”, “emails”, “benghazi”, “collusion”, “podesta”), while community D contains the words “bill”, “women”, and “rapist”, which are likely related to rape allegations against Hillary Clinton’s husband Bill Clinton. Some smaller communities have words that are related to a theme, such as community G (“msm”, “media”, and “debates”) and community H (“cnn”, “video”, and “caught” as well as “anderson”, “cooper”, “msnbc”) , but there are also words in these communities that are not related.

In the combined corpus (Fig. 8), we see communities similar to the dominant ones in the “favor” and “against” corpora. Community A is about voting, community B contains words that are either against Hillary or for Donald Trump, community C contains words relating to scandals involving Hillary Clinton, and community D contains words







**Fig. 8** Word bars obtained for the top 9 communities with the most number of nodes for the “ALL” tweets. The shaded bars give the deviation from normality ( $h_{\Delta,w}^C$ , top x-axis) as computed by Eq. 1 for the community. The unfilled bars give the relative frequency of the word compared to the sum of all the counts of the words in the backbone  $B$  ( $N_w / \sum_{w \in B} N_w$ , bottom x-axis). Note that the bottom x-axis is symmetric about 0 to always make the unfilled bars appear directly below the shaded bars. The number of nodes and the average scores for each community are given above each plot

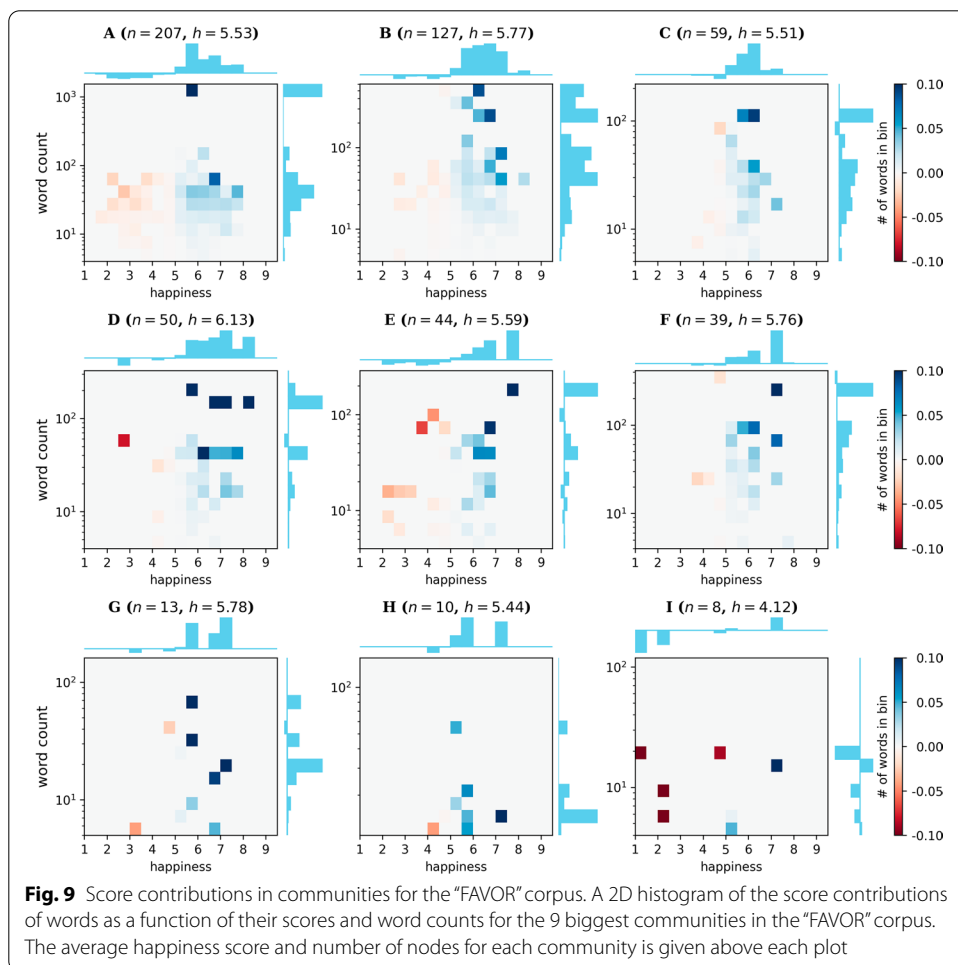
that relate to the rape allegations against Bill Clinton. Some themes in smaller communities also exist. In community G, we see words relating to other Democrats and the Black Lives Matter movement, while community H contains words about the media (“cnn”,

“msnbc”, as well as “nbc”, “foxnews”, “anderson”, “cooper”, “bernie” and “sanders”, referring to Bernie Sanders who was Hillary Clinton’s main opponent in the Democrat primaries).

