

Conversation Graphs in Online Social Media

Marco Brambilla[✉][0000-0002-8753-2434], Alireza Javadian^[0000-0001-9459-2411],
and Amin Endah Sulistiawati

Politecnico di Milano, Dipartimento di Elettronica, Informazione e Bioingegneria,
Via Giuseppe Ponzio, 34, I-20133 Milano, Italy
`firstname.lastname@polimi.it`

Abstract. In online social media platforms, users can express their ideas by posting original content or by adding comments and responses to existing posts, thus generating virtual discussions and conversations. Studying these conversations is essential for understanding the online communication behavior of users. This study proposes a novel approach to retrieve popular patterns on online conversations using network-based analysis. The analysis consists of two main stages: intent analysis and network generation. Users' intention is detected using keyword-based categorization of posts and comments, integrated with classification through Naïve Bayes and Support Vector Machine algorithms for uncategorized comments. A continuous human-in-the-loop approach further improves the keyword-based classification. To build and understand communication patterns among the users, we build conversation graphs starting from the hierarchical structure of posts and comments, using a directed multigraph network. The experiments categorize 90% comments with 98% accuracy on a real social media dataset. The model then identifies relevant patterns in terms of shape and content; and finally determines the relevance and frequency of the patterns. Results show that the most popular online discussion patterns obtained from conversation graphs resemble real-life interactions and communication.

Keywords: Network Analysis · Conversation Graph · Intent Analysis · Social Media · Instagram · Discourse Analysis · Online Conversation.

1 Introduction

According to Qualman, 2011 [23], the emergence of social media (SM) has profoundly changed the perspective of communication, which resulted in a revolution in the way people interact with each other. As technology grows and expands the range of communication, SM becomes a vital tool for daily social interactions. The interactions can take the form of various activities, like sharing links about interesting content, public updates on the profile such as location data or current activities, and commenting or liking posts and updates.

To leverage SM data benefits as a key to crucial insights into human behavior, many studies such as [4, 7, 15, 18] have been done to perform analysis on SM data by scholars, journalists, and governments. Reasons of people relying on

SM platforms include, but are not limited to, interacting within the inner circle of friendship, entertainment purposes, or subscribing to news; also as presented in various work such as [1, 26], evolving widely for knowledge sharing purpose on online learning and Q&A platforms. Many companies adopt SM to utilize this growing trend to gain business values [13]. Schreck *et al.* [25] discuss how leveraging massive amounts of SM data presents many challenges. The data is multimodal and ambiguous in its content. Communication patterns also change rapidly among various SM elements. This defies choosing proper approaches to handle the systems' complexity. Various methods can be used for understanding complex SM systems [6, 16]. The presence of graph libraries simplifies the intricacy analysis of social networks (SNs), yet the workloads to uncover meaningful values from billions of nodes and vertexes have not diminished.

1.1 Problem Statement

Understanding communication behaviors is an essential awareness. The conversations among SM users are the core of virtual communication that deputizes closely to the real/direct communication. Seeing that most studies on SNs are centralized on a user-to-user relationship, they let through the valuable information from generated conversations in order to conceive online communication behavior. Considering a large dataset from SM platforms with its complex structure, the research questions that lead to this work are as follows:

1. How to build a convenient graph to describe the conversations on SM?
2. How to reconstruct conversations from comments that belong to an SM post that does not follow the *reply* feature?
3. How to assign an appropriate category label to an SM comment that represents the author's intention?
4. What frequent patterns can be found in conversation graphs of online SM?

1.2 Objective

This study proposes a new approach for analyzing online conversations. It consists of two main stages. The first step is **intention analysis** on SM comments reflecting the authors' thoughts. Initially, a list of category names is defined using popular keywords based on set bag-of-words. Then, we perform keyword-based classification to assign a label to each SM comment representing its meaning. Finally, human-in-the-loop techniques are involved in improving the initial keywords. The second stage is **network generation** based on the designed nodes and edges from SM data as well as their attributes. Subsequently, by identifying comments connected by a *reply* edge in the generated network, we automatically generate conversation graphs. Therefore, conversation graphs with labeled comments are produced, portraying patterns of communication behavior between the authors. Finally, we perform statistical and matrix analyses on the conversations. We test the proposed methodology on a real SM event — *YourExpo2015*¹, *i.e.*, a game challenge developed for Expo 2015 Milano event.

