Eurographics Conference on Visualization (EuroVis) 2016
K-L. Ma, G. Santucci, and J. J. van Wijk
(Guest Editors)

Volume 35 (2016), Number 3

EmotionFlow: Visual Analysis of Emotion Transitions

Submission ID: 307

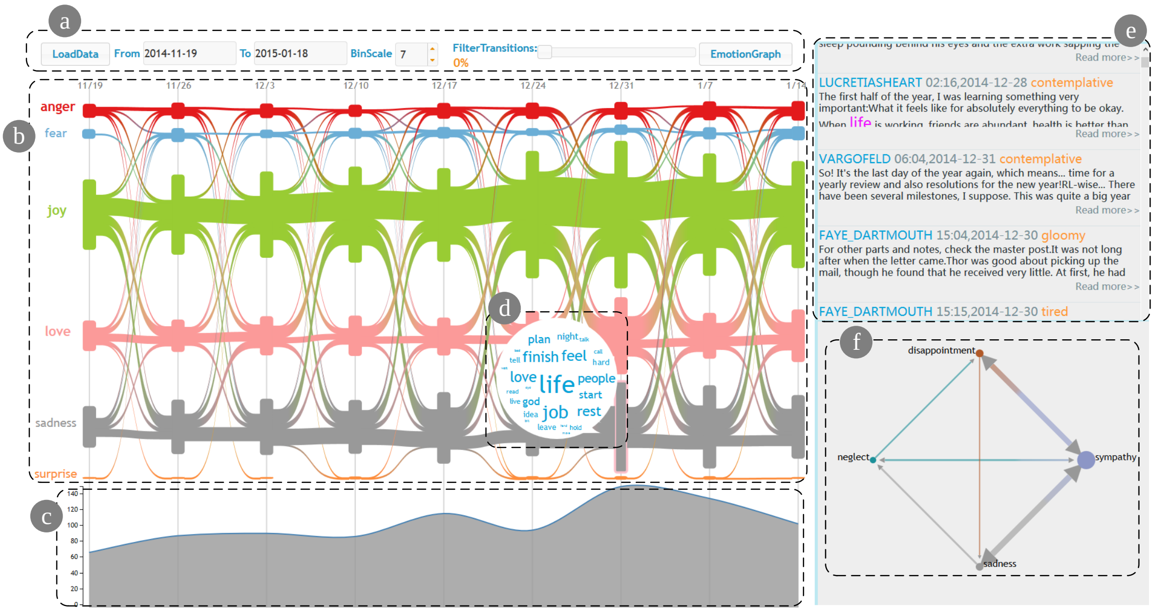


Figure 1: Exploratory analysis of emotion transitions and their triggers in the LiveJournal dataset using EmotionFlow. (a) the control panel. (b) the emotion transition flow with six emotions. (c) the line chart for the amount of one emotion over time. (d) emotion trigger word cloud of one emotion. (e) blog posts and associated emotion labels. (f) the emotion transition graph.

Abstract

Emotional states are not independent, there are transitions and dependencies among them, which are governed by internal, cognitive and external social factors. Describing and modeling emotion transitions can gain a better understanding on human emotions, and provide powerful support on social interactions. Nowadays, a great many of users post online blogs and their emotional labels together, which provide more accurate descriptions of blogger's emotions. We design a timeline-based visual analytic tool, EmotionFlow, for online blogs and their associated emotional label to help users 1) grasp the overall emotions of people, 2) discover emotion transition patterns derived from these emotion data, and 3) find the potential triggers of certain emotions or transitions between them. Case studies are performed to demonstrate the effectiveness and usefulness of EmotionFlow.

Categories and Subject Descriptors (according to ACM CCS): I.3.3 [Computer Graphics]: Picture/Image Generation—Line and curve generation

1. Introduction

Emotions are states of feeling in the sense of an affect, such as anger, joy, and sadness. They are not independent and can transit from one state to another. Those transitions are governed by various factors, such as personal experience or social events [Plu01]. Real-world events have close connection with mood messages expressed by people. Public's emotion transitions can be utilized to detect events, such as major weather phenomena and terror attacks. To monitor people's emotion transition, modeling and displaying the emotion patterns can provide a better understanding of human emotions , and provide powerful support on social interactions For example, individuals could leverage the knowledge in emotional states for self-enhancement [MKK*12], and conditional dependencies can exist among sequential emotional states [SGEMP14].

Recently, social media, such as Twitter and Facebook are getting more popular. As one of the social media's key functions is for users to express their views or opinions, the text posted on social media contains a rich of sentiment information. Therefore, those text corpuses are ideal data sources to extract emotion transitions. So far there have been many efforts in sentiment analysis of text. Most of them focus on automatic detection of emotion in texts [LK06]. Based on the results of sentiment analysis, many visual analytic researches on emotions have been proposed. For example, Zhao et al. provided a multi-dimensional analysis tool by visualizing the personal emotion transitions [ZGWZ14]. However, at the best of our knowledge, few of previous work emphases exploring transition patterns between emotions and discovering possible triggers for them.

In this paper, we present a visual analytic system, EmotionFlow, for progressive exploration in emotion transitions and their possible triggers. We adopt both categorical and dimensional model in psychology to build our emotion model based on emotion data from social media. The emotion data is text corpus annotated with emotion labels which are mostly tagged by the social media users to reveal their emotional states. A hierarchical emotion structure is established by clustering the emotion labels. This structure allows analysis from multi-granularity. We also apply a machine learning method to perform emotion classification for unlabeled blogs. Each user's emotions are then sequentially connected to form a chain. In order to capture both timeline-based and order-based transition patterns, we employ two different approaches to aggregate all users' emotion chains. The timeline-based approach characterizes time-related information from the data, while the order based approach emphases general patterns in emotion transitions.