**Communities and scores**

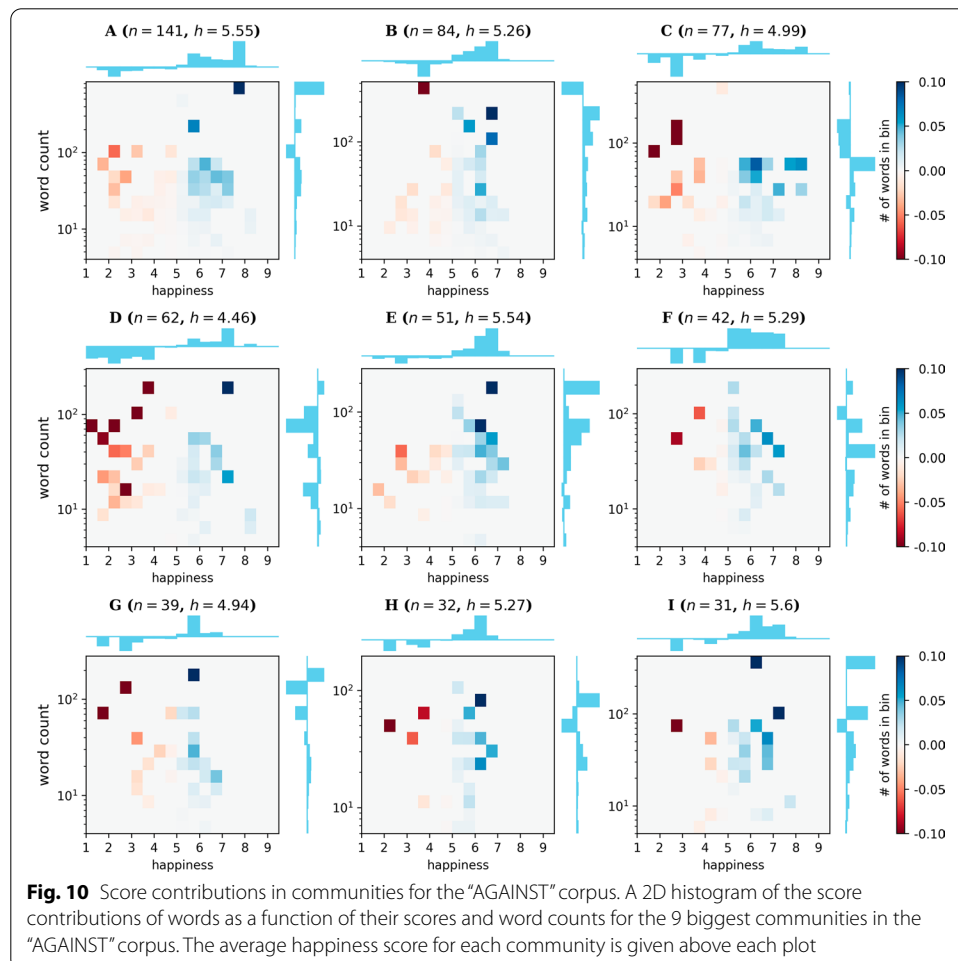
We have shown how disjoint communities in the network correspond to certain themes, with themes that are for or against Hillary Clinton clearly separated in the different corpora. We now explore whether communities of different themes differ in terms of the scores of the words they contain.

The biggest communities in “favor” corpus are dominated by positive contributions from normality (Fig. 9, while notable negative words include “nasty” (community D), “wikileaks” (community E), and “rape” (community I). As discussed earlier, the word “nasty” (in the context of “nasty woman”), though inherently negative, has been used by supporters of Hillary Clinton. Closer examination of tweets reveal that although the hashtag “#imwithher” is known mainly to express opinions favorable to Hillary Clinton, it is also hijacked by those who express negative opinions about her, and we find that community detection on the backbone of the co-occurrence network was able to isolate these negative themes.



**Fig. 9** Score contributions in communities for the “FAVOR” corpus. A 2D histogram of the score contributions of words as a function of their scores and word counts for the 9 biggest communities in the “FAVOR” corpus. The average happiness score and number of nodes for each community is given above each plot