¹ <http://www.socialmediaexpo2015.com/yourexp/>

1.3 Contribution

This study is designed for companies or organizations that desire to analyze their audiences' *communication behaviors* on SM platforms. Using text classification on SM comments, we can obtain the *most discussed topics*. Accordingly, exploiting the illustrated comment-to-comment relationships, *patterns from conversation graphs* are gained. Considering the obtained patterns, they can better understand the most *frequent conversations*. Moreover, an *automatic reply* feature is possibly generated based on the analysis result.

The rest of the work is as follows. Sec. 2 discusses the related work. Sec. 3 details the methodology. Sec. 4 presents a set of experiments on a real case dataset. Sec. 5 discusses the results. Finally, Sec. 6 concludes the work.

2 Related Work

2.1 Graph Analysis on Social Network

Myers *et al.* [21] investigate the structural characteristics of Twitter's follow graph to understand how such networks arise. Zhao *et al.* [30] formulate a new problem of specialized finding in Q&A platforms. Buntain *et al.* [8] present an identification method to find a social role based on the user interactions' graph on Reddit. McAuley *et al.* [20] develop a model for detecting circles in ego networks. Rao *et al.* [24] propose a new algorithm for community detection using graph techniques. Yang *et al.* [28] model the statistical interaction between the network structure and the node attributes.

2.2 Conversation Graphs on Social Media

Ning *et al.* [22] utilize graph analysis to better support Q&A systems. Aumayr *et al.* [3] explore classification methods for recovering the reply structures in forum threads. Cogan *et al.* [11] propose a method to reconstruct complete conversations around initial tweets. Zayats *et al.* [29] predict the popularity of comments on Reddit discussions. Kumar *et al.* [17] propose a mathematical model for the generation of basic conversation structure to explore the model humans follow during online conversations. Aragon *et al.* [2] investigate the impact of threading the messages instead of linearly displaying them. Work [12, 14] show how individuals' contribution increases when they feel unique. *Reply* and *Mention* functions can be employed for this purpose.

2.3 Proposed Network Analysis on Conversation Graphs

This study offers a novel network analysis to learn conversation graphs on SM by automatically detected *reply* comments. Besides, we further perform analysis on users' intentions. Note that intent analysis is different from sentiment analysis (*positive*, *neutral*, or *negative* [10]), while the proposed intent analysis studies several classes that are most relevant for the comments. Finally, the constructed conversations with labeled members bring exciting information, such as finding common patterns. To analyze networks, we use SNAP [19] and Gephi [5].

3 Methodology

Fig. 1 illustrates the main methods implemented in this study. Initially, the data gathering from the Internet is constructed to extract data from SM platforms; afterward, the data is stored in the database. Then, we do text processing to perform intent analysis. The next step is to develop a multigraph network’s design to construct conversation graphs.



Fig. 1: High-level overview of the employed methods.

3.1 Text Classification Design

The text preprocessing pipeline consists of two main activities. At first, it applies text cleaning and stemming in order to produce bag-of-words. Then it constructs the TF/IDF to obtain the word/document weight matrix. After preprocessing, as illustrated in Fig. 2, the list of comment categories is initially defined. After we specify the classes’ label, we use keyword-based classification to assign each media comment label. Then, we apply Naïve Bayes and SVM to increase our intent analysis accuracy. Finally, human-in-the-loop is involved for validation.

