

Tweether: A Visualization Tool Displaying Correlation of Weather to Tweets

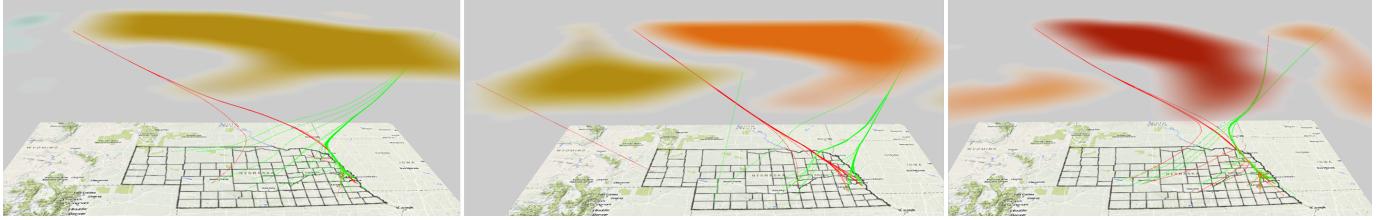


Fig. 1. Tweether showing the real time mood of Nebraska correlated to the weather patterns over three time steps in the morning, afternoon, and night, from left to right, respectively.

Abstract—As the generation of social media we can instantly express how our day is going; however, unknowingly the weather can play a key role in how we are feeling. The weather dictates our lives regardless of what may be happening. The relationship between weather and mood has been immensely studied to show that the weather does play a major factor regarding our emotions. However, how much weather affects us and how we display the relationship using social media remain interesting questions. Based on the natural correlation between weather and mood we propose *Tweether*, a real-time weather and tweet visualization, to see how Twitter users are feeling. This visualization displays a current reflection of emotions in a set of select geographic regions and also predicts possible emotions in these regions in response to the weather forecast. The visualization uses multiple layers to show the connection between locations, weather, and emotions. By aggregating multiple users with similar emotions, we create an aesthetic design that is relatively free of visual clutter and is simple to understand in a 3D manner.

Index Terms—Time series, weather, clustering, sentiment classification, prediction, line bundling, correlation

1 INTRODUCTION

Weather affects our daily lives, from what we wear, what activities we do, what type of transportation we use, what we eat, or even how we feel. With the increasing accuracy of weather forecasts, people can gain an idea on the type of weather they can expect for upcoming days. Activities are usually planned according to the weather outside (e.g., weddings) and alternative plans must be made in case of inclement weather. How people dress is also affected by weather; when the temperature drops people need to wear coats to stay warm. The economy is also greatly affected by the weather. Certain weather conditions can lower crop yield and cause higher prices in stores. Disastrous weather phenomena such as hurricanes, tornadoes, or even floods can cause devastation in communities resulting in homelessness, death, and destruction. Inclement weather can also cause delays in transportation on roads or via flights. We can also choose to ride our bike to work instead of driving the car if the temperature is warm enough. One thing that is an effect of all these items is how we feel.

- Are you sad that you cannot enjoy the outdoors due to rain?
- Do you love that it's raining so you can bundle up and read your favorite book?
- Do you love the snow because it's close to Christmas?
- Do you hate the winter because you want it to be spring?

These feelings are all brought out by the weather outside. One person can feel positive about a certain type of weather and one person can feel negative. In this work, we showcase a novel tool, named *Tweether*, a visualization of real-time Twitter and weather data to show the

feelings of current users and how their emotions could fluctuate. Having the weather forecast in the future, the emotions in current regions can be predicted.

We see if the weather has any correlation to the majority of the population and we try to predict the future feelings of given weather. For example, when thinking about warm and sunny weather we naturally assume the majority of the population will be happier in comparison to dreary or cold temperatures. We plan to examine if the majority of the population follows such patterns. We will also examine how the overall sentiment changes when we filter tweets so only tweets in regards to weather are displayed.

It is not enough to just determine the correlation between emotion and weather, but a novel visualization is necessary. Our work showcases a 3D map which highlights select clusters of weather. The correlation of tweets to weather is represented by a graph. We use line bundling to visualize the graph to reduce visual clutter. Introducing a clear relationship between weather and tweets, the design presents a natural manner of representing correlation.

2 RELATED WORK

Visualizations correlating sentiment and weather are highly sparse, and the presence of live visualizations is also non-existent. However, there are works showing the two portions of this work. Clustering of weather data has been done many times in the past. There is also a vast number of visualizations which indicate the sentiment of different locations.

2.1 Psychological Studies

The natural correlation of weather with emotion has been studied profusely [1, 4, 7, 10, 16]. With various factors among the different research, the conclusions attained were broad. In general, humidity, sunshine, and temperature have the greatest effect on mood [1].

In some research on weather and mood, correlations have been debunked. There is no consistency due to seasons and time spent outside [4]. Having certain emotions regarding the season has strong

links to seasonal affective disorder (SAD), where people are depressed in regards to changes of the seasons, which usually occurs during the winter. However, most psychologists believe that the weather has an impact on psychological intentions [7]. When observing serotonin levels in regards to sunshine there were strong relationships to being happy [10]. It has been found that weather may not play a big role in the positive attitude, but the negative attitude can have a correlation with weather [16].

2.2 Social Media, Weather, and Emotions

When dealing with the correlation of weather with emotion the research is fairly sparse. Few works present the use of Twitter data for their social media feed and some form of weather data. However in comparison to our work the following research used past data and developed a 2D graph implementation for visualization.

Work has been done using two to four years of Twitter data and correlating it with meteorological data from NOAA [7, 14]. Using urban areas in the United States as the area of interest, the tweets are passed to a sentiment analyzer that has a multi-level process. The researchers first determine keywords which are identified from public events (e.g., entertainment or natural disasters), they then identify the mood state, and finally assign sentiment scores. To correlate the weather with the tweets they use a Generalized Mixed Model to display the non-linear relationship between emotion and weather. Using multiple variables for weather (temperature, temperature change, precipitation, snow depth, wind speed, solar energy, and hail), they determine the connection to hostility-anger, depression-dejection, fatigue-inertia, and sleepiness-freshness. Their results indicate that the warmer temperatures create an angrier atmosphere, lower depression, and less sleepiness, and they determine the influence of temperature to mood is trivial. Their visualization is limited to graphs [7]. Other than using urban areas in USA the researchers see relationships between temperature, humidity, and atmospheric pressure for tweets in the United States and weather data from Weather Underground. Using Linguistic Inquiry and Word Count for sentiment classification they see a pattern with temperature and emotion of every state in the United States. Using regression analysis they find that the warmer states have a happier mood than the colder states. Their visualization is limited to a bar chart.

These works are limited in visualization and usability study. Using past data is useful for our training and testing model; however, having a live view of what Twitter users feel is what we aim for in this work.

2.3 Sentiment Analysis in Social Media

Sentiment analysis has been studied vastly. There are various methods to detect the sentiment of a sentence. However, in regards to tweets sentences may be incomplete, because a tweet is limited to 140 characters and the need to express oneself is limited to short meaningful phrases. We find abbreviations, neologisms, acronyms, hashtags, emoticons, and URL's throughout most tweets.