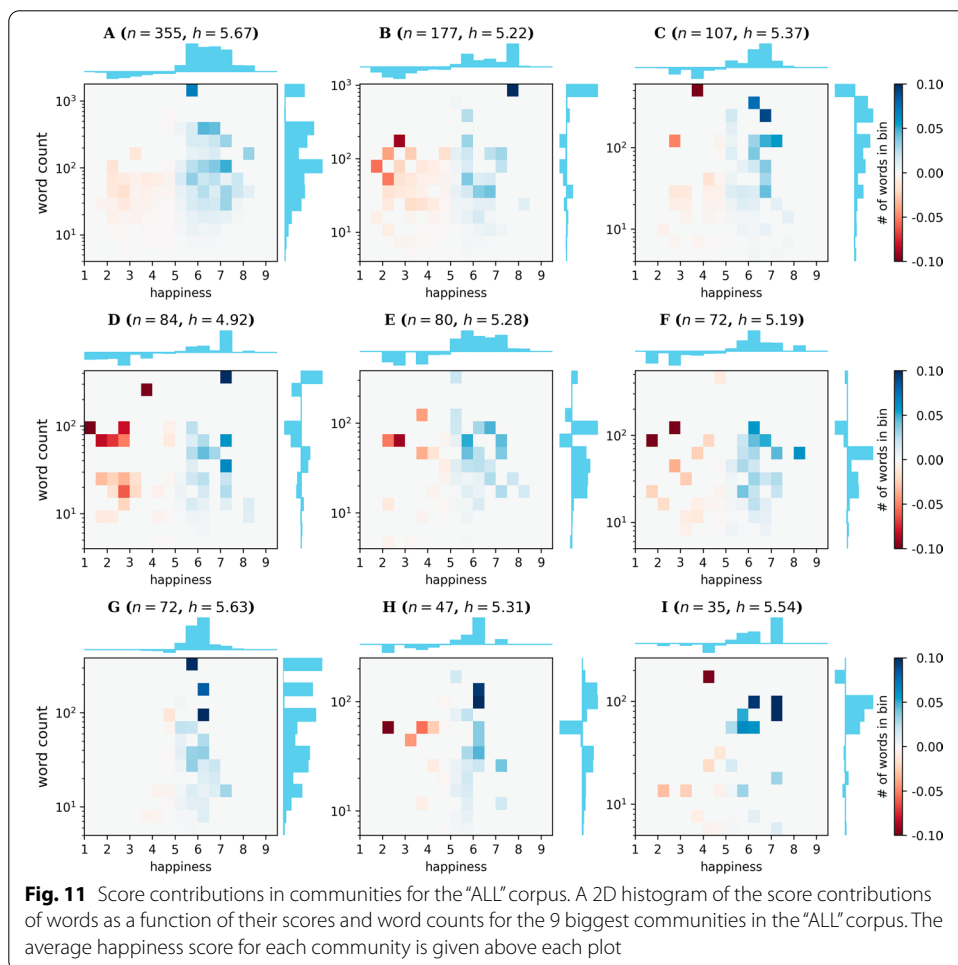
The biggest community (community A) in the “against” corpus is characterized by words of support for Donald Trump, primarily by the acronym “maga” (“make America great again”), Donald Trump’s slogan in the 2016 elections (Fig. 10. Community B contains words relating to the scandals involving Hillary Clinton (“wikileaks”). It also contains “emails” which has a positive score in the labMT dataset, despite it having a negative connotation in the context of the 2016 US presidential elections. Community C has the negative words “lies”, “lying”, and “jail”, but also the positive words “love” and “great”. We note that although “love” and “great” are some of the most commonly used words on Twitter, they are not as frequently used in this corpus and were therefore not filtered out. Closer examination of the raw tweets show that these positive words are used in a negative or sarcastic sense against Hillary Clinton. Community D has an extremely low negative score due to the words associated with “rape”. The word “bill”, which was mostly used in the context of “Bill Clinton”, has a score of 3.64, as it was probably interpreted by the labMT respondents to mean “invoice”, as in “medical bill” or “electric bill”. Regardless, even if “bill” were not scored, we see that the words in community D were mainly negative, consistent with its association with “rape”. A notable exception is the word “women”, which has a happiness score of 7.12, even though it is used often with the word “rape” in this context.