On top of emotion data processing, we design three main visualization methods for interactive analyzing emotions, transitions, and their possible triggers. Emotion transition flow focuses on presenting time-related transitions between emotions. Emotion transition word cloud reveals the information of personal experience and social events behind the emotion transition. Emotion transition graph is designed for showing general emotion transition patterns.

The main contributions of this paper are:

- A comprehensive model for characterizing emotions of text corpus from social media. Emotions are organized in a hierarchical structure and users' emotion chains can be aggregated either by time or by order.
- Enabling visual analysis in patterns of aggregated emotion transitions. Emotion data are visual encoded in an intuitive manner, and the visualization can interactively reveal the trend of emotion changes according to time or the general transition patterns between emotions.
- Enabling visually discovering possible triggers of aggregated emotion transitions. Key words texts are well organized and emotion transition graph provides supplementary information for understanding emotion transitions.

2. Related Work

Emotion analysis and visualization has been widely investigated in diverse areas. This section reviews a few research areas, including emotional psychology, visual analysis of emotions, and visual analysis of temporal event sequences.

2.1. Emotional Psychology

Categorical and dimensional model [CMK13] are two commonly used emotional models. Parrott's tree structured emotional model [Par01] is a typical categorical model, where emotions are ordinarily classified into three to eleven categories. In contrast, dimensional models are often characterized emotions as a small number of basic dimensions, such as the PAD (Pleasure, Arousal, and Dominance) model proposed by Mehrabian [Meh80]. Typically, categorical model is more understandable, because people tend to use affective labels to refer to emotions. On the other hand, dimensional model emphasises the fundamental components in understanding emotions. In this paper, we leverage both of the two models to cluster various emotional labels. Emotions do not happen in isolation, and generally transfer from one state to another. These transitions usually connect with other actions or factors to form a loosely connected chain. For example, a feeling state can lead to a certain action and the action may influence the emotion in return, as described in the emotion feedback loop [Plu01]. Modeling the transitions and dependencies among emotional states can gain insights into the nature of human emotions and improve the sentimental analysis efforts. In Psychology, Eaton et al. [EF01] studied the stability of emotional experience and measured the rate of change in intraindividual emotional experience across time. Filipowicz et al. [FBM11] found that the influences of emotional transitions on social interactions may be different than corresponding steady-state emotions via several laboratory studies. In addition, Sudhof et al. [SGEMP14] developed a theory of conditional dependencies between emotional states to characterize affective transitions. Our work is based on these previous theories in emotion psychology, while we focus on visually exploring emotion transitions and discovering their potential triggers.

2.2. Visual analysis of Emotions

A variety of techniques have been developed for sentiment detection and analysis from text. Mohammad [Moh15] summarized previous techniques to detect valence, emotions, and other affectual states from text. Although most of these techniques focus on determine valence (positive and negative dimension) in text, there are some work on automatically detecting basic emotions. Leshed et al. [LK06] applied SVM to recognize emotions in LiveJournal's blog posts. Aman and Szpakowicz [AS07] employed annotated blog posts to develop supervised machine learning algorithms to classify emotions in text. In addition, there are efforts on finding the triggers for certain emotion from text. For example, Mihalcea et al. [Mih06] employed 'linguistic ethnography' to extract actions and other kind of factors highly related to happiness from a blog corpus.

Visualization has been successfully applied to analyze and explore the results of sentiment analysis. TwitInfo [MBB*11] is designed with several visual metaphors and timeline-based interaction to detect social events as well as the sentiment from corpus of tweets. Zhao et al. [ZGWZ14] proposed PEARL, a multi-dimensional emotion analysis tool on social media. They combined Plutchik's four pairs of primary emotions with the dimensional PAD model to model the emotions of individuals from a finer level, and revealed the temporal evolution of single person's emotion and its possible trigger. This paper focuses on public's emotion transitions instead of individual's emotions. SentiCompass [WSK*15] uses a supervised classifier together with an affective dictionary to detect emotions of tweets, and applies a cylindrical tunnel and a cyclic layout of various emotions to display time-varying emotions.

Most previous methods apply machine learning methods to detect emotions in tweets. In contrast, we use blog corpus annotated with emotional labels, which are mostly posted by the bloggers themselves. This emotion data is more accuracy than the results of sentiment analysis. Furthermore, we support analysis of both public's emotions and emotion transitions over time in a unified framework.

2.3. Visual Analysis of Temporal Event Sequences

Temporal event sequences have been widely explored in visual analytics. Previous methods usually focus on visualizing how multiple entities evolve after going through different events and their outcomes. For example, LifeFlow [WGGP*11] aggregated temporal event sequences to a tree and acyclic directed graph. It then aligns the events according to corresponding timeline. OutFlow [WG12] extends LifeFlow by arranging those events through their layers in the graph, and events elapse time is encoded to the width of the rectangle which represents an aggregation of same events appeared to different entities. In addition, EventFlow [MWP*12] provides an approach for exploring the relationship between the temporal event sequences and their outcomes. OpinionFlow [WLY*14] combines a tailored density map with the Sankey graph to represent the diffusions of public's opinions among corresponding topics. A stacked tree is used to facilitate the comparison the diffusion patterns across different topics. VAiRoma [CDW*16] combines timeline view with geographic map and topic content to explore the relationships between events, places and time in Roman history.

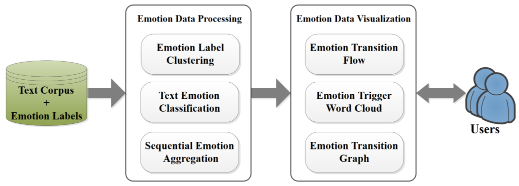


Figure 2: Pipeline of EmotionFlow.

Similarly, we aggregate emotions by time and use flow to intuitively show how emotions vary over time and interactively analyze the triggers behind emotion transitions. We also provide a graph to discover the general emotion transition patterns.