Certain features need to be extracted and some need to be filtered out. Filtering of URL's, usernames, Twitter special words, and emoticons may be needed in certain scenarios [19]. Stop words (i.e., a, an, the, etc.) are also removed due to not adding any extra sentiment information. For classifying the tweets, a number of different methods are used, and the most prevalent one is the Naive Bayes classifier [19]. Emoticons are used as basis for sentiment classification for classifying tweets as positive or negative for the training purposes [14, 19].

Using emoticons for classifying training data is novel; however, most tweets gathered in the live feed have a very low count of emoticons. Thus, emoticons are not employed in our method. Filtering of URL's and usernames is used for our method. We also add the filtering process to drop phrases beginning with hashtags.

2.4 Visualizations

2.4.1 Clustering

Although clustering of data has been extensively studied, it remains a non-trivial task to deal with temporal and time series data. Weather data needs to be clustered based on values, proximities, and changes

throughout the given time span. There have been various visualization techniques for time series data. Visualizing time series data using spirals for large data sets can better identify periodic structures in data [19]. Using wavelet to transform data along a multi-resolution temporal representation to find clusters with similar trends is a useful method for exploring data in a time series fashion [26]. Applying smooth data histograms for visualizing clusters in self-organizing maps is a simple method for 2D data sets [20].

The mass majority of clustering visualizations uses k-means clustering on the basis of their algorithms [17, 25]. We choose to follow this pattern as well, as K-means is a widely chosen algorithm that gives reasonably good results [25, 12].

2.4.2 Bundling

The correlation between multiple entities (e.g., various weather and emotion patterns in our study) can be fundamentally represented as a graph. However, graph visualization remains a challenging task. Severe visual clutter can be easily incurred if we directly draw all edges as straight lines, even with some optimization techniques, such as force-directed placement of vertices or clustering of vertices [13].

To address this issue, Holen [8] proposed a concept of *edge bundling* that groups the related edges of a hierarchical graph together as a set of smooth curved bundles, and thus can significantly reduce visual clutter. Holen et al. [9] extended the original edge bundling method and presented force directed edge bundling (FDEB) for a general graph without hierarchy. Other researchers have also made similar efforts to generalize edge bundling [3, 24, 5, 6]. Few efforts have been dedicated to creating extensions in 3D space. Lambert et al. [15] presented a 3D edge bundling to visualize geographical networks on the Earth surface. Böttger et al. [2] presented mean-shift edge bundling to visualize 3D functional connectivity across the cortical regions of the brain. Their method combines FDEB and kernel density estimation edge bundling (KDEEB) [11] with an improved numerical stability.

2.5 Predictions

Using the past Twitter data set, it is necessary to predict the mood for the next days. Predicting the stock market based on Twitter mood has been studied [1]. The notion that the mood of Twitter users can correlate to stocks immediately is not present; however, the mood is reflected when a few days have passed. Because the general public has strong connections with the outcome of a man-made entity, there will be some form of correlation available. In our case, however, the weather is not a man-made entity. Thus, finding a correlation between weather and mood, and then predicting the mood for the future can show zero correlations. We will address this idea later in this paper.

3 APPROACH

Implementing Tweether has multiple steps which are composed of individual and interconnected components. We use both the weather data and the Twitter data in our study (Section 3.1). Based on the clustering result of the weather data (Section 3.2) and the sentiment classification of each tweet (Section 3.3), we correlate these two entities (Section 3.4). We also predict the sentiments for future times (Section 3.5).

Tweether in its simplest form takes tweets and assigns a sentiment value which is then correlated to the nearest weather cluster. Each tweet is aligned to the map according to its geographic location. The current hour visualization and the prediction visualization have the same user interface. We represent the derived correlation as a graph, and embrace the natural link of weather up in the sky to the tweets down on the ground to implement our visualization (Section 3.6). After we implement each of the necessary steps we can clearly see how the visualization adopts a layered design which effectively highlights the correlation of weather and emotion in space-time. The major steps of Tweether are illustrated in Figure 2.

3.1 Data Sets

We use two main data sets with respect to weather and Twitter in this work. The weather data set is generated from a climate simulation

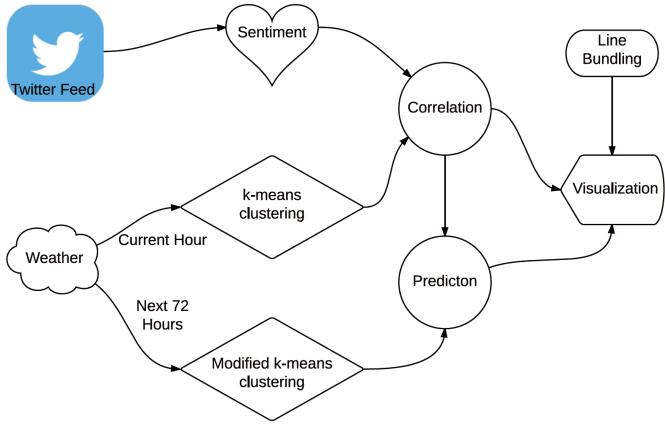


Fig. 2. The major steps of Tweether.

using Weather Research and Forecasting (WRF) Model [18]. The data provides hourly forecasts with up to 72 hours in the future. Each WRF file contains multiple variables in regards to weather (i.e., temperature, precipitation, wind speed, etc.). For this visualization, we choose to focus on the surface skin temperature (TSK) variable. Each TSK file is represented via a 2D array covering a regular geographic region. In this work, we use a WRF data that geographically corresponds to the state of Nebraska.

The Twitter data set contains the live data feed from Twitter users throughout Nebraska. The fetched Twitter data is synchronized with the weather data. Only users that have opted-in to turn on the feature of Tweeting With Location are selected. In addition, because Nebraska has a fairly low population with most of the land being barren, only the most populous cities are chosen. These cities include Omaha, Lincoln, Grand Island, Kearney, Fremont, North Platte, Norfolk, Columbus, and Scottsbluff. We use a geographic filtering process to select these cities. The Twitter data is stored in JSON format where we need to extract the coordinates of each tweet and the tweet itself. Due to some cities being on the border of Nebraska (such as Omaha), the Twitter data needs to have a second filter which removes any tweets that do not have any relation with Nebraska.

3.2 Weather Clustering

We use clustering to extract different weather patterns from the WRF data and identify their geographic coverage. The clustering of weather differs depending on if looking at the current hour or the predicted values for the next 72 hours. For the current hour of weather, we use the k-means clustering algorithm. For the forecasted weather, we use a modified k-means algorithm.