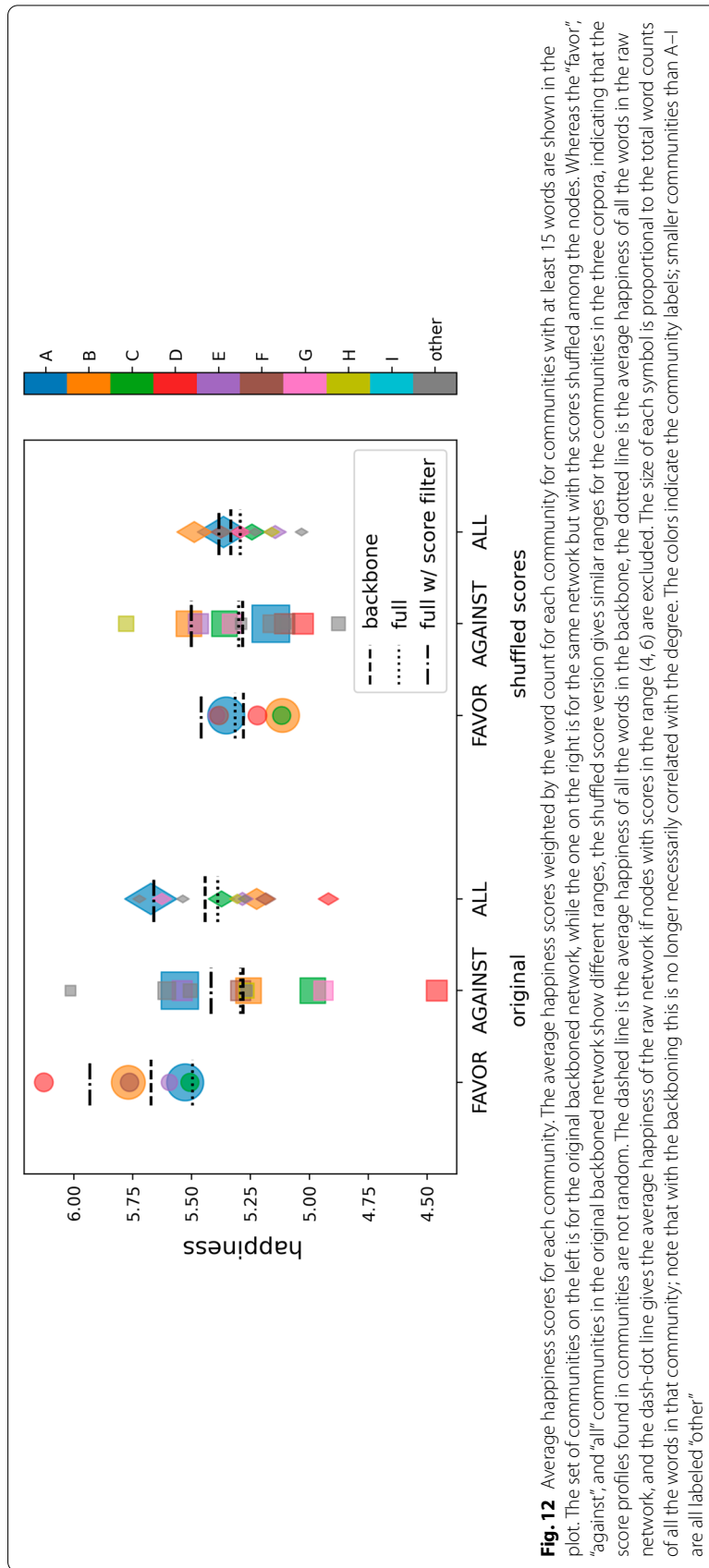




The biggest community (community A) in the combination of the two corpora (“all”) is mainly positive (Fig. 11, which is consistent with the fact that the most frequently used words in this community are in support of Hillary Clinton. Community B has a fair share of negative words that are used against Hillary Clinton (“lies”, “corrupt”, “criminal”) but are also used with Trump’s slogan “maga”, which carries a positive score. Community C, being about particular scandals such as “wikileaks”, has fewer negative words. Further, some words that are used in a negative sense in this context are in general either positive or neutral words (“emails”, “spiritcooking”). Community D contains many negative words associated with “rape”, but also positive words such as “women”, which in turn commonly occurs with “nasty” (in the context of “nasty women”). Communities E and F both have dominant negative words, but the words in these communities do not share coherent themes. Communities G and H both have dominant negative words, but the words in these communities do not share coherent themes. Community I has a fair share of negative words that are used against Hillary Clinton (“lies”, “corrupt”, “criminal”) but are also used with Trump’s slogan “maga”, which carries a positive score.

Figure 12 shows the average happiness scores weighted by the word count for each community with at least 15 nodes. For comparison, we also include the average scores per community obtained if the scores were shuffled across the unfiltered network, but the network structure (and thus the communities) were retained. In the shuffled version, the range of the community scores are similar across stances, which is not the case in the original network, indicating that the score profiles found for the communities are





**Fig. 12** Average happiness scores for each community. The average happiness scores weighted by the word count for each community for communities with at least 15 words are shown in the plot. The set of communities on the left is for the original backbone network, while the one on the right is for the same network but with the scores shuffled among the nodes. Whereas the “favor”, “against”, and “all” communities in the original backbone network show different ranges, the shuffled score version gives similar ranges for the communities in the three corpora, indicating that the score profiles found in communities are not random. The dashed line is the average happiness of all the words in the backbone, the dotted line is the average happiness of all the words in the raw network, and the dash-dot line gives the average happiness of the raw network if nodes with scores in the range (4, 6) are excluded. The size of each symbol is proportional to the total word counts of all the words in that community; note that with the backbone this is no longer necessarily correlated with the degree. The colors indicate the community labels; smaller communities than A–I are all labeled “other”

not random. We also include in the figure the average happiness scores computed if no network backboning was implemented, but if words close to neutral ( $4 < h < 6$ ) were disregarded, as is customary in studies using happiness scores (Dodds et al. 2011, 2015). While removing these words increases the scores across all corpora, reflecting the dominance of positive words in text (Dodds et al. 2015), network backboning only increased the scores in the “favor” corpus and not in the “against” corpus, resulting in a moderate increase in the “all” corpus compared to the average happiness score derived from the complete network.

### Discussion and summary

From the word co-occurrence network, we examined how the co-occurrence structure of words in political tweets is related to the word happiness scores. In particular, we looked at three corpora that relate to the American politician Hillary Clinton: one collected using the hashtag “#imwithher”, which is associated with being in favor of Clinton; another collected using the hashtag “#crookedhillary”, which is associated with being against her; and a combination of both corpora. The degree, node strength, word count and tweet count distributions are all heavy-tailed, consistent with what is expected from Zipf’s law (Ryland Williams et al. 2015; Williams et al. 2015). Although it is expected that neutral words would dominate the corpus without any backboning or filtering, we find little assortative mixing outside the neutral range. Contrary to our expectation, while positive or negative words are more likely to co-occur with neutral words, positive or negative words are not more likely to co-occur with words of similar polarity. A clear demonstration of this is the use of the hashtag “#maga”, which on its own exhibits a highly positive sentiment. However, as a slogan for Donald Trump, it is also used together with words with a negative sentiment against Hillary Clinton. Further, we compared the score profiles with null models that change the network structure or the scores, and find that network structure plays a much smaller role in the score profile than the score distribution.

Given that several words in the corpus do not occur frequently, we extracted a more robust network structure using network backboning. We implemented it in two passes, first by removing frequently occurring words on Twitter, and second, by using the disparity filter to remove edges between words that co-occur less frequently. Performing community detection on the resulting backbone results in well-defined communities that correspond to themes in favor or against the target (Hillary Clinton), even for the “favor” or “against” corpora that lean more towards only one stance. Effectively, the Louvain algorithm, a non-overlapping community detection model, functions like a topic model when applied to the word co-occurrence network when co-occurrence is taken at a tweet level.

The average happiness scores of the communities also correspond to the theme that they are associated with. Communities dominated by words in favor of Hillary Clinton tend to have higher happiness scores, while those dominated by words against her tend to have lower happiness scores. A few context-specific words whose scores do not reflect their meaning are “nastywoman”, “nasty”, and the name “bill”, but even if these words are disregarded, we still see similar behavior for the rest of the words in their communities. We also note that both positive and negative words are present in each

community. Notably, even in the most negative communities in the “against” corpus, we find positive words, reaffirming the positivity bias of human language (Dodds et al. 2015) found even in negative statements (Aithal and Tan 2021).