3. Overview

Our emotion data is collected from LiveJournal, a social networking website where users can post a blog, journal or diary. The website provides a selection toolbar including 132 emotional labels for its members to choose from to describe their emotional states when they are posting a blog, and they can also write short texts themselves to express their emotional states. EmotionFlow, as shown in Figure 1, is designed to assist users achieve a better understanding of the dependencies between emotions and discovering of their potential triggers. The pipeline of our system is illustrated in Figure 2, and it comprises two main parts: emotion data processing and emotion data visualization.

The input of our system is text corpus and their associated emotion labels. In emotion data processing, emotion labels are first clustering according to both Parrot's emotion classification model of Emotion and the transition property of each emotion. All emotion labels are clustered into six emotion categories. In LiveJournal, one user can also post a blog without an emotion label due to some reasons, such as forgetting to choose the emotion label. In this case, we can train models based on the blogs with emotion labels to classify these unlabeled blogs. Thus, the amount of valid blogs will become larger to make the analysis more trustworthy. Finally, we aggregate emotions of all users by time to construct public' emotion transitions.

After data processing, we design three main visualization methods to interactively analyze emotion transitions and discover their potential triggers. Emotion transition flow in Figure 1(b) is used to show the evolutions of emotion transitions over time. Emotion trigger word cloud in Figure 1(d) extracts the key words in the associated blogs and is used to explore the possible trigger of one emotion or emotion transition during a period of time. Emotion transition graph in Figure 1(f) shows the accumulation of transitions among the emotions for discovering the general emotion transition patterns. Our system also provides rich interactions, such as filtering, and highlighting, for users to intuitively analyze emotions and their transitions. The details of these methods will be discussed in the following sections.

4. Emotion Data Processing

Emotion data in our system is the blog posts and their associated emotion labels. In the data processing stage, emotion labels are first clustered, unlabeled blogs are then assigned with emotion categories, and finally emotions are aggregated sequentially to construct general emotion transitions.

4.1. Emotion Label Clustering

There are 132 emotional labels for selection in LiveJournal, and the user can also express the emotion in her own words. In our emotion data, we actually collect more than 1000 kinds of emotional labels, most of which are appear only a few times. It is difficult or even impossible to analyze and visualize so many kinds of emotion labels directly. Thus, we need to cluster these emotional labels to several general emotion categories. The emotion labels within the same category should satisfy two criteria: 1) being similar according to their VAD values. The VAD value of an emotional word is a three dimensional vector, which represents the emotion's inner character of valence, arousal, and dominance. Emotions with similar VAD score have similar degree to certain dimensions, thus, they share a common emotional profile [Rus03]; and 2) having the similar compressed transition probabilities (CTP values). The first criterion is to make sure the clustering results are reasonable and easier to understand, while the other criterion is to ensure that the transition patterns are more obvious when visualizing them.

Before we cluster emotion labels, we first remove low-frequency emotion labels. The filtering process should comply with two requirements: 1) discarding the untypical emotion labels whose occurrence frequencies are low and 2) keeping the size of corpus not affected much. We first sort the labels according to their occurrence frequencies. The labels' frequencies and their rank follow a power law distribution. According to the statistics rules of data that fits the power law, there exists a proportion α , and the top α of labels contain more than 1 - α of the total sum of occurrence frequencies. Emotion labels, whose sum of frequency amount to α percent of the total (α = 85% in this paper), are preserved. After the filtering process, we finally obtain 99 emotion labels in our emotion data.

We then choose Parrott's emotion classification model [Par01] to cluster our emotion labels into emotion categories. This model is one of the most nuanced emotion classification models so far, and emotions are conceptualized under a 3-level hierarchical structure in this model. Users can easily characterize a detailed emotional state using this classification and understand a variety of emotions relationship under different levels. Besides, its hierarchical structure is very suitable for progressively exploration. To be specific, general patterns of emotion transitions can be drawn from the six emotion labels in the first level, while transitions between the second level's emotions provide insights when focusing on one general emotion. Blogs containing the most detailed information are labeled with emotions in the third level. Thus the exploration can be done progressively.

Joy	Love	Surprise	Anger	Fear	Sadness
43.58%	17.93%	0.51%	6.97%	5.12%	25.89%

Table 1: The occurrences of emotions in the first level of Parrott's emotion model in our emotion data.

However, the model does not cover all of 99 emotional labels in our emotion data. Inspired by [ZGWZ14], for the labels that do not appear in Parrott's model, we refer to the labels with their synonyms through WordNet [Fel98]. If the synonyms appear solely in one category of Parrott's model, we will assign the emotion label to this category. Otherwise, we will refer to the ANEW lexicon for their VAD scores [BL99]. After that, we deploy a KNN-like classification algorithm to the remaining emotion labels. The distance of two emotional labels is determined by both the VAD values (criterion 1) and CTP values (criterion 2). This approach considers both dimensional characteristics of emotions and dependencies between emotional states, so emotion categories can reflect similarities in the conscious experience along with transition patterns among their members. Emotion transitions between such categories can be revealed more clearly.

Based on the above clustering method, 99 emotion labels are classified into Parrott's emotion model. The occurrences of emotions in the first level in our emotion data are summarized in Table 1. Blog posts labeled with positive emotions (joy, love, and surprise) account for 62.01%, far more than blog posts labeled with negative emotions (sadness, anger, and fear). This can be explained by positive bias in social media [FY15], as people are more inclined to share and favorite positive contents.

4.2. Text Emotion Classification

Our emotion data has unlabeled blogs where the user do not specify the emotion label. In order to obtain more users' emotions and their associated blog posts, we complement the unlabeled blog with one of Parrott's emotions instead of the original emotion labels. This would greatly improve the accuracy of emotion classification.