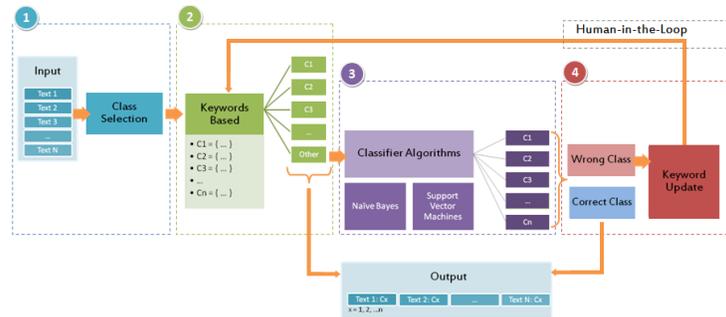


Fig. 2: Intent analysis procedure.

3.2 Network and Conversation Graph Design

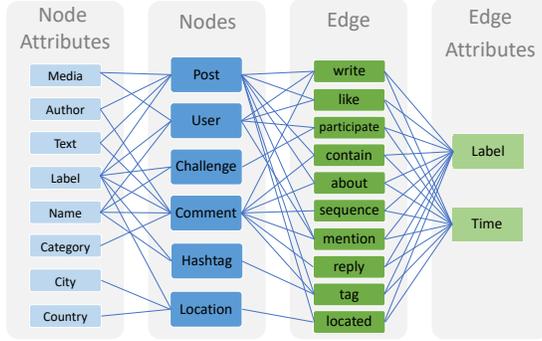


Fig. 3: Network design for Social Media platforms.

In Fig. 3, we present a general SN design representing relationships among all components, such as SM posts, users, comments, locations, *etc.* Fig. 4 displays a graph illustration of a post on SM. The path destination is needed, for instance, to describe the relationship between comment nodes within a conversation and to track which comment’s sequence. This is the reason for designing a directed multigraph for this study. Meanwhile, a multigraph is selected since there are possibly multiple edges connecting two nodes. Attributes of each node and edge from the graph depict the information needed for our analysis. Finally, the generated graph is stored in a graph file to be used for the analysis.

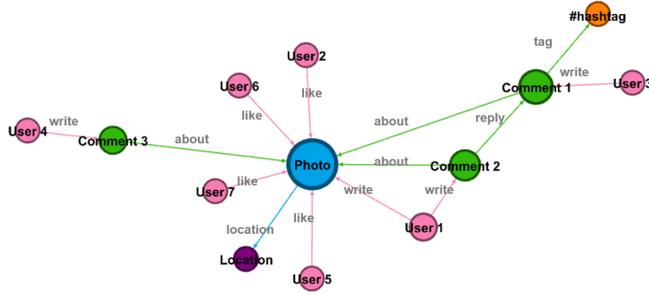


Fig. 4: Graph visualization of a Social Media post from the case study.

To date, most SMs adopt the comment *reply* feature, making it easier to recognize relationships between comments in a post. However, it is possible that the users do not use this feature. They can write a comment intended to reply to the preceding one in a new section as if it is a brand new comment. Also, when a post contains some comments that all together come within a short period and

have the same intention, (e.g., “congratulation”), the author may do not reply to them singly but only reply in a new comment by mentioning those users.

In this study, we design a methodology to recognize a comment that is intentionally linked to the previous comment. The method is described as follows.

1. *User mention recognition*: The idea is to recognize whether a comment has one or more mentioned users. A mentioned/tagged user can be extracted by identifying a term initiated by “@” character in a comment or caption.
2. *Search tagged users*: We examine a list of authors from all comments posted before the current one to find a similar user from the mentioned users list.
3. *Reply assignment*: Once a comment is found that its author mentions in the recent comment, the reply edge is assigned between the two comments.

4 Experiments

4.1 Case Study and Data Collection

Expo 2015, hosted in Milan, Italy, was a universal exposition and a part of the International Registered Exhibition. During the six months of the exhibition, 145 countries participated by running their exhibition. The exhibition successfully attracted more than 22M visitors and derived many marketing campaigns to promote the event. Also, an SM game challenge—*YourExpo2015* was proposed. The game was based on Instagram posts, which are tagged by specific hashtags published every week by Expo 2015. During the whole challenge cycle of nine weeks, more than 15K photos and 600K actions were generated. This study is applied to 15,343 posts containing 98,924 comments related to the challenge.