We use the k-means clustering algorithm to partition the 2D array of each time step of TSK into a set of clusters. Some clusters can be dispersed, resulting in random patterns or outliers. These outliers are removed because we assume that moods are affected mostly by comparably dominant weather patterns, and the outliers are small in space and can be changed dynamically in time. To remove outliers, we use a filtering process based on the number of data points in each cluster. If there exists a cluster which has less than one percent of the overall number of clustered elements, this cluster is removed and the data points which belong to a cluster are clustered again. This process is continued until there is no cluster which has less than one percent of the overall cluster count. For the forecasted weather, we use a modified k-means algorithm, where we use the difference between two successive hours for the similarity measures.

To smooth any randomness in the cluster values of the 2D array, we use a low-pass filter where the cluster value of an element is determined by the values of surrounding elements. If at least 5 out of 8 neighbors have the same value, the cluster values of the element is kept; otherwise, the value is removed. We found that repeating this operation 4 times can give us sufficient smooth outcome amongst the

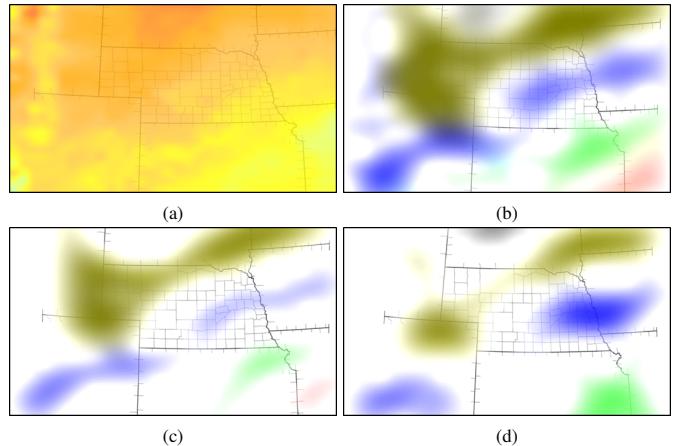


Fig. 3. (a) The geographic distribution of TSK variable at a time step. (b) The k-means result with outliers. (c) The comparably dominant weather patterns. (d) The clustering result of forecasted weather using the modified k-means clustering. The color values in (a) are mapped to TSK. The color values in (b)-(d) are used to distinguish different clusters.

majority of 72 hours. Figure 3 (a) shows the distribution of TSK at a certain time step. A direct use of k-means can generate many dispersed small regions, as shown in Figure 3 (b). Our method can clearly extract the dominant TSK patterns, as shown in Figure 3 (c). Figure 3 (d) shows the result of forecasted weather.

3.3 Sentiment Classification

Each tweet can contain different attributes other than plain sentences. Due to each tweet being limited to 140 characters, the majority of users tend to use abbreviations, neologisms (e.g., noob, troll), acronyms hashtags, emoticons, or URL's. Abbreviations, acronyms, and neologisms are taken into account for training our classifier. However, a few items are filtered from certain tweets. The filtering process removes emoticons, URL's, usernames, and hashtags. In some situations, it is known that hashtags can provide instant insight as to what the users are feeling [14]. However, most hashtags that we encountered contain meaningless text or sentences for tags instead of keywords.

We use the Naive Bayes classifier [19] to determine the sentiment of tweets. Robert Plutchik's theory [22] states that there are eight basic emotions within two categories:

- negative - fear, anger, sadness, depression, disgust
- positive - joy, trust, anticipation, surprise

These emotions are the basic training portion of the classification of tweets. The synonyms for each category are taken into account and this sets up the basic foundation for the tweet classifier.

Other than acronyms we also need to take into account profanity. The use of profanity in social media is very high and it may lead to a positive or negative emotion depending on the situation. To take into account how profanity is used in sentences, we fetched tweets to explore the usage of these words. We found that using these tweets to train the classifier gave a high accuracy rate in regards to profanity. In the beginning, we tried to remove any tweet with profanity, which however drastically lowered our tweet count. We then tried to remove the occurrences of profanity in the tweet and used the remaining words as a judge of emotion, which however only worked in a few cases. For the majority of these tweets, we believe that profanity gave insight to negative moods, and thus decided to take into account profanity.

Due to tweets using abbreviations and incomplete sentences, sentiment calculation is a non-trivial task. The classifier is trained using around 10,000 tweets, where each tweet is given a positive or negative score, and there were rare occurrences of duplicate tweets. There is an equal portion of positive and negative tweets in regards to words where sentiment could go either way. Figure 4 shows the sentiment classification results of two time steps.

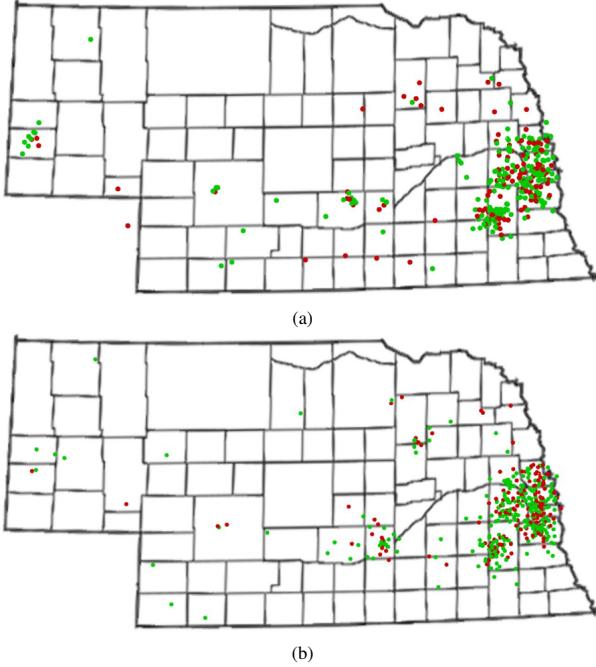


Fig. 4. The sentiment classification results and their geographic distribution of two time steps over Nebraska, where a green point corresponds to a tweet with positive sentiment and a red point corresponds to a tweet with negative sentiment.

3.4 Correlation

We investigate the correlation between the patterns generated from the weather clustering and the sentiment classification. These patterns are characterized with geographic distributions. Figure 5 (left image) illustrates an example of two tweets, T_1 and T_2 , and four weather clusters, W_1 , W_2 , W_3 and W_4 . It is intuitive that the sentiment derived from a tweet is mostly affected by its overlapped weather cluster. We call such a cluster as the *primary* cluster of a tweet. If there is a naturally geographic overlap between a tweet and a weather cluster, the mapping is pure, such as T_1 and W_1 in Figure 5. In situations where there is no direct overlap for a tweet to any of the weather clusters the nearest cluster is used, such as T_2 and W_1 in Figure 5.

Other than the natural link between the primary cluster and a tweet, we explore the similarity of connections of the tweet to other clusters. This is because the sentiment of the tweet can be also affected by its vicinal weather clusters. Therefore, disregarding the primary cluster, we quantify the correlation of the tweet to the other clusters to indicate what other clusters the tweet could map to. In particular, we use the location of a tweet and the TSK value at the location to determine the correlation value to the points of a weather cluster using the Pearson product-moment correlation coefficient:

$$\rho_{X,Y} = \frac{\text{cov}(X,Y)}{\sigma_X \sigma_Y}, \quad (1)$$

where X is the location and the TSK value of the tweet, Y is the locations and the TSK values of all points of a weather cluster, cov is the covariance, and σ_X (σ_Y) is the standard deviation of X (Y). The correlation value of $\rho_{X,Y}$ ranges from -1 to 1. If the value is 1 (-1), it indicates a perfect positive (negative) linear relationship between X and Y . If the value is 0, it means that there is no linear relationship between X and Y .