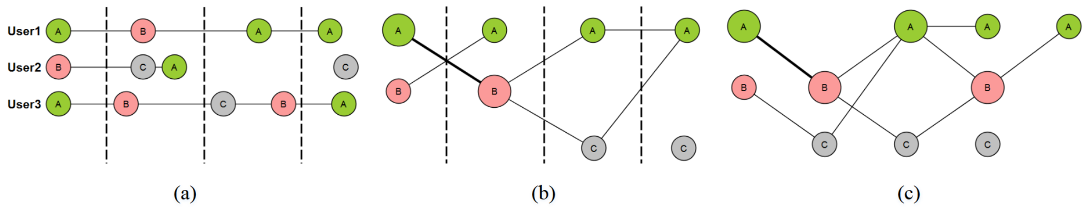


Figure 3: Sequential emotion aggregation. (a) three users' original emotional labels in four time bins. (b) the result of the timeline-based aggregation. (c) the result of the order-based aggregation

This is basically an emotion classification problem. A lot of methods have been proposed for emotion classification, and most of them use supervised machine learning methods. In our case, emotion labels are organized under a hierarchical structure, and blog posts generally have more content than tweets. The hierarchical method introduced [KI09] has been proved useful in classification for the long blog posts when classes have hierarchical structure. To be specific, the hierarchical approach first trains one classifier to classify unlabeled data into categories in the first level of the hierarchy. Then multiple classifiers are trained for each category in the first level to classify among its subcategories (which are in the second level). This top-down process proceeds till the finest level of the hierarchy. This approach corresponds to the 3-level hierarchical structure of our emotion, so we choose this approach to perform the text emotion classification.

Our feature set incorporates bag-of-words (BoW) features, lexicon-based semantic features and length-related features. Besides the lexicon resource in [KI09], we also use SentiWordNet [ES06], a lexical resource derived from the WordNet with scores indicating sentiment polarity of words, to calculate the sentiment orientation score for each post. SentiWordNet provides the synset for each word in the text, and each term in the synset is associated with a score ranging from 0 to 1. We use the average score of each term to show the sentiment bias of that word. Since social media often contains a lot of slangs and misspelled words, we further apply twitter lexicon [HL04, LHC05] to calculate sentiment orientation score.

Similar to the hierarchical method, we use SVM as the classifier. We randomly select 80% of all labeled blogs as training set and the remains as test set to evaluate the result. The average accuracy of the six emotion labels in the first level is 49.50%.

4.3. Sequential Emotion Aggregation

After each blog is assigned with one of Parrott's emotions, each user's emotions can be sequentially connected according to their blogs' posted times. These sequential emotions comprise an emotion chain, which describes the user's emotion transitions. However, if the time interval between two neighbor emotions is too large, the dependencies between them will be weak. Thus, we propose the upper bound of the length of time interval as T. Suppose the post time of two neighbor emotions are t_i and t_{i+1}, if t_{i+1} - $t_i > T$, then these two emotions does not form a chain. In our implementation, the upper bound T is set to 7 days. When the chain has been formed, we need to aggregate all users's emotion chains to emotion transitions, since we focus on general patterns of emotion transitions regardless of certain user.

We first aggregate emotion chains based on timeline. The time period is first sliced into time bins with the same length, and the same emotions in each bin are accumulated. If the time bin is not small enough, there could be more than one emotion expressed by the same user within the time bin, such as User 2 in the second time bin in Figure 3(a). In this case, we choose the emotion with the largest elapsed time as the representative emotion in this time bin. Figure 3(b) is the timeline-based emotion aggregation result. The area of each node is proportional to the number of users in one emotion, and the users in one emotion can have different emotions in the next time bin. This aggregation result can be used to analyze time-based emotion transitions and their potential triggers.

Besides the timeline-based aggregation, we can also organize and accumulate emotions according to their orders in emotion chains. It can capture all of emotion transitions within the time period. Figure 3(c) shows the aggregation result of this order-based method. We can further accumulate transitions among emotions to construct a graph. Each node in the graph is one emotion, and each edge represents all transitions between two emotions in the time period. This graph can be used to explore general emotion transition patterns.

5. EmotionFlow Design

In this section, we describe our design choices concerning EmotionFlow, as shown in Figure 1. The timeline-based aggregation result is displayed by the emotion transition flow. Key words are extracted from the emotion associated blog posts and visualized by word cloud to discover the possible triggers of the emotion and emotion transition. The node-link graph shows the order-based aggregation result to analyze emotion transition patterns.

5.1. Emotion Transition Flow

The flow or river is one of the most intuitive metaphors to display time-oriented data, and is getting familiar by general users. Thus, we choose the flow to show emotion transitions over time, as shown in Figure 1(b).

The time line in the horizontal axis is sliced into several time bins, which have the same time range. The emotions are placed on the vertical axis. Each emotion in one time bin is represented by a vertical bar, called the emotion bar. The color scheme in Plutchik's wheel of emotions [Plu01] is used to encode different emotions. We make some modifications to this scheme; for example *sadness* is encoded by grey instead of blue. The height of the emotion bar is proportional to the number of users in this emotion. The transition between emotions in adjacent time bins is displayed by the cubic BÃľzier curve, called the emotion path. For the transition between two different emotions, a color gradient is used to fill the path to imply the transition from one emotion to another. Similarly, the width of the path is proportional to the number of users belonging to this emotion transition. The width in the middle of the path is drawn thinner to reduce the visual clutter of emotion transitions.

As described in Section 4.3, a user's emotion chain can break between two time bins. Thus, the emotion transition cannot be formed depending on the blogging frequency of the user. Within a time bin, there may be starting-points and end-points of emotion chains or even isolated emotion which does not compose any transition. Thus, both the incoming flow and the outgoing flow are narrower than the emotion bar in a time bin.

As it is very difficult to reduce the visual clutter just by minimizing the number of edge crossing, We come up with various designs for interacting with the emotion transition flow. When hovering on an emotion path, a line chart will appear at the bottom to indicate how the amount of this emotion transition evolves over time intuitively, as shown in Figure 1(c). Users can also hover on the name of the emotion to see the number of blogs in that emotion over time in the same line chart. If users finds it hard to find the point of the pattern because there are too many edges, a filter slider bar is provided to filter out the edges whose width is under the specified value. Besides, if users are attracted by an emotion transition, they can hover the mouse over the transition so that other edges will fade gradually.

