Visual Analysis of Meetup Group Event Topics

Team 08
CSE 6242: Data and Visual Analytics
Fall 2019

Alejandro Ardila
aardila3@gatech.edu
Georgia Institute of Technology
Atlanta, Georgia

Jashua Thejas Arul Dhas
jdhas3@gatech.edu
Georgia Institute of Technology
Atlanta, Georgia

Joshua Demusz
jdemusz3@gatech.edu
Georgia Institute of Technology
Atlanta, Georgia

Justin Lewis
jlewis356@gatech.edu
Georgia Institute of Technology
Atlanta, Georgia

Khanh Tran
ktran93@gatech.edu
Georgia Institute of Technology
Atlanta, Georgia

Tri Nguyen
trinpm@gatech.edu
Georgia Institute of Technology
Atlanta, Georgia

1 INTRODUCTION

Event marketing is an attractive research opportunity due to emerging, new technologies that are narrowing the gap between offline social interactions and online event hosting [19]. According to the latest Event Marketing Benchmarks and Trends report, overperforming companies and organizations invest between 20% to 50% of the marketing budget in hosting live events [1]. Therefore, hosting an event is one of the most effective channels for accomplishing business goals. However, planning an event can be stressful. Besides strenuous management effort, it is critical for the event organizers to understand the target market to maximize their chances of attracting people with relevant interests. Thanks to data available from event-based social networks like Meetup, event organizers can gain valuable insights about their audiences.

To make use of this data we provide data visualizations to explore how events are distributed in a geographic area, as well as providing custom metrics to help users decide what kind of events may be successful in the area. By leveraging sentiment analysis of user comments on events and topic extraction from event names we can inform users what the sentiment towards certain topic categories has been. We believe that visualizations will help event planners alleviate the complexity of understanding the market trends in specific sectors and make more informative decisions on their upcoming event.

2 SURVEY

Leveraging visual and analytical techniques on data has proven to be useful in virtually every industry, and social media based events are no exception [8]. The significance of understanding Meetup event characteristics can be remarkably beneficial to event planners in identifying factors such as market trends and under-represented yet high potential topics [5, 13].

In terms of utilizing Meetup data in analytics, most of the current research projects emphasize heavy analysis rather than user-friendly visualizations. Examples include using deep neural networks to predict the popularity of a Meetup group, or creating complex equations to quantify event location competitiveness [12, 15, 18]. Balancing the analysis of Meetup data with functional visualizations will contribute to the overall success of the project since visuals have proven to be effective in cultivating the decision making process [4, 7, 11].

Furthermore, sentiment analysis has proven to be successful for websites such as Twitter, Yelp, and Amazon in extracting positive/negative experiences, or common perception about a certain product or topic [10]. Thus, it is promising that such viable information about a Meetup topic could also be mining from the event comments [2, 6, 16]. When analyzing Meetup event comments, expected output is an insight of a certain city attitude towards a particular topic, which could be crucial knowledge for an event planner to utilize [9, 17].

In terms of risk, the existence of meetup location and user data security issues could pose a threat to the longevity of the project depending on how data access is managed in the future [3, 14].

3 PROBLEM DEFINITION

As leadership is increasingly understanding the value of hosting live events, their main interests reside in where should an event be hosted in order to leverage connectivity between a company and their attendees. Unfortunately, this decision still lacks the appropriate tools to analyze actual trends.

Given the opportunity that Event Marketing has in finding people with similar interests and promoting social interactions between them, we find crucial to know beforehand how such interest in different topics are clustered in geographical regions to maximize the engagement rate of hosting an event in a particular place. Our main intent will focus on applying topic analysis and relating them to the overall sentiment of previous hosted events. The measurements will be exposed in a set of visualizations so that event planners can take the best decision on where to host their next event.

4 PROPOSED METHOD

4.1 Data Extraction

The main goal of the data extraction process is to create a lightweight script that subscribes to the Meetup API and extracts and transforms data from topics, groups, events and comments into the a relational schema and dump them into a SQLite database.