This correlation is computed for the current hour. For each tweet, we use the top two values of correlation to other clusters which we deem the *secondary* and *tertiary* clusters, respectively. The secondary and tertiary mapping have the same sentiment since the best fit is a cluster close by with a similar TSK value. Figure 5 illustrates the

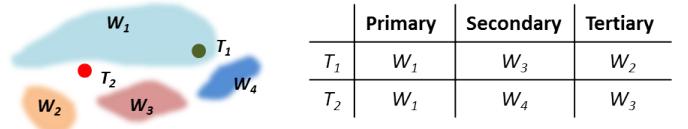


Fig. 5. The correlation between the tweets, T_1 and T_2 , and the weather clusters, W_1 , W_2 , W_3 and W_4 . The left image shows the geographic distribution of the tweets and the weather clusters. The right table shows the primary, secondary, and tertiary weather clusters of T_1 and T_2 .

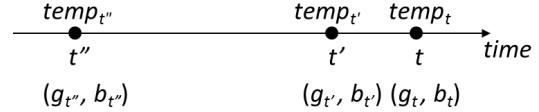


Fig. 6. The prediction model.

primary, secondary and tertiary clusters of T_1 and T_2 respectively.

3.5 Prediction

Predicting the future is mostly based on facts. We have at our disposal the current mood and the current temperature, all of the previous days tweets and the temperatures for each hour, and the predicted temperature for each hour for the next two days. Using these facts, we try to determine what the sentiment at each tweet location will be for the next two days.

Determining the mood of the current locations up to 72 hours in the future is non-trivial. Our prediction technique is based on the current hour and the previous day. We choose not to use data from earlier times because the trends today are definitely not the same a year ago, let alone a month ago. In addition to this, we should state that the long-term weather is relatively unpredictable for the state of Nebraska.

We start with a simple method by only comparing the TSK difference. If the current temperature is closer to the predicted temperature, we use the sentiment for the current hour to show the prediction. If the sentiment of the hour we are trying to pick has a closer temperature to the same hour of the previous day, we use the sentiment from the previous day.

However, this simple method neglects the time. Alternatively, we can consider the time difference and state that if the current hour is closer to the predicted hour we will place a higher weight on using the current hours values; when the current hour goes further from the predicted hour, we can place more weight on the previous day value. Using this method we take the sentiment based on the percentage of the weight in terms of time.

The final method we use is a combination of above methods. As shown in Figure 6, we want to predict the sentiments at a future hour t , and we have the following information:

- t' is the current hour;
- t'' is the same hour as the predicted hour on the previous day, i.e., $t'' = t - 24$;
- $\text{temp}_{t'}$, $\text{temp}_{t'}$ and $\text{temp}_{t''}$ are the TSK values at t , t' , and t'' , respectively;
- $g_{t'} (g_{t''})$ is the total number of tweets with positive emotions at t' (t'') correlated to the clusters of $\text{temp}_{t'} (\text{temp}_{t''})$;
- $b_{t'} (b_{t''})$ is the total number of tweets with negative emotions at t' (t'') correlated to the clusters of $\text{temp}_{t'} (\text{temp}_{t''})$.

Based on this information, we first compute the differences between the TSK values:

$$p_{t'} = |\text{temp}_{t'} - \text{temp}_{t''}|; \quad p_{t''} = |\text{temp}_{t'} - \text{temp}_{t''}| \quad (2)$$

According to the difference values, $p_{t'}$ and $p_{t''}$, we determine the weights of the two sentiment sets:

$$w_{t'} = \frac{p_{t''}}{p_{t'} + p_{t''}}; \quad w_{t''} = \frac{p_{t'}}{p_{t'} + p_{t''}} \quad (3)$$

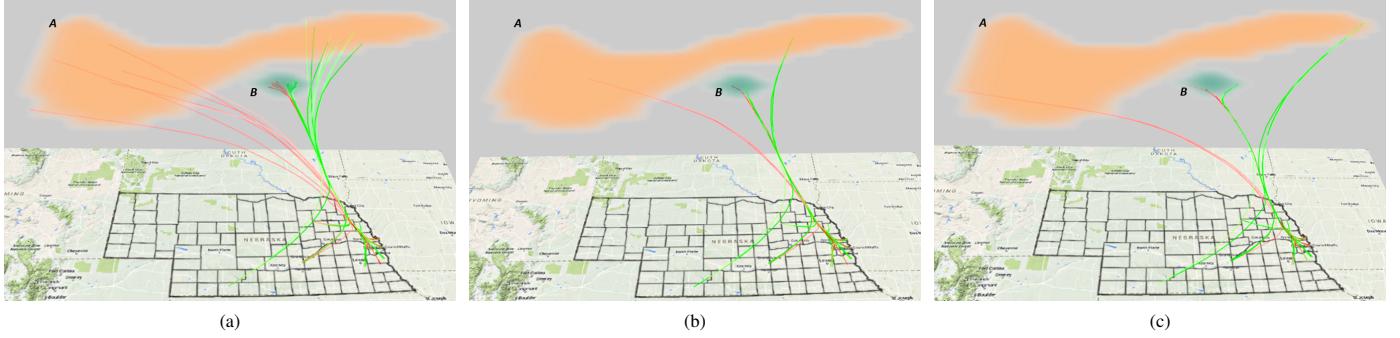


Fig. 7. Determining the best positions for indicating negative and positive tweets in the cloud representations of clusters. (a) Use the different point of a cluster for each edge. It can easily generate visual clutter. (b) Use the center points of the two halves of a cluster. This method is effective for a large cluster A, but makes it difficult to distinguish lines for a small cluster B. (c) Use the top rightmost point and the bottom leftmost point of a cluster. This method is effective for both the large and small clusters, A and B, respectively.

Finally, we estimate the amount of tweets with positive and negative emotions, g_t and b_t , at t with respect to temp_t as:

$$g_t = g_t' w_{t'} + g_t'' w_{t''}; \quad b_t = b_t' w_{t'} + b_t'' w_{t''} \quad (4)$$

The method accounts directly for both the time and the temperature. We will demonstrate the details regarding how well these three methods in Section 4.2.

3.6 Visualization

We design a novel 3D visualization to picture the analytics of weather clustering, sentiment classification, and the corresponding correlation and prediction. The visualization is comprised of multiple layers, and there is a natural visual connection between the layers such that the weather above affects the people below. The layers are connected via line bundles to dictate how the majority of the Twitter users in Nebraska are feeling. This visualization provides a novel and more intuitive way to present the data.