The emotion transition flow enables users to visually develop an understanding of the number of blogs in each emotion category and the scale of emotion transition over time. It aids users in analyzing the proportion of emotions and the emotion transition patterns, which are both time-related.

5.2. Emotion Trigger Word Cloud

With emotions and emotion transitions over time in the emotion transition flow, the user may be interested in one of the

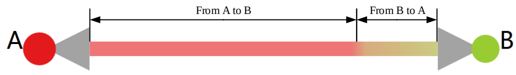


Figure 4: Transition between two emotions A and B. Length of a color bar is in positive correlation with the proportion of corresponding transitions.

emotions or emotion transitions, and try to analyze the cause of this pattern. As the blogger selects the emotion label when posting a blog, each emotion has a large number of associated blog posts in the investigated time period. The reason why the user has this emotion may be expressed in the blog post, and we can derive potential triggers from these blogs.

An emotion trigger word cloud is designed for an overview of related blog posts, as shown in Figure 1(d). The word cloud will appear when the user double clicked the emotion bar. We first search for all blogs labeled with this emotion during the selected time bin. A set of distinctive keywords are selected from these blog posts to provide the user an intuitive description of the emotion. These keywords should occur frequently and commonly in this emotion's blog posts and rarely appears in others. Thus, we choose the typical TF-IDF (Term Frequency inverse document frequency) to pick the keywords from the blogs. Then the words are placed aside the clicked position in a circle by a randomized greedy algorithm.

We also provide a text view to list the blogs labeled with the emotion in the selected time bin to support detailed analysis of potential triggers, as shown in Figure 1(e). If the user clicks on keyword in the word cloud, the blog posts will be sorted according to the keyword frequency in the blog. Thus, the user can first read the most related blogs to further analyze potential triggers.

5.3. Emotion Transition Graph

Parrott's classification model organizes emotions to a 3-level hierarchical structure, and the flow shows a timeline based description to transitional patterns among six emotions in the first level. In order to provide the user with more detailed transition information, we provide an emotion transition graph to show the accumulation of transitions among emotions during a time period, i.e., the order-based aggregation result.

As shown in Figure 1(f), emotions are represented by nodes, whose colors are also from Plutchik's wheel of emotions [Plu01]. These nodes are positioned along a circle path. These emotions can be the six first-level emotions or among one of theirs subcategories. The edge indicates the transition between two emotions, and the edge width is proportional to the accumulated number of transitions between two emotions. Since they are directed edges, we fill each edge with two colors, which are the colors of the end nodes. As illustrated in Figure 4, the transition between two emotions A and B are represented by the edge with two colors.

The nodes are positioned along a circle, and it is desired to highlight the edges with a larger amount of transitions while reducing the edge crossing as much as possible. Thus, we propose the following cost function to optimize the positions of nodes:

$$E = -\beta \#intersections + \alpha \sum w_e \times l_e, \qquad (1)$$

where $\#intersections$ is the number of the edge crossings, w_e and l_e indicate the width and length of the edge e, respectively, and α and β are two parameters to balance two terms. In our implementation, α and β are both 0.5. We seek out to highlight the edge with a broader width by positioning its end nodes as far as possible.

We first lay nodes randomly along a circle, and then iteratively exchange the positions of neighbor nodes to maximize the proposed cost function until the layout is stable or the iteration number reaches the threshold. The graph can help the user to understand emotion transitions in different emotion levels.

6. Evaluation

In this section, we conduct three case studies to demonstrate the usefulness and effectiveness of EmotionFlow. Our system is a browser/server-based system. The visualization and interaction at the client side is implemented with javascript taking the advantage of D3.js [BOH11], while emotion data processing at the server side is built with Python and Django. Our emotion data is collected from LiveJournal. The data set contains 635164 blogs by 37185 users.

6.1. Case Study 1: Analyzing Emotion Transition Patterns

Our system can help psychologists to visually explore emotions and their transitions based on large emotion data and to testify some psychological arguments. We show one example in this section.

As shown in Figure 1, it is clear that the emotion bars labeled with *joy* and *sadness* are obviously larger than others. Each emotion has a widest emotion path to itself, i.e., there is a tendency for bloggers in LiveJournal to post consecutive blogs under the same emotion. On average,during the specific month, the transition from *joy* to *joy* takes 30% of all blogs in one time bin, while the transition from *sadness* to *sadness* takes 22.41%, and the transition from *fear* to *fear* is only 13.36%. We can generate a hypothesis that compared to other emotions, *fear* tends to transform to other emotions more quickly. Thus, we are particularly interested in the emotion *fear* and its emotion transitions.

We then try to make more efforts to demonstrate this tendency. Figure 5(a) shows three emotions *fear*, *joy*, and *sadness* from November 20th 2004 to December 20th in 2004. When we select the emotion bar of *fear* at November 26th, it can be observed that the proportion of emotion transition from *fear* to *joy* is quite large compared to *fear* to *fear* itself. *Joy* can characterize the normal emotional state in daily life [NVVMK08]. We then let η_E denotes the average value of transition from emotion E to E itself divided by transition from E to *joy* during the specific time period. η_{fear} equals 1.46 which is the largest among all six values. This is corresponding to our former hypothesis.