4.1.1 *OAuth2 authorization:*

At present, Meetup API supports only Oauth 2.0 authentication

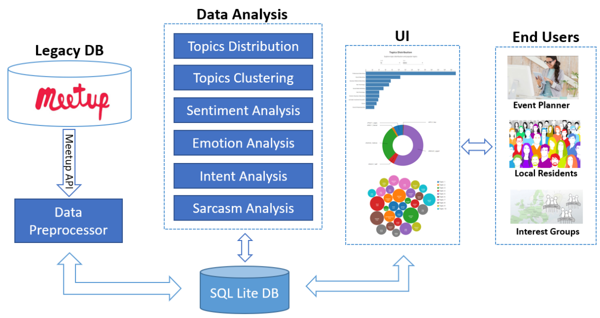

Figure 1: System architecture

protocol, which is a two-step process. First there is an authorization request, which generates a one-time code to bind the state of the requester to an authorization token. Afterwards, the client uses the code along with API keys and secrets to generate an expendable access token. The generated Access token is used for querying the API and extracting a raw response.

4.1.2 Data Extraction and Throttling policies:
Groups, Topics, Events, and Event Comments are the necessary datasets for performing various data analysis. For extracting Group and topic response data, latitude and longitude of a particular city need to be passed as an essential parameter. In our use case, we have considered San Francisco for exploring data and perform sentiment analysis. Using the latitude and longitude of San Francisco, we retrieved group response data, which has the keys attributes such as Group ID, group name, group URL, members, and category details in addition to other attributes. Several groups will fall under the same category.

Similarly, topic response data is also extracted, which contains key attributes such as Topic ID, Topic name, and the category associated with the topic.For matching Group ID retrieved from the group response data, event response data with attributes event ID and event name is collected. Event ID from event dataset and group URL from group dataset is used collectively to retrieve comments response data from comments URL.

The Meetup API has a clever throttling mechanism that prevents their servers from suffering a denial of service. If the client does not interact properly with the response header info, Meetup could block further requests from that specific consumer. Our script supports the policies defined by meetup and makes intelligent decisions on request handling.

4.1.3 Data ingestion into SQLite Database:
Extracted response data is loaded into an SQLite database named meetup. For each dataset of groups, events, comments, and topics corresponding table have been created. These tables serve as Dimensions with the necessary primary key and foreign keys. Group table will have Group ID as the primary index, and the same Group ID will be a foreign key in the events table. Similarly, events will have Event ID as the primary key, and topics will have Topic ID as the primary key. Data cleansing is performed on top of the ingested table to remove redundant data and to retain the comments for those events. Final curated data with Group ID, Event ID, Topic name, and comments are used for further sentiment analysis and topic saturation.

This summarizes our data stored in SQLite database:

- #Events description: 18,821
- #Public meetup groups : 12,530
- #Group/topics relationship: 19,362
- #Categories of topic: 2,833
- #Topics: 4,123

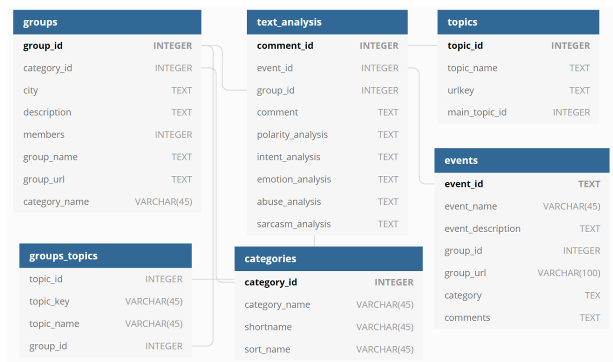

Figure 2: Database schema

4.2 Analysis

4.2.1 Text Analysis. Initially, we planned on analyzing the sentiment of Meetup event comments to see how positive or negative the general attendee attitude was towards an event topic. After further research, we decide to utilize different flavors of sentiment analysis to obtain deeper insights on comments given by meetup event's attendees. We implemented an interface on top of ParallelDot Python library to provide following analysis on Meetup's comments:

- Polarity analysis: Each comment classified into labels defined in $S_{polarity} = \{positive, negative, neutral\}$.
- Emotion analysis: Classify feelings and emotions of commenters about events into states defined in $S_{emotion} = \{angry, happy, excited, fear, sad, bored\}$.
- Intent analysis: Detect intent of comments about events to categories in $S_{intent} = \{news, query, spam, marketing, feedback\}$.
- Sarcasm analysis: Identify presence of sarcasm in the comments; i.e., $S_{sarcasm} = \{sarcastic, non-sarcastic\}$.
- Abusive analysis: Detect abusive and offensive language from comments; i.e., $S_{abusive} = \{abusive, hate{-}speech, neither\}$.

Accessing this model through the front-end visualizations is a key innovation of the project since it transforms the monotonous process of reading through individual comments in order to aggregate useful feedback into a streamlined process that can be performed in seconds.

4.2.2 Topic Extraction from Events. After conceptualizing which information might be useful to a Meetup event planner, we originally planned on analyzing Meetup event topics for a selected location. For example, the category "Technology" in San Francisco might have topics "Java", "Kafka", and "Spark". In reality, however, the topics underneath a category are still very general (e.g. the topics under "Technology" include "New Technology", "Web Technology", "Mobile Technology"). Thus, in order to determine more specific topics, we needed to extract this information from the events themselves.

In order to accomplish this, we implemented a topic-extraction model using the LDA (latent Dirichlet allocation) model that clusters groups into more granular topics. This model takes a list of pre-processed event names & descriptions as input, and returns a weighted distribution of topics based on frequently occurring words. This distribution of topics will allow us to identify popular Meetup topics that were not obvious based on the original data set, and identify similarities between Meetup groups and events. In terms of innovation, this topic-extraction model will allow us to organize Meetup groups and identify connections between groups in a way that is not readily available on the site or app, and will provide Meetup event planners with a new way to define their own events and identify similar events.

$$P(W, Z, \theta, \varphi, \alpha, \beta) = \prod_{i=1}^{K} P(\varphi_i; \beta) \prod_{j=1}^{M} P(\theta_j; \alpha) \prod_{t=1}^{N} P(Z_{j,t} | \theta_j) P(W_{j,t} | \varphi Z_{j,t})$$

Figure 3: Total probability equation for the LDA model.

5 USER INTERFACE

As mentioned in the Survey section, MeetUp.com and other existing approaches are not focusing on delivering enriched visualization. Specifically, there is a lack of information comparisons, which is limiting group leaders' ability to identify area opportunities. Therefore, our user interface design is inspired to enhance the existing practice as followed:

5.0.1 Topic distribution. One goal of Meetup group leaders and event planners is to increase membership and attendance of their groups & events. Thus, displaying information regards to the popularity and member attendance of Meetup topics could have valuable implications for such users.

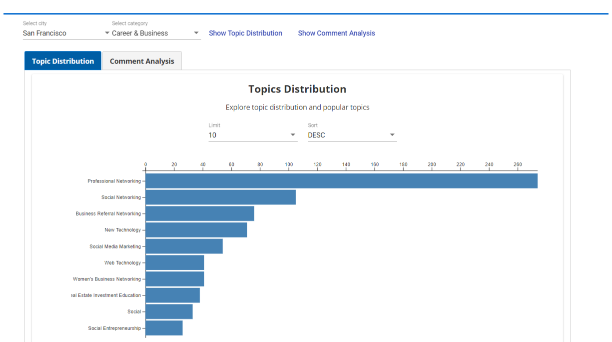

Figure 4: Distribution of topics under Tech events in San Francisco

In the 'Topic Distribution' tab, after the user makes an initial city & category selection, an interactive bar chart of Meetup subcategories is displayed. The user can then select the order and number of subcategories displayed. If the user selects one of the bars, a bubble matrix displays that allows the user to see popular topics generated from the topic extraction model. The size of the bubbles correspond with the topic word's weight. Furthermore, a table displays below this matrix with extra information, including the number of members underneath said topics.