In summary, we found that raw co-occurrence networks in tweets are dominated by neutral words as well as connections between them, and no score homophily is observed. Network backboning reduces the influence of neutral words, but does not uncover assortative mixing. However, when split into communities, the average happiness score of the community corresponds with the theme. Further, opposing themes are found, not just when the opposing tweets are equal in number (as in the “all” corpus), but also when an opposing theme is in the minority. This indicates that the word co-occurrence network with backboning and community detection can be used to detect the presence of opposing sentiments within a corpus.

We have only tested our methods on political tweets, and also for a limited subset. Some studies of Twitter’s free-tier Streaming API, which takes a 1% random sample of all tweets, have shown that bias may arise when dealing with very popular terms, likely due to rate limits set by Twitter (Tromble et al. 2017; Campan et al. 2018), but that having a higher coverage of the entire set of tweets reduces the problem (Morstatter et al. 2013). A comparison between the free-tier API (1% random sample) and subsamples of the Decahose (10% random sample) API also showed that the number of new entities (hashtags, mentions, users, and proper nouns) in the sample plateaus with increasing coverage, so that the advantage of increasing the sample size from 1 to 2% is greater than increasing it from, for example, 5–6% (Li et al. 2016). Although we found no direct comparisons between the Decahose API and the Firehose API (100% of all tweets), we can infer from these previous studies that the expected bias due to subsampling will be mitigated by using the Decahose API, as we have done in this study. We also note that the Decahose API is the best available option offered by Twitter after the comprehensive but prohibitively expensive Firehose API.

Because generating the word co-occurrence network does not involve any filters, applying our analysis to a larger dataset, such as an entire day’s worth of tweets or a more popular anchor such as all tweets mentioning the words “Trump” or “BTS”, which have extremely high word counts on Twitter (Dodds et al. 2021), would be computationally expensive. Implementing a suitable filter in the network generation step may be necessary and is an avenue for future work. We would also like to expand our work in the future to analyze and compare tweets from different domains, as well as include higher order n-grams in our analysis.

### Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1007/s41109-022-00446-2>.

**Additional file 1:** Supplementary Material.

### Acknowledgements

M.I.F. is grateful to Takayuki Hiraoka for helpful discussions.

### Authors’ contributions

MIF: Designed the study, performed the analysis, and authored the manuscript. TA: Contributed code and research expertise, edited the manuscript. MVA: Contributed research expertise and edited the manuscript. CMD and PSD: Contributed funding and research expertise, aided in study design, and edited the manuscript. All authors read and approved the final manuscript.

**Funding**

Computations were performed on the Vermont Advanced Computing Core supported in part by NSF Award No. OAC-1827314. The authors acknowledge support from NSF Award No. 2117345 and a gift from MassMutual Life Insurance to the Vermont Complex Systems Center.

**Availability of data and materials**

The data used to generate our results are available at <https://doi.org/10.6084/m9.figshare.19136933>.

**Declarations****Competing interests**

The authors declare that they have no competing interests.

**Author details**

<sup>1</sup>Computational Story Lab, Vermont Complex Systems Center, MassMutual Center of Excellence for Complex Systems and Data Science, University of Vermont, Burlington, VT, USA. <sup>2</sup>Advanced Bioimaging Center, UC Berkeley, Berkeley, CA, USA. <sup>3</sup>Department of Mathematics and Statistics, University of Vermont, Burlington, VT, USA. <sup>4</sup>Department of Computer Science, University of Vermont, Burlington, VT, USA.