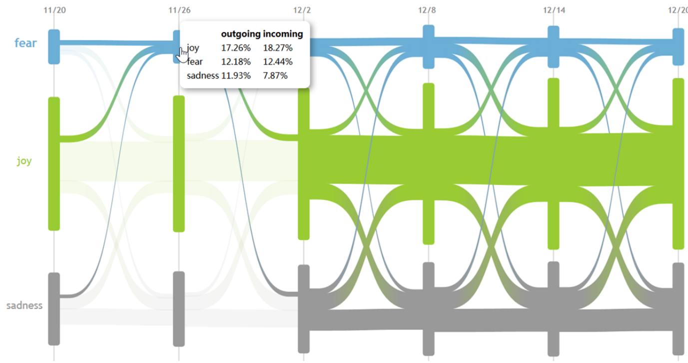
Then we generate the emotion transition graph (Figure 5(b)) with all the data, so that the graph can reveal general emotion transition patterns. Most edges have a rough fifty-fifty split, which means emotion transitions between each pair are balanced in the long range. As in Table 2 each cell T_{ij} represents the percentage of transition from emotion in row i to emotion in column j. However, we find that transitions related to *fear* is still special. Among all the pairs related to *fear* except the *fear-surpirse* pair, the number of transitions from *fear* is larger than the number of transitions into *fear*. This also verifies the hypothesis in the emotion transition flow, indicating a higher tendency for the emotion *fear* to transform quickly. This hypothesis is consistent with a psychological argument on the biased manner in individual's emotion evolvement. This argument points that fear, as a negative affect with high activation [BL99] will endure a faster fading process across time [WPR04]. Thus, the results of our visual exploration provide a proof for one psychological argument.

6.2. Case Study 2: Exploring Emotion Transitions During New Year Breaks

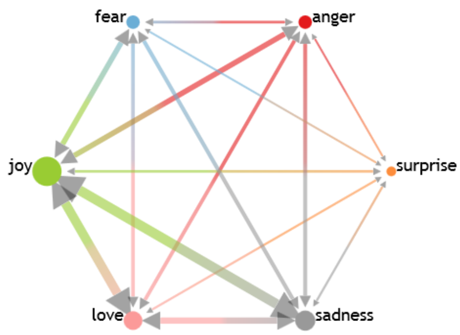
The second case study focuses on emotions and their transition patterns during the period around New Year's Day.

We first specify the time range from November 19th 2014 to January 18th 2015, and select the scale of time bin as one week. The result is shown in Figure 1. We hover on the name of the emotion to check the line chart on the bottom 6(a)(d). The line chart show that Emotions *joy* and *sadness* increase a lot during the time period about December 31th, and then drop down to the normal level gradually while no obvious changes in other emotions. It seems that people are more apt to post blogs with two basic polarities of emotion during New Year breaks. It is not surprising that people tend to be happier on festival. However, from the emotion transition flow, we can also see that the number of bloggers with a negative emotion also increases.

In order to find possible triggers, we double click the emotion bar of *sadness* at the timebin around December 31th, and check the word cloud generated from all blogs labeled with *sadness*. The biggest two words in the word cloud are "finish" and "life" as shown in Figure 1, which conclude the reasons that make people unhappy on the happy moment. We first click the word "finish" to obtain the blogs that are sorted according to the frequency of the word for detailed



(a)



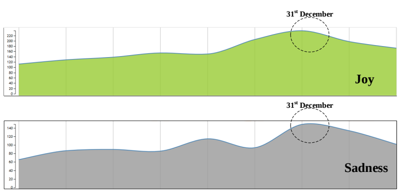
(b)

Figure 5: (a): Emotion Flow of *fear*, *joy* and *sadness* from November 20th to December 20th in 2004. (b):Emotion Transition Graph generated by the whole dataset.

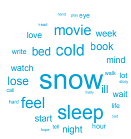
	Anger	Fear	Joy	Love	Sadness	Surprise
Anger		48.38	53.86	51.15	52.13	44.56
Fear	51.62		54.20	50.82	51.76	47.16
Joy	46.14	45.80		46.78	47.22	46.00
Love	48.85	49.18	53.22		50.26	51.32
Sadness	47.87	48.24	52.78	49.74		48.10
Surprise	55.44	52.84	54.60	48.68	51.90	

Table 2: Emotion transition proportions.

analysis. We find that most of the bloggers tend to summarize their works last year on the approaching of New Year, and those blogs are mainly labeled with emotion labels like "tired" or "pensive", which are classified to the first level emotion *sadness*. For example, there are lots of blog posts like "*Just wanted to do one last post for the year, finished 60 books this year, up from 54 last year. 60! A new record for me*" or "*early 2014 seems like it's years ago! Think the first thing I finished was...*". The visual exploration about the word "life" is similar. From the associated blog posts, we find that for lots of bloggers, the coming of New Year does not bring them a positive attitude to their lives, and they tend to complain. For example, one blogger writes "*I understand that life can't be all sunshine and rainbows. Even if it wasn't my worst year, 2014 was horrible.*"



(a)



(b)

Figure 6: (a): Line chart of *joy* and *sadness*. Both emotions have a sharp increase on December 31th. (b)Word cloud of *sadness* from December 31th to January 6th.

We then explore the emotion transition pattern. We find that the transition from *joy* to *sadness* reaches its maximum number during the time bin from December 31th to January 6th. We make use of the emotion trigger word cloud to uncover why a great many of people's emotion change in this way. We double click the emotion bar *sadness* to show the word cloud (Figure 6(b)). There are two potential triggers derived from the word cloud and related blogs. on the one hand, it was snowing in lots of places during that time period and temperature was low. On the other hand, as it is the first week to get back work, lots of bloggers are not prepared for it. This is not shown directly in the word cloud, but we conclude this from the blog posts.

Thus, through the interactive exploration with EmotionFlow, we obtain blog users' emotional states along with the transition patterns during important events. In addition, reasons behind the emotion patterns are revealed by further analyzing the emotion trigger word cloud and the blog texts.

6.3. Case Study 3: Exploring Real-world Events Correlated with Emotions

A special event may affect people's emotions. For example, there will mass increase in the level of *sadness* around major weather phenomena. Thus, we can analyze the potential trigger for certain emotion transition and derive some events happened in the time period.