3.6.1 Layers

The visualization is represented by three main layers. The base layer contains the map of Nebraska and neighboring states. We have chosen to show the counties in Nebraska to gain insight on the population distribution in the state, and choose to keep the other states plain to indicate emphasis of correlation to Nebraska. Using the WRF data set for the longitude and latitude values for the map, we compute the values at intermediate locations via interpolation. We use the longitude and latitude location of each tweet to place it on the map.

The topmost layer shows the weather clusters by lifting each cluster vertically from its geographic region on the ground of map. We visualize the weather clusters using a cloud representation (Section 3.6.2). The middle layer contains the edges which starts from the Twitter user location and ends at the weather cluster they correspond to above, according to the correlation results. We use the edge bundling method to succinctly display the complex edges while suppressing visual clutter (Section 3.6.3).

3.6.2 Cloud Representation

In our visualization, the weather clusters appear as clouds. Each cluster is blurred at the boundaries to make the graphical interface cleaner and aesthetically pleasing. The color of each weather cluster is mapped from indigo to dark orange with respect to lower and higher TSK values. Clusters which are not connected of the same color indicate the same conditions (i.e., similar temperature) as other locations of the state.

Because each cluster can be possibly correlated to the tweets with positive and negative emotions, we need to locate two types of edge points on the cluster for the tweets to link up to. We explore three designs on how the points should be picked. First, we assign half of one cluster to the positive emotions and the other half to the negative

emotions. This visualization can easily generate a cluttered output, as shown in Figure 7 (a). Second, we split a cluster in half and use the center points of the two halves. This method is effective until we come into situations where the cluster is represented by a very small amount of data points, causing a clutter visualization again, as shown in Figure 7 (b). Finally, we map the tweets with positive emotions to the top rightmost point of a cluster, and map the tweets with negative emotions to the bottom leftmost point of the cluster. This method can significantly reduce visual clutter and help in situations where the cluster size is fairly small, as shown in Figure 7 (c).

3.6.3 Edge Bundling

Our design naturally forms a graph to represent the correlation between weather and tweets. However, visual clutter can easily occur if we directly display each edge as a straight line, as shown in Figure 8 (a). To address this issue, we adopt FDEB [9] to visualize the graph of correlation. We bundle the related edges with high compatibilities, and iteratively subdivide the edges to generate smooth curves with coherent shapes. This method can effectively reduce visual clutter in 3D.

The original paper of FDEB [9] proposed four criteria, angle, scale, position, and visibility, for edge compatibility measures. In our design, as the edges are mostly oriented along the vertical direction between cloud and ground, the variation in angle or scale is relatively marginal compared to a general graph. In addition, as we display the edges in 3D, the visibility of an edge can be changed from different views. Therefore, only position (i.e., distance between midpoints of edges) is considered for computing edge compatibility in our design.

Similar to FDEB, we use an iterative simulation to refine the bundling. The simulation starts with P_0 subdivision points for each edge, and then performs C simulation cycles. During each cycle, a specific number of iteration steps I is conducted to move the subdivision points to reach an equilibrium between forces. The number of iteration steps during the first cycle is I_0 . After performing a cycle, the number of subdivision points is doubled to smoothen the edges, and the number of iteration steps I is decreased by a factor R . We found that a configuration of $P_0 = 1$, $C = 6$, $I_0 = 50$, and $R = \frac{2}{3}$ leads to appropriate results in our design. Figure 8 (b) shows the edge bundling result of the same graph as in (a).

Each bundle is composed of multiple lines where each line represents one tweet. Every line corresponds to a positive or negative sentiment. The sentiment is represented in two ways. First, from the weather cluster perspective, we indicate the sentiment according to which corner of a cluster the bundle is linked to. In this way, we can easily examine the sentiments affected by a certain cluster, as shown in the two highlighted boxes as in Figure 9. Second, from the tweet perspective, we associate the positive sentiments to the green bundles and the negative sentiments to the red bundles. In this way, we can intuitively convey the sentiments distribution among the tweets, as shown in the highlighted box B in Figure 9.

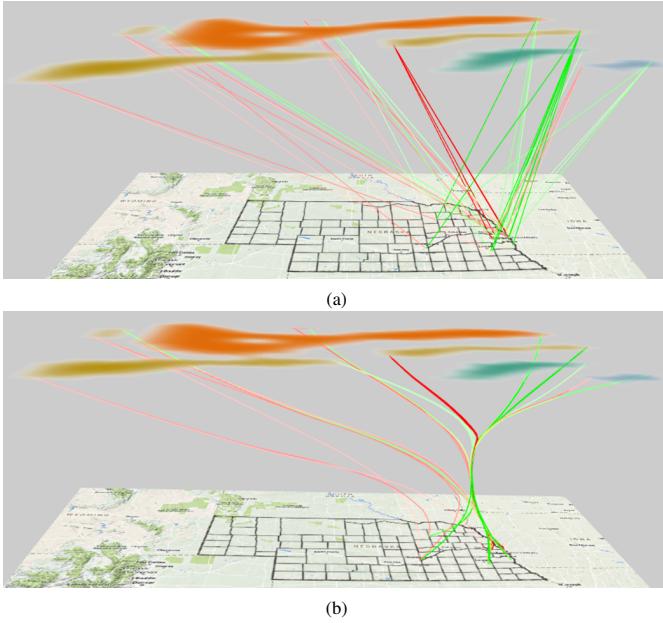


Fig. 8. (a) A direct visualization of the correlation graph. (b) The edge bundling result.

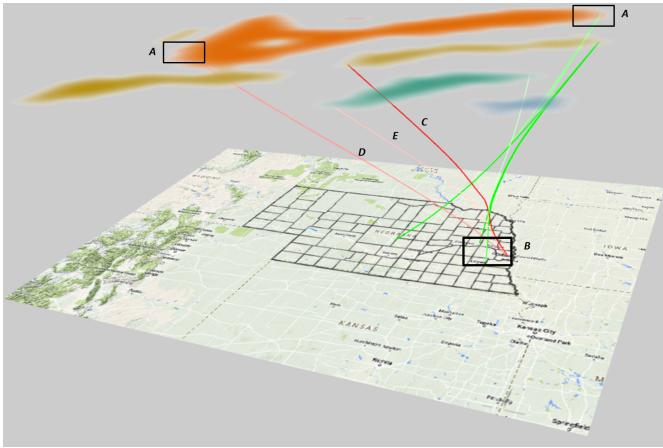


Fig. 9. The detailed configuration of line bundles.

The opacity of each bundle represents the intensity of the correlation between the cluster and the tweets. The bolder the line the stronger the relationship between the two points is. The primary cluster has a stronger hue in comparison to the secondary and tertiary clusters. These links associated with the secondary and tertiary clusters have a staggered decrease of intensity of color. Thus, we state that the more prominent the link between the tweet and the cluster the higher the opacity of the line. The lines C, D, and E in Figure 9 show an example of the opacity values of lines associated with the primary, secondary, and tertiary clusters, respectively.