Received: 23 October 2021 Accepted: 31 January 2022

Published online: 14 February 2022

**References**

- Ahn Y-Y, Bagrow JP, Lehmann S (2010) Link communities reveal multiscale complexity in networks. *Nature* 466(7307):761–764. <https://doi.org/10.1038/nature09182>
- Aithal M, Tan C (2021) On positivity bias in negative reviews. arXiv:2106.12056 [cs]. Accessed 1 July 2021
- Al Rozz Y, Hamoodat H, Menezes R (2017) Characterization of written languages using structural features from common corpora. In: Gonçalves B, Menezes R, Sinatra R, Zlatic V (eds) *Complex networks VIII*. Springer proceedings in complexity. Springer, Cham, pp 161–173. [https://doi.org/10.1007/978-3-319-54241-6\\_14](https://doi.org/10.1007/978-3-319-54241-6_14)
- Alshaabi T, Adams JL, Arnold MV, Minot JR, Dewhurst DR, Reagan AJ, Danforth CM, Dodds PS (2021) Storywrangler: a massive exploratorium for sociolinguistic, cultural, socioeconomic, and political timelines using Twitter. *Sci Adv* 7(29):6534. <https://doi.org/10.1126/sciadv.abe6534>
- Alshaabi T, Van Oort C, Fudolig MI, Arnold MV, Danforth CM, Dodds PS (2022) Augmenting semantic lexicons using word embeddings and transfer learning. *Front Artif Intell* 4:66. <https://doi.org/10.3389/frai.2021.783778>
- Alstott J, Bullmore E, Plenz D (2014) powerlaw: a python package for analysis of heavy-tailed distributions. *PLoS ONE* 9(1):85777. <https://doi.org/10.1371/journal.pone.0085777>
- Baccianella S, Esuli A, Sebastiani F (2010) SentiWordNet 3.0: an enhanced lexical resource for sentiment analysis and opinion mining. [http://www.lrec-conf.org/proceedings/lrec2010/pdf/769\\_Paper.pdf](http://www.lrec-conf.org/proceedings/lrec2010/pdf/769_Paper.pdf). Accessed 28 June 2021
- Bestgen Y, Vincze N (2012) Checking and bootstrapping lexical norms by means of word similarity indexes. *Behav Res Methods* 44(4):998–1006. <https://doi.org/10.3758/s13428-012-0195-z>
- Blondel VD, Guillaume J-L, Lambiotte R, Lefebvre E (2008) Fast unfolding of communities in large networks. *J Stat Mech Theory Exp* 10:10008. <https://doi.org/10.1088/1742-5468/2008/10/P10008>
- Bojanowski P, Grave E, Joulin A, Mikolov T (2017) Enriching word vectors with subword information. *Trans Assoc Comput Linguist* 5:135–146. [https://doi.org/10.1162/tacl\\_a\\_00051](https://doi.org/10.1162/tacl_a_00051)
- Bradley MM, Lang PJ (1999) Affective Norms for English Words (ANEW): instruction manual and affective ratings. Technical Report C-1, The Center for Research in Psychophysiology, University of Florida
- Campan A, Atrafu T, Truta TM, Nolan J (2018) Is data collection through Twitter Streaming API useful for academic research? In: 2018 IEEE international conference on big data (big data), pp 3638–3643. <https://doi.org/10.1109/BigData.2018.8621898>
- Chen H, Chen X, Liu H (2018) How does language change as a lexical network? An investigation based on written Chinese word co-occurrence networks. *PLoS ONE* 13(2):0192545. <https://doi.org/10.1371/journal.pone.0192545>
- Clauset A, Shalizi CR, Newman MEJ (2009) Power-law distributions in empirical data. *SIAM Rev* 51(4):661–703. <https://doi.org/10.1137/070710111>
- Cody EM, Reagan AJ, Dodds PS, Danforth CM (2016) Public opinion polling with Twitter. arXiv:1608.02024 [physics]. 2 Aug Accessed 2021
- Cong J, Liu H (2014) Approaching human language with complex networks. *Phys Life Rev* 11(4):598–618. <https://doi.org/10.1016/j.plrev.2014.04.004>
- Coppersmith G, Dredze M, Harman C (2014) Quantifying mental health signals in Twitter. In: *Proceedings of the workshop on computational linguistics and clinical psychology: from linguistic signal to clinical reality*. Association for Computational Linguistics, Baltimore, Maryland, pp 51–60. <https://doi.org/10.3115/v1/W14-3207>. Accessed 29 June 2021
- Coscia M, Neffke FMH (2017) Network Backboning with noisy data. In: 2017 IEEE 33rd international conference on data engineering (ICDE), pp 425–436. <https://doi.org/10.1109/ICDE.2017.100>
- Dodds PS, Harris KD, Kloumann IM, Bliss CA, Danforth CM (2011) Temporal patterns of happiness and information in a global social network: hedonometrics and Twitter. *PLoS ONE* 6(12):26752. <https://doi.org/10.1371/journal.pone.0026752>