When we specify the time range from October 23rd 2014 to November 23rd 2014, the visualization result is shown in Figure 7. We first look into the line chart to check the number of each emotion during this time period. The number of *anger* shows a clear increase around November 6th, while the number of *joy* and *love* drop a few. We then check the emotion path between October 30th and November 6th, which has *anger* as the end point. It can also be seen that the transition from *joy* or *love* to *anger* increases during this time period. It is reasonable to assume that some events may account for this emotion transition.

We display the emotion trigger word cloud to uncover the reasons, as shown in Figure 7. In the word cloud, the largest

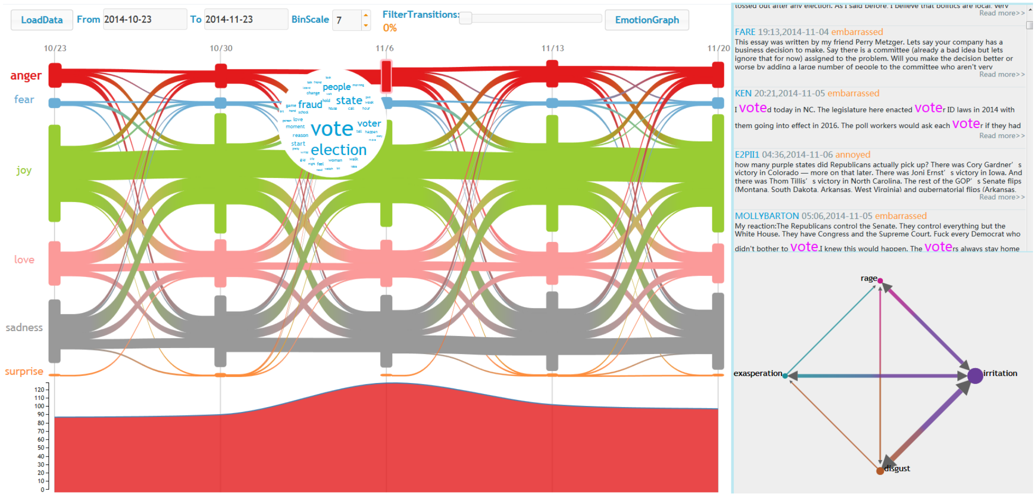


Figure 7: Overview of EmotionFlow during 2014 United States midterm election.

word is "vote", while other important words contain "election", "state", "party", "fraud", and "voter". We refer to the search engine and find out that the United States midterm election took place on 4th November 2014. Republic Party gained control of the Senate after the election. Several news reported that the fraudulence related issues during the election. We may make an assumption that the fraud in 2014 U.S. midterm election has made many bloggers feel *anger*. This assumption is drawn by people's common sense and by further exploring our system we can decide whether this assumption stands.

We double click on "vote" and "election" to read related blog posts containing these words. It turns out that among these blogs, many are comments on the result of the election, such as "*I admit that I had more confidence in the Democratic electorate to get out and vote this year, and that they would win the races they were supposed to. I admit I was wrong.........*", which shows the discontent of Democratic Party supporters. Some blogs are long article expressing criticism to politics with the midterm election as opening words. Some other users mention the election voting but their focus is not on the election itself. Thus, on the topic of midterm election, we can conclude some types of reasons explaining users' emotion transition to *anger*. Then, we check blog posts with the keyword "fraud". Lots of users show their hate towards the fraudulent activities. But still, there are some users have the emotion of *anger* because they think the fraud issues have been exaggerated. For example a blogger says "*there has been so much exposure of fraudulent voting–because of OUR countries alternative media's great reporting: this is what worrisome to me!*

We also observe the emotion transitions of *anger*'s subcategories in the emotion transitions graph. We find that *irritation* and *disgust* are the biggest two emotion nodes, and the transition from *irritation* to *disgust* outweighs the other. This is consistent to the above finding. We also observe the subcategories' transitions in the graph. Thus, by exploring emotion data with our system, we can discover an event through a special emotion transition. We also find some potential triggers for this transition, which are more comprehensive than explanations drawn by common sense.

7. Conclusion

In this paper, we have proposed a visual analytic tool for analyzing emotions and their transitions. Based on online blogs and their associate emotional labels, emotional labels are first clustered according to Parrott's emotion classification model, and classified into a hierarchical emotion category. The blogger's sequential emotions are aggregated based on the timeline, and then they are visual encoded by emotion transition flow to explore the overall emotions and their transitions over time. Particular emotion or transition can be explained through emotion trigger word cloud and original blog posts. The emotions can be aggregated only by emotions themselves to analyze general emotion transition patterns in emotional psychology. Case studies from both emotional psychology and social events demonstrate the effectiveness of our data aggregation and visual design in EmotionFlow. As the future work, we would like to show our EmotionFlow to psychologists and socialists to further verify our method and findings.

References

[AS07] Aman S., Szpakowicz S.: Identifying expressions of emotion in text. In *Text, Speech and Dialogue* (2007), vol. 4629 of *Lecture Notes in Computer Science*, pp. 196–205. 3

[BL99] Bradley M. M., Lang P. J.: *Affective norms for English words (ANEW): Instruction manual and affective ratings*. Tech. rep., Technical Report C-1, The Center for Research in Psychophysiology, University of Florida, 1999. 4, 7

[BOH11] BOSTOCK M., OGIEVETSKY V., HEER J.: D^3 data-driven documents. *Visualization and Computer Graphics, IEEE Transactions on 17*, 12 (2011), 2301–2309. 7

[CDW*16] CHO I., DOU W., WANG D. X., SAUDA E., RIBARSKY W.: Vairoma: A visual analytics system for making sense of places, times, and events in roman history. *Visualization and Computer Graphics, IEEE Transactions on 22*, 1 (2016), 210–219. 3

[CMK13] CALVO R. A., MAC KIM S.: Emotions in text: dimensional and categorical models. *Computational Intelligence 29*, 3 (2013), 527–543. 2