4 RESULTS

Our Tweether system succinctly shows regional weather characteristics and local sentiment classification using a layered design. The edge bundling method can create smooth curves with coherent shapes and avoid visual clutter. The resulting visualization is visually aesthetic as well as directs a user's focus of attention to salient relationship between weather and mood over geographic locations and time. We implemented Tweether using JavaScript and WebGL such that the system can be viewed from any devices supporting HTML5 and WebGL.

4.1 Tweet Weather Correlation

We use the weather and Twitter data from two back to back weekends (from March 6th to March 8th, and from March 13rd to March 15th) to see if there is any relationship between the temperature and the overall mood of people. We are lucky that the days we chose show warmer changes with certain fluctuations in weather. Being on the brink of spring, we predict that the overall tweets for the warmer weekend will have a more positive sentiment in comparison to the colder weekend. This weather pattern also can show any potential correlation for seasonal affective disorder (SAD) [4].

4.1.1 All Tweets

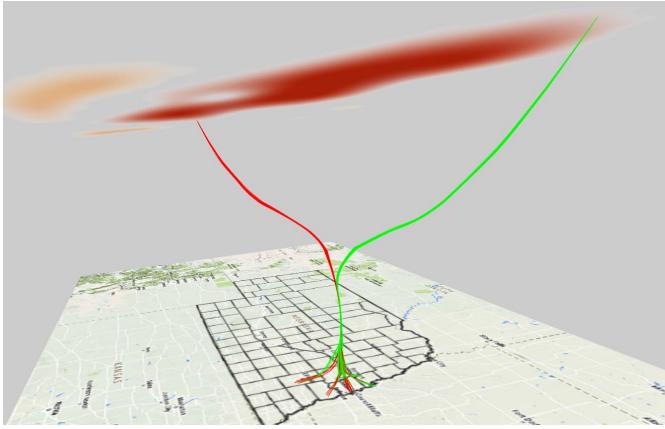
In the first study, we use all collected tweets whether a tweet contains weather-related terms or not. Figure 1 shows the visualization results of three time steps across the state. Of all the cities, we see that Omaha and Lincoln have the most tweets, and thus we use these two cities to show the details of our visualization. We choose to look at times where most of the population is awake (i.e., noon to 11 PM). Figure 10 shows the visualization results at two time steps over the two weekends. The color values from indigo to dark orange are computed according to the local TSK values. From Figure 10 (a), we can see that the tweets have an equal amount of positive and negative sentiments, and they are all mapped to the same primary weather cluster that has the highest local TSK. We note this cluster is the dominant weather pattern nearly covering the entire state of Nebraska. From Figure 10 (b), we can see that the majority of tweets have the positive sentiment within these two cities. They are all mapped to the same primary weather cluster that has the lowest local TSK. In addition, from both the images, we can see that the most red lines (negative sentiment) came from Omaha. By visualizing multiple time steps, we observe that both the temperature and the total number of tweets become higher with an increase of positive sentiment and a decrease of negative sentiment.

To explain our observation, we examine the detailed weather patterns, as shown in Figure 11 (a). Overall it shows a dramatically growing trend of temperature, roughly from 20°F to 80°F, as well as some fluctuations around the early March. From Figure 11 (b) we can see the tweets and their sentiment for select days through the two weeks, where more emphasis is placed on the first week and less on the later. Each day has a positive and negative count, where we clearly see throughout, the pattern that the positive tweets outweigh the negative tweets. Another pattern that we see is that as the temperature increases the number of tweets increases as well. We also observe that more people tweet on Saturday than on Sunday. We think that this may be related to a number of factors: people go to church on Sunday where they choose to be among family, they do work around the house, or they do not have any eventful things which they feel they need to share.

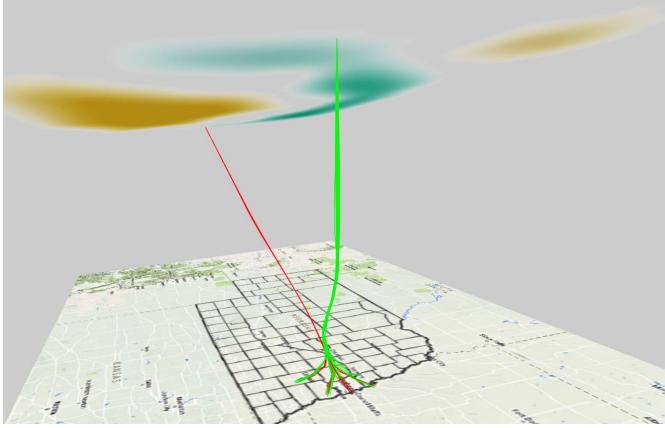
From Figure 11 (a) and (b), we do not see that with the change of temperature there is a clear trend of more positive tweets versus negative tweets. The number of positive tweets is at least more than half the amount of negative tweets. We observe that the percentage of the negative sentiment is higher in the colder temperatures than in the warmer temperatures. For example, the weekend of March 14th has the highest temperatures as shown in Figure 11(a), and the percentage of negative sentiment is lowest as shown in Figure 11(b). However, from the statistics in Figure 11 (b), we cannot be certain if there is a clear correlation. To address this issue, we look at tweets only with words related to weather to see if we find any trends.

4.1.2 Weather Related Tweets

Other than using all the tweets we collected, we filter them for weather-related terms (i.e., snow, sunny, warm, cold, rain, etc.). We want to see if there is a stronger correlation in this situation compared to using all the tweets. As shown in Figure 11(c), we see that the number of tweets drastically reduces to around 160 tweets for one day. We note that there is also a spike of tweets when there is warmer weather (e.g., the weekend of March 14th). This pattern is seen in both sets of tweet data sets. However, we are now able to confirm that the correlation is due to weather. As the weather increased slightly on March



(a)



(b)

Fig. 10. The visualization results at two time steps, (a) 1pm on March 8th and (b) 1pm on March 15th.

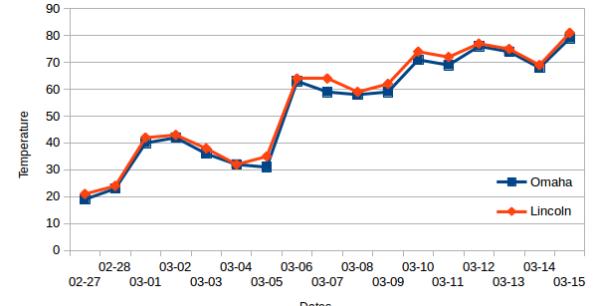
1st we see a spike in amount of tweets overall (Figure 11(b)) and also when the tweets had weather-related words (Figure 11(c)). We can also see that in Figure 11(c) that with the warmer weather the number of tweets in relation to the weather increases by approximately twice the amount. We still see the trend of more positive tweets in comparison to negative tweets when using the filtered data set.

However, this trend can be clearly conveyed in our visualization (Figure 10) without an examination of detailed statistics which can be cumbersome for large tweet data.