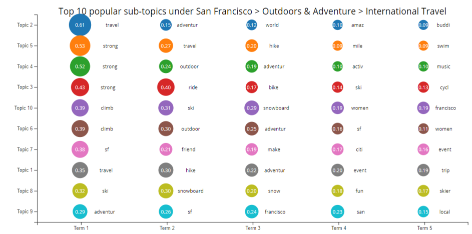

Figure 5: Our application visualizes topics generated by the topic extraction model. The size of the bubble corresponds to the weight of the topic.

Search

Group Name	# of Members	Created On	Category	Topics
SF New Tech	8381	3/17/2006 21:46	Tech	New Technology,Social Software,Web Technology
Startup Addicts SF	1631	8/16/2006 17:44	Tech	New Technology,Semantic Technologies,Social Networking,Web Technology
SF Beta	1945	9/9/2006 19:31	Tech	New Technology,Professional Networking,Social Media,Social Software,Web Technology
TED-San Francisco	1089	12/8/2006 14:20	Tech	New Technology,Web Technology
Bay Area Futurists	1477	2/12/2007 5:47	Tech	New Technology,Technological Singularity

Figure 6: Groups under Tech events in San Francisco

Users can use this information to help gauge the popularity of topics underneath a Meetup category, in addition to important metrics pertaining to group membership. Meetup event planners could use this data in order to predict which event topics may be more or less successful in their particular city.

5.0.2 Comment analysis. Similarly, event planners can increase membership by improving the quality of their events by listening to attendee feedback. In the 'Comment Analysis' tab, the user can select to analyze the comments under four different levels of granularity: category, cluster (from the topic extraction model), groups, or events. If the user selects cluster, group or event, they will have go on to make a more specific selection.

Once the user makes their selection, they can see a labeled pie chart depicting the comment analysis breakdown for sentiment, emotion, intent, sarcasm, and abuse. For example, the "sentiment" option is broken down into "negative", "positive", and "neutral". The sentiment pie chart would thus display a slice for each of these three qualities, along with labels indicating their percentage. If "negative" was the largest slice of the pie, that would indicate a majority negative sentiment for all the selection's attendee comments. This visualization will allow the user to analyze qualities of event comments on multiple levels of granularity, and will grant them information that will help them approve their events.

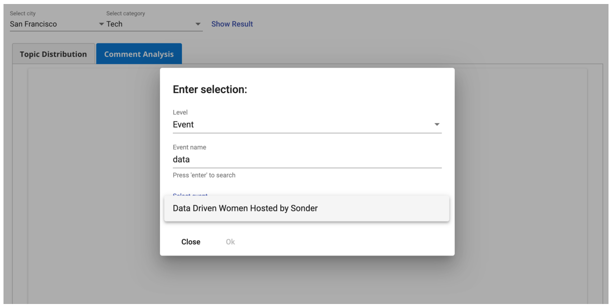

Figure 7: Example of 'Event' selection for comment analysis.

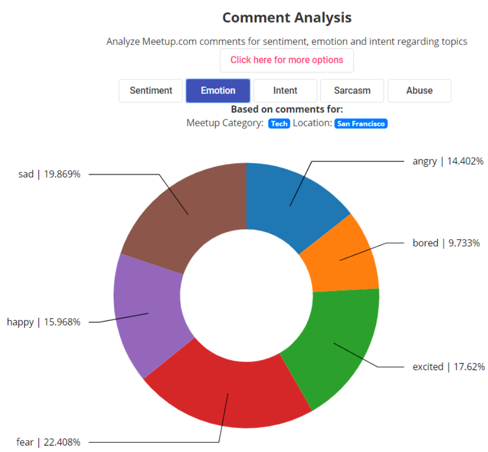

Figure 8: 'Emotion' analysis pie chart.