[EF01] EATON L. G., FUNDER D. C.: Emotional experience in daily life: valence, variability, and rate of change. *Emotion 1*, 4 (2001), 413. 2

[ES06] ESULI A., SEBASTIANI F.: Sentiwordnet: A publicly available lexical resource for opinion mining. In *Proceedings of LREC* (2006), vol. 6, Citeseer, pp. 417–422. 5

[FBM11] FILIPOWICZ A., BARSADE S., MELWANI S.: Understanding emotional transitions: the interpersonal consequences of changing emotions in negotiations. *Journal of personality and social psychology 101*, 3 (2011), 541. 2

[Fel98] FELLBAUM C.: *WordNet: An Electronic Lexical Database*. Bradford Books, 1998. 4

[FY15] FERRARA E., YANG Z.: Quantifying the effect of sentiment on information diffusion in social media. *PeerJ Computer Science 1* (2015), e26. 4

[HL04] HU M., LIU B.: Mining and summarizing customer reviews. In *Proceedings of the tenth ACM SIGKDD international conference on Knowledge discovery and data mining* (2004), ACM, pp. 168–177. 5

[KI09] KESHTKAR F., INKPEN D.: Using sentiment orientation features for mood classification in blogs. In *Proceedings of the IEEE International Conference on Natural Language Processing and Knowledge Engineering (IEEE NLP-KE 2009)* (2009). 5

[LHC05] LIU B., HU M., CHENG J.: Opinion observer: analyzing and comparing opinions on the web. In *Proceedings of the 14th international conference on World Wide Web* (2005), ACM, pp. 342–351. 5

[LK06] LESHED G., KAYE J.: Understanding how bloggers feel: recognizing affect in blog posts. In *CHI'06 extended abstracts on Human factors in computing systems* (2006), ACM, pp. 1019–1024. 2, 3

[MBB*11] MARCUS A., BERNSTEIN M. S., BADAR O., KARGER D. R., MADDEN S., MILLER R. C.: Twitinfo: aggregating and visualizing microblogs for event exploration. In *Proceedings of the International Conference on Human Factors in Computing Systems, CHI 2011, Vancouver, BC, Canada, May 7-12, 2011* (2011), pp. 227–236. 3

[Meh80] MEHRABIAN A.: *Basic Dimensions for a General Psychological Theory*. Oelgeschlager, Gunn& Hain Inc., 1980. 2

[Mih06] MIHALCEA R.: A corpus-based approach to finding happiness. In *In AAAI 2006 Symposium on Computational Approaches to Analysing Weblogs* (2006), AAAI Press, pp. 139–144. 3

[MKK*12] MCDUFF D., KARLSON A., KAPOOR A., ROSEWAY A., CZERWINSKI M.: Affectaura: an intelligent system for emotional memory. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (2012), ACM, pp. 849–858. 2

[Moh15] MOHAMMAD S. M.: Sentiment analysis: Detecting valence, emotions, and other affectual states from text. *Emotion Measurement* (2015). 3

[MWP*12] MONROE M., WONGSUPHASAWAT K., PLAISANT C., SHNEIDERMAN B., MILLSTEIN J., GOLD S.: Exploring point and interval event patterns: Display methods and interactive visual query. *Human Computer Interaction Lab, University of Maryland* (2012). 3

[NVVMK08] NEZLEK J. B., VANSTEELANDT K., VAN MECHELEN I., KUPPENS P.: Appraisal-emotion relationships in daily life. *Emotion 8*, 1 (2008), 145. 7

[Par01] PARROTT W. G.: *Emotions in social psychology: Essential readings*. Psychology Press, 2001. 2, 4

[Plu01] PLUTCHIK R.: The nature of emotions. *American Scientist* (2001). 1, 2, 6

[Rus03] RUSSELL J. A.: Core affect and the psychological construction of emotion. *Psychological review 110*, 1 (2003), 145. 4

[SGEMP14] SUDHOF M., GOMÉZ EMILSSON A., MAAS A. L., POTTS C.: Sentiment expression conditioned by affective transitions and social forces. In *Proceedings of the 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (New York, NY, USA, 2014), KDD '14, ACM, pp. 1136–1145. 2

[WG12] WONGSUPHASAWAT K., GOTZ D.: Exploring flow, factors, and outcomes of temporal event sequences with the outflow visualization. *Visualization and Computer Graphics, IEEE Transactions on 18*, 12 (2012), 2659–2668. 3

[WGGP*11] WONGSUPHASAWAT K., GUERRA GÓMEZ J. A., PLAISANT C., WANG T. D., TAIEB-MAIMON M., SHNEIDERMAN B.: Lifeflow: visualizing an overview of event sequences. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (2011), ACM, pp. 1747–1756. 3

[WLY*14] WU Y., LIU S., YAN K., LIU M., WU F.: Opinionflow: Visual analysis of opinion diffusion on social media. *Visualization and Computer Graphics, IEEE Transactions on 20*, 12 (Dec 2014), 1763–1772. 3

[WPR04] WENZEL A., PINNA K., RUBIN D. C.: Autobiographical memories of anxiety-related experiences. *Behaviour Research and Therapy 42*, 3 (2004), 329–341. 7

[WSK*15] WANG F. Y., SALLABERRY A., KLEIN K., TAKATSUKA M., ROCHE M.: Senticompass: Interactive visualization for exploring and comparing the sentiments of time-varying twitter data. In *Visualization Symposium (PacificVis), 2015 IEEE Pacific* (2015), IEEE, pp. 129–133. 3

[ZGWZ14] ZHAO J., GOU L., WANG F., ZHOU M.: Pearl: An interactive visual analytic tool for understanding personal emotion style derived from social media. In *Visual Analytics Science and Technology (VAST), 2014 IEEE Conference on* (Oct 2014), pp. 203–212. 2, 3, 4