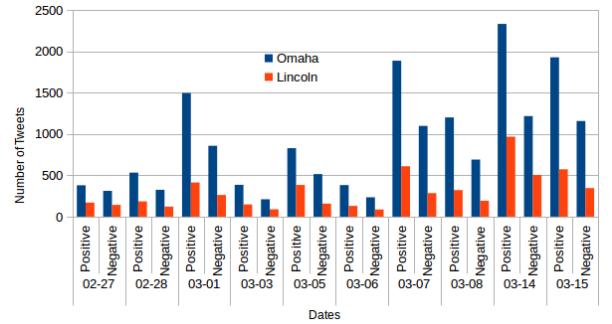
4.2 Prediction

Using the same data set we select a subset of two weekends to determine if the predicted value is close to the outcome. We choose March 6th, 7th and 8th for weekend one and March 13th, 14th and 15th for weekend two. The first weekend has between 10 to 20 degrees difference to the second weekend to take into account any change in weather patterns. We collect our predictions for Saturday and Sunday and compare them to the real sentiment of each hour. Again we focus on the two most populous cities in Nebraska, Lincoln and Omaha.

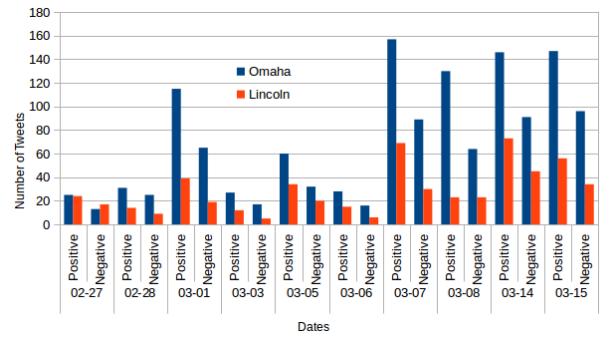
Figure 12 shows the results of the three methods described in Section 3.5 at two time steps. The ground truth is shown in Figure 10. Figure 12 (a) and (b) are generated using the first method where we only consider the temperature difference. As we use the same locations, intuitively the images are close to the ground truth. Figure 12 (c) and (d) are generated using the second method where we only consider the time difference, while Figure 12 (e) and (f) are based on the third method with considering both the temperature and time differences. We note that the thickness of line bundles in (e) and (f) are most similar to the ground truth.



(a)



(b)



(c)

Fig. 11. (a) The temperature of Omaha and Lincoln over two weeks. (b) The sentiment analysis of the overall tweet data set. (c) The sentiment analysis of the weather-related tweets.

Table 1 provides a quantitative measure of errors for the three methods. We compare the number of predicted positive and negative tweets at each time step to the ones of the ground truth. The percentage number is the relative error. We can see that the third method has the least errors for both positive and negative tweets. The second method has the worst error percentage. This is primarily due to the fact that we do not take into account the weather. Because it uses the temperature, the first method has the slight better performance than the second method.

5 DISCUSSION

There are a few aspects in regards to tweets that enforces a limit as to what we can assess. One limitation from tweets is that with a state population as small as Nebraska the chance of a user with their location enabled is relatively low. This limits the amount of tweets we can attain each hour. Tweet length imposes another limitation because the complete expression of a user's feelings can be curbed in some situations. Other than Twitter's limits we need to examine the suppressed feelings of users. As Twitter is a social media hub, the need to express oneself in 140 characters may place a mask to what the user may be feeling. As proposed before there is no distinct way to determine the true feelings of a person [7]. Given this phenomena, we exper-

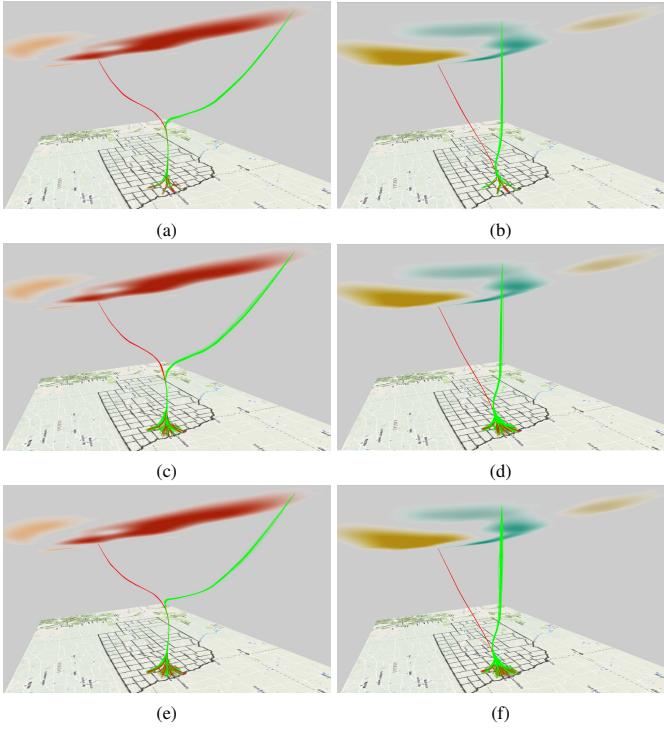


Fig. 12. The results of three prediction methods. (a), (c), and (e) correspond to the predicted results of 1pm on March 8th. (b), (d), and (f) correspond to the predicted results of 1pm on March 15th.

Table 1. The percentage errors of three prediction methods.

	Positive	Negative
Method 1	23%	27%
Method 1	26%	28%
Method 1	23%	26%

mented the usage of only weather related tweets because there is a lower chance of users masking their emotions. Even in this case we saw that the majority of users use their tweets in a positive manner, so to expect a negative tweet is cynical in some sense.

Classifying the sentiment always brings errors. Humans are not known to be the most accurate at identifying the sentiment of written statements [21]. There are always certain words that can make the sentiment have a positive value instead of negative. Tweets are known to have sarcasm in terms of introducing a form of satire or irony [23]. In situations like this there are some misclassifications of sentiment.

We have experienced winter and spring weather in our current study. When we have the change of seasons it is easy to see that there is some correlation with SAD. Like most of the population, the end of winter brings new changes, and thus new feelings. We will study the sentiment changes when summer, fall, or winter approaches. Nebraska is known for its intense heat and harsh winters, but these seasons also entail vacations and family gatherings. Analyzing these aspects can also provide more insight regarding our data sets and other filtering processes that we may need.

6 CONCLUSION

In this paper we have presented a 3D visualization tool, named Tweether, that can assist analysts to explore the correlation of weather to tweets in geospace and time. We extract the dominant patterns from the weather data and classify the sentiments from the synchronized tweets. We identify the relation between weather and tweets using their geospatial relationship and the Pearson product-moment correlation coefficient. We tailor FDEB to our layered visual design and

employ line bundles to show detailed graphs of the connection between the weather clusters and the tweets. With a fine tune on graph vertex placement and edge properties, our line bundling can significantly reduce visual clutter and clearly reveal the sentiment distribution with respect to different weather conditions in 3D. In addition, we develop a prediction model to estimate the future emotions among the tweets according to the forecasted weather. Our Tweether system integrates these data analysis models and interactive visualization, and facilities users to visually examine the connection between weather and mood and identify possible relationship. Although researchers have presented visualizations showcasing social media sentiment or clustering of weather patterns, to the best of our knowledge we have presented here the first application to visualize the correlation between the two.