- Dodds PS, Clark EM, Desu S, Frank MR, Reagan AJ, Williams JR, Mitchell L, Harris KD, Kloumann IM, Bagrow JP, Meg-erdoomian K, McMahon MT, Tivnan BF, Danforth CM (2015) Human language reveals a universal positivity bias. *Proc Natl Acad Sci* 112(8):2389–2394. <https://doi.org/10.1073/pnas.1411678112>
- Dodds PS, Minot JR, Arnold MV, Alshaabi T, Adams JL, Dewhurst DR, Reagan AJ, Danforth CM (2021) Fame and ultraframe: measuring and comparing daily levels of 'being talked about' for United States' presidents, their rivals, God, countries, and K-pop. arXiv:1910.00149 [physics]. Accessed 30 June 2021
- Fang X, Zhan J (2015) Sentiment analysis using product review data. *J Big Data* 2(1):5. <https://doi.org/10.1186/s40537-015-0015-2>
- Ferrer i Cancho R, Solé RV, Köhler R (2004) Patterns in syntactic dependency networks. *Phys Rev E* 69(5). <https://doi.org/10.1103/PhysRevE.69.051915>
- Garg M, Kumar M (2018) The structure of word co-occurrence network for microblogs. *Phys A Stat Mech Appl* 512:698–720. <https://doi.org/10.1016/j.physa.2018.08.002>
- Gerlach M, Peixoto TP, Altmann EG (2018) A network approach to topic models. *Sci Adv* 4(7):1360. <https://doi.org/10.1126/sciadv.aag1360>
- Grady D, Thiemann C, Brockmann D (2012) Robust classification of salient links in complex networks. *Nat Commun* 3(1):864. <https://doi.org/10.1038/ncomms1847>
- Gray E (2016) How 'Nasty Woman' became a viral call for solidarity. *Huffington Post*. Section: Women. Accessed 23 June 2021
- Hollis G, Westbury C (2016) The principals of meaning: extracting semantic dimensions from co-occurrence models of semantics. *Psychonom Bull Rev* 23(6):1744–1756. <https://doi.org/10.3758/s13423-016-1053-2>
- Hollis G, Westbury C, Lefsrud L (2017) Extrapolating human judgments from skip-gram vector representations of word meaning. *Q J Exp Psychol* 70(8):1603–1619. <https://doi.org/10.1080/17470218.2016.1195417>
- Howard PN, Duffy A, Freelon D, Hussain MM, Mari W, Maziad M (2011) Opening closed regimes: what was the role of social media during the Arab Spring? SSRN Scholarly Paper ID 2595096, Social Science Research Network, Rochester, NY. <https://doi.org/10.2139/ssrn.2595096>. Accessed 29 June 2021
- Jiang J, Yu W, Liu H (2019) Does scale-free syntactic network emerge in second language learning? *Front Psychol*. <https://doi.org/10.3389/fpsyg.2019.00925>
- Joulin A, Grave E, Bojanowski P, Mikolov T (2017) Bag of tricks for efficient text classification, pp 427–431. <https://www.aclweb.org/anthology/E17-2068>. Accessed 28 June 2021
- Klimiuk K, Czoska A, Biernacka K, Balwicki L (2021) Vaccine misinformation on social media-topic-based content and sentiment analysis of Polish vaccine-deniers' comments on Facebook. *Hum Vacc Immunotherap* 17(7):2026–2035. <https://doi.org/10.1080/21645515.2020.1850072>
- Küçük D, Can F (2020) Stance detection: a survey. *ACM Comput Surv* 53(1):1–37. <https://doi.org/10.1145/3369026>
- Li Q, Shah S, Thomas M, Anderson K, Liu X, Fang R (2016) How much data do you need? Twitter dechase data analysis. In: The 9th International conference on social computing, behavioral-cultural modeling & prediction and behavior representation in modeling and simulation
- Liu H, Cong J (2013) Language clustering with word co-occurrence networks based on parallel texts. *Chin Sci Bull* 58(10):1139–1144. <https://doi.org/10.1007/s11434-013-5711-8>
- Loper E, Bird S (2002) NLTK: the natural language toolkit. In: Proceedings of the ACL-02 workshop on effective tools and methodologies for teaching natural language processing and computational linguistics—vol 1, pp. 63–70. Association for Computational Linguistics, Philadelphia, PA. arXiv:0205028v1. 6 July Accessed 2021
- Mikolov T, Sutskever I, Chen K, Corrado GS, Dean J (2013) Distributed representations of words and phrases and their compositionality. *Adv Neural Inf Process Syst* 26:66
- Mishev K, Gjorgjevikj A, Vodenska I, Chitkushev LT, Trajanov D (2020) Evaluation of Sentiment Analysis in Finance: From Lexicons to Transformers. *IEEE Access* 8:131662–131682. <https://doi.org/10.1109/ACCESS.2020.3009626>. Conference Name: IEEE Access
- Mohammad S (2018) Obtaining reliable human ratings of valence, arousal, and dominance for 20,000 English words. In: Proceedings of the 56th annual meeting of the association for computational linguistics (volume 1: long papers), pp 174–184. Association for Computational Linguistics, Melbourne, Australia. <https://doi.org/10.18653/v1/P18-1017>. Accessed 30 April 2021
- Mohammad SM, Sobhani P, Kiritchenko S (2016) Stance and sentiment in tweets. arXiv:1605.01655 [cs]. Accessed 26 March 2021
- Morstatter F, Pfeffer J, Liu H, Carley KM (2013) Is the sample good enough? Comparing data from Twitter's streaming API with Twitter's firehose. In: 7th International AAAI conference on weblogs and social media, Boston, USA
- Pagolu VS, Reddy KN, Panda G, Majhi B (2016) Sentiment analysis of Twitter data for predicting stock market movements. In: 2016 International conference on signal processing, communication, power and embedded system (SCOPES), pp 1345–1350. <https://doi.org/10.1109/SCOPES.2016.7955659>
- Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, Blondel M, Prettenhofer P, Weiss R, Dubourg V, Vanderplas J, Passos A, Cournapeau D (2011) Scikit-learn: machine learning in python. *J Mach Learn Res* 12:2825–2830
- Pennebaker JW, Boyd RL, Jordan K, Blackburn K (2015) The development and psychometric properties of LIWC2015. Technical report, University of Texas at Austin, Austin, TX (September 2015). Accepted: 2015-09-16T13:00:41Z. <https://repositories.lib.utexas.edu/handle/2152/31333>. Accessed 28 June 2021
- Pennington J, Socher R, Manning CD (2014) GloVe: global vectors for word representation, pp 1532–1543. <https://doi.org/10.3115/v1/D14-1162>. Accessed 28 June 2021
- Raghupathi V, Ren J, Raghupathi W (2020) Studying public perception about vaccination: a sentiment analysis of tweets. *Int J Environ Res Public Health* 17(10):3464. <https://doi.org/10.3390/ijerph17103464>
- Recchia G, Louwse MM (2015) Reproducing affective norms with lexical co-occurrence statistics: predicting valence, arousal, and dominance. *Q J Exp Psychol* 68(8):1584–1598. <https://doi.org/10.1080/17470218.2014.941296>
- Reece AG, Reagan AJ, Lix KLM, Dodds PS, Danforth CM, Langer EJ (2017) Forecasting the onset and course of mental illness with Twitter data. *Sci Rep* 7(1):13006. <https://doi.org/10.1038/s41598-017-12961-9>