5.1 Tools

Tool	Usage
Angular	Front-end framework wrapped around D3
Flask	Back-end framework for Python
Gensim	Extract Meetup data
Meetup API	Data extraction
Meetup Kaggle dataset	Utilized until API consumer was ready
NLTK	Python library used for text processing
Parallel dots	Sentiment analysis
Python	Back-end language
SQLite	Data storage

5.2 User Evaluation

Since the success of this application is dependent on how useful the information provided is to Meetup event planners, it is crucial to evaluate the application from the user's perspective. We decided to evaluate the tool through online surveys, which are multiple-choice questions to avoid subjective responses. Meetup users among our friends and peers have participated in the online survey and provided their feedback.

The following questions were posted in the online surveys.

Questions:

- This tool is easy to navigate
- This tool would help improve the event's quality
- This tool would improve the event's popularity
- This tool helps in decision-making in regards to events or groups
- The tool provides non-trivial insights which I cannot find in meetup.com
- This tool is aesthetically pleasing

The user's provided feedback by choosing either of the below options.

- 7 - Strongly Agree
- 6 - Agree
- 5 - Somewhat agree
- 4 - Neither agree nor disagree
- 3 - Somewhat disagree
- 2 - Disagree
- 1 - Strongly disagree

We performed a qualitative evaluation of the result set we received from the survey through seventeen participants and calculated the average rating for each survey question. We can visualize the average feedback rating from the users in the below graph.

From the survey feedback, we can infer that Tool is easy to navigate for users and can provide non-trivial insight to guide an event planner's decision making. Specifically, this Tool expands upon current methods since it offers novel data-driven insights such as new organizational topics/clusters as well as comment analysis metrics.

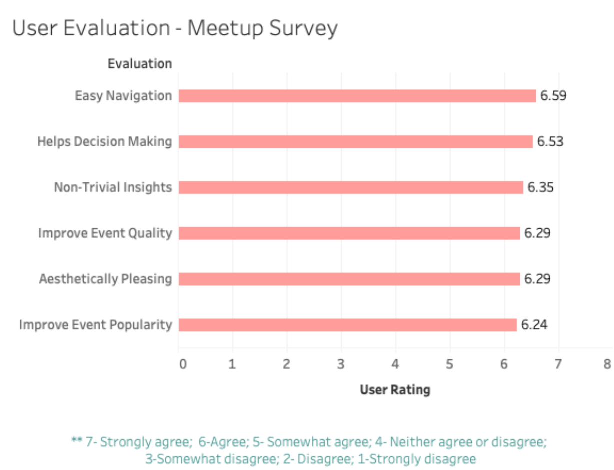

Figure 9: User Evaluation - Online Survey

On the contrary, users less frequently thought that the Tool would directly increase the popularity of a Meetup event. This makes sense, as the tool simply provides event planner with new information. The event planner themselves is responsible for applying these insights to improve their groups and events. Furthermore, based on the survey the participants also believe that the user interface's appearance has room for improvement. In order to improve in this category, it may be beneficial to acquire more detailed feedback since liking the look and feel of an application is more subjective than other categories.

5.3 Topic Extraction Model Evaluation

5.3.1 Experiment. When creating the LDA (latent Dirichlet allocation) topic extraction model, the number of words in a topic as well as the input corpus needed to be chosen in order to optimize the model. In order to select values for these parameters, the model was built, ran, and analyzed using different combinations of the parameters in question, specifically:

Number of words: 1, 2, 3, 4, 5, 6, 7, 8, 9, 10
Input corpus: event name, event description, both

During this experiment, the model's perplexity and coherence score (defined below) were recorded for each test and used to evaluate how the model performed under different parameter configurations.

Perplexity: measure of how well a probability model predicts a value **(lower is better)** [20]
Coherence score: assesses the coherence of a learned topic **(higher is better)** [20]

5.3.2 Results. The results of 10 iterations of this experiment, depicted below, were averaged in order to improve accuracy.