The future of this research can be expanded immensely. Although we focus on the relation to temperature in this work, the general design philosophy can be applied to precipitation, humidity, wind speed, or any other weather variables. We will conduct research regarding what the overall sentiment of people to rain, storms, and strong wind conditions. For example, there may be a significant correlation between rain amount and negative tweets [7]. We also believe that we can determine if there is a better correlation of weather to other social phenomena or events, such as road accidents, calamities from natural disasters, or even medical episodes like the flu, which can be directly impacted by weather. Given the possible uncertainty when determining if a person's mood is truly affected by weather, we plan to involve other social phenomena or events to obtain more concrete correlation results. This work can be also expanded to other states. The outcomes in more populous states may be different. We are also interested in the states which experience less seasonal changes, and study the corresponding correlation between weather patterns and sentiment.

Other than visualization we would like to improve sentiment classifier by mining large tweet data to gain information regarding new words, such as determining sarcasm, irony, or satire, and find new linguistic patterns. This can also help improve the sentiment prediction.

REFERENCES

- [1] J. Bollen, H. Mao, and X. Zeng. Twitter mood predicts the stock market. *Journal of Computational Science*, 2(1):1–8, 2011.
- [2] J. Böttger, A. Schafer, G. Lohmann, A. Villringer, and D. S. Margulies. Three-dimensional mean-shift edge bundling for the visualization of functional connectivity in the brain. *Visualization and Computer Graphics, IEEE Transactions on*, 20(3):471–480, 2014.
- [3] W. Cui, H. Zhou, H. Qu, P. C. Wong, and X. Li. Geometry-based edge clustering for graph visualization. *Visualization and Computer Graphics, IEEE Transactions on*, 14(6):1277–1284, 2008.
- [4] J. J. Denissen, L. Butalid, L. Penke, and M. A. Van Aken. The effects of weather on daily mood: a multilevel approach. *Emotion*, 8(5):662, 2008.
- [5] O. Ersoy, C. Hurter, F. V. Paulovich, G. Cantareiro, and A. Telea. Skeleton-based edge bundling for graph visualization. *Visualization and Computer Graphics, IEEE Transactions on*, 17(12):2364–2373, 2011.
- [6] E. R. Gansner, Y. Hu, S. North, and C. Scheidegger. Multilevel agglomerative edge bundling for visualizing large graphs. In *Pacific Visualization Symposium (PacificVis), 2011 IEEE*, pages 187–194. IEEE, 2011.
- [7] A. Hannak, E. Anderson, L. F. Barrett, S. Lehmann, A. Mislove, and M. Riedewald. Tweetin' in the rain: Exploring societal-scale effects of weather on mood. In *ICWSM*, 2012.
- [8] D. Holten. Hierarchical edge bundles: Visualization of adjacency relations in hierarchical data. *Visualization and Computer Graphics, IEEE Transactions on*, 12(5):741–748, 2006.
- [9] D. Holten and J. J. Van Wijk. Force-directed edge bundling for graph visualization. In *Computer Graphics Forum*, volume 28, pages 983–990. Wiley Online Library, 2009.
- [10] E. Howarth and M. S. Hoffman. A multidimensional approach to the relationship between mood and weather. *British Journal of Psychology*, 75(1):15–23, 1984.
- [11] C. Hurter, O. Ersoy, and A. Telea. Graph bundling by kernel density estimation. In *Computer Graphics Forum*, volume 31, pages 865–874. Wiley Online Library, 2012.
- [12] A. K. Jain. Data clustering: 50 years beyond k-means. *Pattern recognition letters*, 31(8):651–666, 2010.

- [13] M. Kaufmann and D. Wagner. *Drawing Graphs: Methods and Models*. Springer, 2001.
- [14] M. C. Keller, B. L. Fredrickson, O. Ybarra, S. Côté, K. Johnson, J. Mikels, A. Conway, and T. Wager. A warm heart and a clear head the contingent effects of weather on mood and cognition. *Psychological Science*, 16(9):724–731, 2005.
- [15] A. Lambert, R. Bourqui, and D. Auber. 3d edge bundling for geographical data visualization. In *Information Visualisation (IV), 2010 14th International Conference*, pages 329–335, July 2010.
- [16] G. Lambert, C. Reid, D. Kaye, G. Jennings, and M. Esler. Effect of sunlight and season on serotonin turnover in the brain. *The Lancet*, 360(9348):1840–1842, 2002.
- [17] J. Li, X. Wang, and E. Hovy. What a nasty day: Exploring mood-weather relationship from twitter. In *Proceedings of the 23rd ACM International Conference on Conference on Information and Knowledge Management*, pages 1309–1318. ACM, 2014.
- [18] J. Michalakes, J. Dudhia, D. Gill, T. Henderson, J. Klemp, W. Skamarock, and W. Wang. The weather research and forecast model: Software architecture and performance. In *Proceedings of the 11th ECMWF Workshop on the Use of High Performance Computing In Meteorology*, pages 25–29, 2004.
- [19] A. Pak and P. Paroubek. Twitter as a corpus for sentiment analysis and opinion mining. In *LREC*, volume 10, pages 1320–1326, 2010.
- [20] E. Pampalk, A. Rauber, and D. Merkl. *Using smoothed data histograms for cluster visualization in self-organizing maps*. Springer, 2002.
- [21] B. Pang, L. Lee, and S. Vaithyanathan. Thumbs up?: Sentiment classification using machine learning techniques. In *Proceedings of the ACL-02 Conference on Empirical Methods in Natural Language Processing - Volume 10*, EMNLP ’02, pages 79–86, Stroudsburg, PA, USA, 2002. Association for Computational Linguistics.
- [22] R. Plutchik. *Emotions and Life: Perspectives from Psychology, Biology, and Evolution*. American Psychological Association (APA), 2002.
- [23] E. Riloff, A. Qadir, P. Surve, L. De Silva, N. Gilbert, and R. Huang. Sarcasm as contrast between a positive sentiment and negative situation. In *EMNLP*, pages 704–714, 2013.
- [24] A. Telea and O. Ersoy. Image-based edge bundles: Simplified visualization of large graphs. In *Computer Graphics Forum*, volume 29, pages 843–852. Wiley Online Library, 2010.
- [25] M. Weber, M. Alexa, and W. Müller. Visualizing time-series on spirals. In *Information Visualization, IEEE Symposium on*, pages 7–7. IEEE Computer Society, 2001.
- [26] J. Woodring and H.-W. Shen. Multiscale time activity data exploration via temporal clustering visualization spreadsheet. *Visualization and Computer Graphics, IEEE Transactions on*, 15(1):123–137, 2009.