- Ryland Williams J, Lessard PR, Desu S, Clark EM, Bagrow JP, Danforth CM, Sheridan Dodds P (2015) Zipf's law holds for phrases, not words. *Sci Rep* 5(1):12209. <https://doi.org/10.1038/srep12209>
- Sanh V, Debut L, Chaumond J, Wolf T (2019) DistilBERT, a distilled version of BERT: smaller, faster, cheaper and lighter. In: Proceedings of the 7th international conference on neural information processing systems. MIT Press. arXiv: 1910.01108. Accessed 28 June 2021
- Serrano MA (2005) Weighted configuration model. In: AIP conference proceedings, vol 776, pp 101–107. AIP, Aveiro (Portugal). <https://doi.org/10.1063/1.1985381>. ISSN: 0094243X. Accessed 11 June 2021
- Serrano MA, Boguna M, Vespignani A (2009) Extracting the multiscale backbone of complex weighted networks. *Proc Natl Acad Sci* 106(16):6483–6488. <https://doi.org/10.1073/pnas.0808904106>
- Shaoul C, Westbury C (2006) Word frequency effects in high-dimensional co-occurrence models: a new approach. *Behav Res Methods* 38(2):190–195. <https://doi.org/10.3758/BF03192768>
- Shivaprasad TK, Shetty J (2017) Sentiment analysis of product reviews: a review. In: 2017 International conference on inventive communication and computational technologies (ICICCT), pp 298–301. <https://doi.org/10.1109/ICICCT.2017.7975207>
- Slater PB (2009) A two-stage algorithm for extracting the multiscale backbone of complex weighted networks. *Proc Natl Acad Sci* 106(26):66. <https://doi.org/10.1073/pnas.0904725106>
- Smailovic J, Grcar M, Lavrac N, Znidarsic M (2013) Predictive sentiment analysis of tweets: a stock market application. In: Holzinger A, Pasi G (eds) Human–computer interaction and knowledge discovery in complex, unstructured, big data. Lecture notes in computer science. Springer, Berlin, pp 77–88. [https://doi.org/10.1007/978-3-642-39146-0\\_8](https://doi.org/10.1007/978-3-642-39146-0_8)
- Stupinski AM, Alshaabi T, Arnold MV, Adams JL, Minot JR, Price M, Dodds PS, Danforth CM (2022) Quantifying language changes surrounding mental health on Twitter. *JMIR Mental Health* (in press). <https://doi.org/10.2196/33685>
- Tromble R, Storz A, Stockmann D (2017) We don't know what we don't know: when and how the use of Twitter's public APIs biases scientific inference. SSRN Scholarly Paper ID 3079927, Social Science Research Network, Rochester, NY (2017). <https://doi.org/10.2139/ssrn.3079927>. Accessed 21 Jan 2022
- Turney PD, Littman ML (2002) Unsupervised learning of semantic orientation from a hundred-billion-word corpus. arXiv:cs/0212012. Accessed 27 June 2021
- Turney PD, Littman ML (2003) Measuring praise and criticism: inference of semantic orientation from association. *ACM Trans Inf Syst* 21(4):315–346. <https://doi.org/10.1145/944012.944013>
- Wang X, Zhang C, Ji Y, Sun L, Wu L, Bao Z (2013) A depression detection model based on sentiment analysis in micro-blog social network. In: Li J, Cao L, Wang C, Tan KC, Liu B, Pei J, Tseng VS (eds) Trends and applications in knowledge discovery and data mining. Lecture notes in computer science. Springer, Berlin, pp 201–213. [https://doi.org/10.1007/978-3-642-40319-4\\_18](https://doi.org/10.1007/978-3-642-40319-4_18)
- Wang W, Zhou H, He K, Hopcroft JE (2017) Learning latent topics from the word co-occurrence network. In: Du D, Li L, Zhu E, He K (eds) Theoretical computer science. Communications in computer and information science. Springer, Singapore, pp 18–30. [https://doi.org/10.1007/978-981-10-6893-5\\_2](https://doi.org/10.1007/978-981-10-6893-5_2)
- Warriner AB, Kuperman V, Brysbaert M (2013) Norms of valence, arousal, and dominance for 13,915 English lemmas. *Behav Res Methods* 45(4):1191–1207. <https://doi.org/10.3758/s13428-012-0314-x>
- Westbury C, Keith J, Briesemeister BB, Hofmann MJ, Jacobs AM (2015) Avoid violence, rioting, and outrage; approach celebration, delight, and strength: using large text corpora to compute valence, arousal, and the basic emotions. *Q J Exp Psychol* 68(8):1599–1622. <https://doi.org/10.1080/17470218.2014.970204>
- Williams JR, Bagrow JP, Danforth CM, Dodds PS (2015) Text mixing shapes the anatomy of rank-frequency distributions. *Phys Rev E* 91(5):052811. <https://doi.org/10.1103/PhysRevE.91.052811>
- Wolfsfeld G, Segev E, Sheaffer T (2013) Social Media and the Arab Spring: politics comes first. *Int J Press Polit* 18(2):115–137. <https://doi.org/10.1177/1940161212471716>
- Wu HH, Gallagher RJ, Alshaabi T, Adams JL, Minot JR, Arnold MV, Welles BF, Harp R, Dodds PS, Danforth CM (2021) Say their names: resurgence in the collective attention toward black victims of fatal police violence following the death of George Floyd. arXiv:2106.10281 [physics]. Accessed 30 June 2021

## Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Submit your manuscript to a SpringerOpen® journal and benefit from:

- Convenient online submission
- Rigorous peer review
- Open access: articles freely available online
- High visibility within the field
- Retaining the copyright to your article

---

Submit your next manuscript at ► [springeropen.com](https://www.springeropen.com)

---