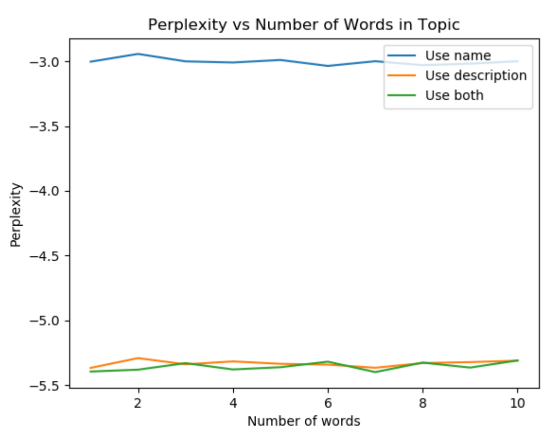

Figure 10: Graph depicting perplexity vs. word count for different input corpora. Lower values indicate a better prediction.

Based on the resulting perplexity and coherence score curves, the topic extraction model is better performing when the input corpus is based off of the event description or both the event name

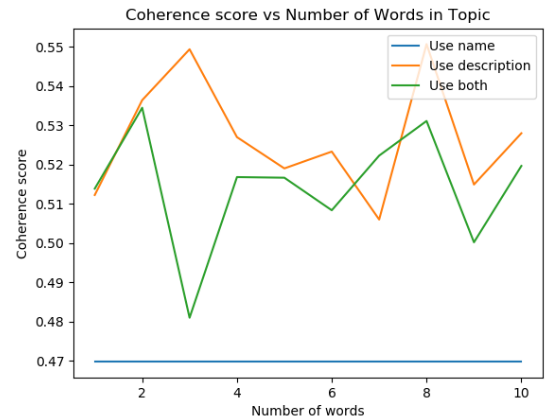

Figure 11: Graph depicting coherence score vs. word count for different input corpora. Higher values indicate better topic coherence.

& description. We therefore chose to use both event name & description as the input corpus to optimize the significance of the model's extracted topics. Word count, on the other hand, does not seem to correlate with perplexity nor coherence score. Thus, we decided to use 1 word per extracted topic for simplicity.

5.4 Division of Work

All team members contributed an equal amount of work.

6 CONCLUSION

Our project aims to alleviate the inherent analytical complexities of event planners when trying to make decisions on hosting an upcoming event. We have collected and analyzed Meetup's API event data in order to create enriched visualizations based on event topics using state of the art text analysis & topic extraction algorithms.

6.1 Future Work

Based on the foundation that our tool was built upon, there are various ways in which it can be expanded in the future.

6.1.1 Data. This project will be further expanded to accommodate the Meetup data from all other cities in the US and significant cities of Non-US countries in the world. Also, real-time data extraction model will be implemented to provide the latest visualization in the meetup visual application.

6.1.2 Analysis. Geographical analysis will be performed to identify which Topic/ Sub-Topic is popular across which region. This could allow high-profile event planners, e.g. from a national or global corporation, to determine which topics would be more or less successful in certain locations or regions.

6.1.3 User Interface. The web application will be tested and launched in various mobile platforms along with login credentials for the event planners to analyze the events and comments. Looking forward, this tool also has the potential to be integrated into the Meetup app and Meetup.com in order to directly provide users with new information & insight in regards to Meetup topics and comments.

6.1.4 Application Delivery. The Application will be delivered to the Meetup Event planners to view the analytical insights of event topics and review the sentiment analysis of the past event comments. Feedback will be collected from event planners for further enhancement on this Application.

REFERENCES

[1] Brandon Rafalson. *Event Marketing – Benchmarks and Trends Report.* https://welcome.bizzabo.com/hubfs/White-Papers-and-Ebooks/Event-Marketing-2019.pdf Bizzabo, 2019.

[2] Chen, Tao, He, Yulan, Wang, Xuan, Xia, Yunqing and Xu, Ruifeng. "Learning User and Product Distributed Representations Using a Sequence Model for Sentiment Analysi". *IEEE Computational Intelligence Magazine* 11, no. 3 (2016): 34-44.

[3] Chung, Taejoong, Han, Jinyoung, Choi, Deajin, Kwon, Ted, Rha, Jong Youn, and Kim, Hyunchul. "Privacy leakage in event-based social networks: A Meetup case study". *Proceedings of the ACM on Human-Computer Interaction* 1, no. CSCW (2017): 35.

[4] Dickson, W. Patrick. "Toward a Deeper Understanding of Student Performance in Virtual High School Courses: Using Quantitative Analyses and Data Visualization to Inform Decision Making". *A synthesis of new research in K-12 online learning.* Michigan Virtual University. April, 2005.

[5] Ding, Hao, Yu, Chenguang, Li, Guangyu, and Liu, Yong. "Event participation recommendation in event-based social networks". *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics).* Vol. 10046 LNCS. Springer Verlag, 2016.

[6] Fang, Xing, and Justin Zhan. "Sentiment analysis using product review data". *Journal of Big Data* Springer Open, 2015.

[7] Hegarty, Mary, Padilla, Lace, Regehr, Sarah, Stefanucci, Jeanine "Decision making with visualizations: a cognitive framework across disciplines". *Cognitive research: principles and implications.* December, 2018.

[8] Najafabadi, Maryam, Villanustre, Flavio, Khoshgoftaar, Taghi, Seliya, Naeem, Wald, Randall Muharemagic, Edin "Deep learning applications and challenges in big data analytics". *Journal of Big Data* Springer Open, 2015.

[9] Kumar, Ravi, Marlow, Cameron, Merugu, Srujana, Olston, Chris, Pang, Bo, Park, Seung-taek, and Reed, Benjamin "Community systems research at Yahoo!". *SIGMOD* Record 36.3, pp. 47-54, 2007.

[10] Liu, Bing "Sentiment Analysis and Subjectivity". *Handbook of Natural Language Processing, University of Illinois at Chicago*, 2010.

[11] Lurie, Nicholas Mason, Charlotte "Visual Representation: Implications for Decision Making". *Journal of Marketing*, 2007.

[12] Lv, Qin and Zhang, Shuo "Event Organization 101: Understanding Latent Factors of Event Popularity". *Eleventh International AAA Conference on Web and Social Media*, 2017.

[13] Müngen, Ahmet Anl and Kaya, Mehmet "A novel method for event recommendation in meetup". *Proceedings of the 2017 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining, ASONAM 2017*, pp. 959-965, 2017.

[14] Mustafizur, A. K.M., Khan, Rahman, Kulik, Lars, and Tanin, Egemen "Location privacy for group meetups". *GIS: Proceedings of the ACM International Symposium on Advances in Geographic Information Systems*, 2016.

[15] Li, Guangyu, Yong Liu, Bruno Ribeiro, and Hao Ding. "On Group Popularity Prediction in Event-Based Social Networks". *International World Wide Web Conference, Lyon, France.* April, 2018.

[16] Pak, Alexander, Paroubek, Patrick "Twitter as a Corpus for Sentiment Analysis and Opinion Mining". *LREc* Université de Paris-Sud, 2010.

[17] Qian, Shengsheng, Zhang, Tianzhu, Xu, Changsheng and Shao, Jie "Multi-Modal Event Topic Model for Social Event Analysis". *IEEE Transactions on Multimedia*, 2016.

[18] Pramanik, Soumajit, Gundapuneni, Midhun, Pathak, Sayan, and Mitra, Bi- vas "Can I foresee the success of my meetup group". *Proceedings of the 2016 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining, ASONAM*, 2016.

[19] Ricken, Stephen, Barkhuus, Louise, and Jones, Quentin "Going on- line to meet offline: Organizational practices of social activities through meetup". *ACM International Conference Proceeding Series Part F128532*, pp. 139 – 148, 2017.

[20] "Topic Modeling with Gensim". *Machine Learning Plus.*