4.2 Intent Analysis

After applying the text preprocessing steps, we obtained the bag-of-words. By analyzing the most frequent and interesting words, with a subjective assumption, we conclude that the suitable categories for the contents associated with the case study are: *thank*, *congratulation*, *agreement*, *positive*, *invitation*, *food*, *greeting*, *question*, *hashtag*, and *other* which cannot be assigned to any other class.

The initial keywords for each category are constructed based on the obtained bag-of-words. The classification method merely is counting scores for each category’s keywords to the comment collection. This method is a simple approach with a consequence of having several comments (20%) being labeled as *other*. Using direct observation to define keyword-based classification’s ground truth, 100 random samples are chosen for each category to be validated by humans. The average accuracy is 97.5%. So the implementation of keyword-based classification is reliable. The misplace labeling on keyword-based classification is due to the lack of consideration for keywords dependence or context meaning.

Next, comments labeled as *hashtag* and *other* were used as a new dataset to be classified using Naïve Bayes and SVM algorithms. As we are not provided the ground truth, we employed the previous result to train Naïve Bayes and SVM. The result states that SVM outperforms Naïve Bayes with 97.67% accuracy.

4.3 Network Analysis

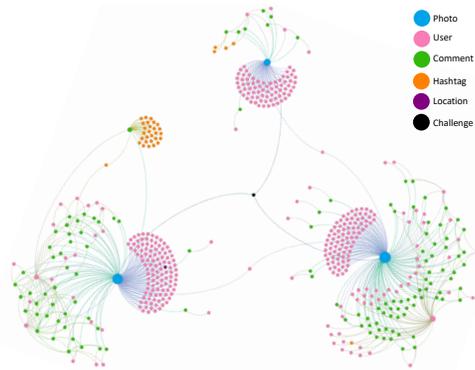


Fig. 5: Graph visualization of three Social Media posts from the case study.

The generated graph is composed of 461,952 nodes and 1,416,751 edges. Fig. 5 presents the visualization of 3 photos. All photos are connected through the challenges node. All nodes are unique, including users. As we can see, a user can publish, like, and comment on more than a photo. Outgoing edges draw user activities; the more outgoing the edges are, the more active the user is.

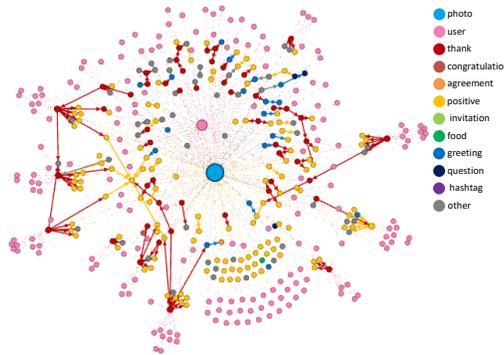


Fig. 6: Visualization of the conversation graphs.

Fig. 6 presents an intent analysis in different colors. Generated relationships inside comments from an Instagram photo portray opinion exchange from the author of those comments. A *reply* edge links two comments. The idea to retrieve conversation graphs is to recognize all connected comments node by *reply* link. From the visualization, we observe that there are some interesting patterns. A node that replies to many comment nodes most likely is a *thank* comment, and a *positive* comment is usually followed either by a *positive* or *thank* comment.

5 Analysis Results

5.1 Statistical Analysis on Conversation

Statistical Analysis The experiment is performed on 15,343 Instagram posts. Table 1 clarifies a statistical analysis of the collection of all comments and retrieved conversations. The number of comments ranges from 0 to 328. If we exclude photos with no comment, the average number of comments is 7. If we include a comment with no relationship with other comments, the maximum number of conversations extracted in all photos is 177. On average, the size of the conversation is 2 nodes. From all conversations in all photos, we obtain that the most extended conversation is the one with the highest size (*i.e.*, 93 nodes).

Table 1: Statistical analysis of comments and conversations.

	# Comments per post (min 1 comment)	# Conversation per post (1 member include)	Size per conversation (1 member exclude)
Mean	7.45	5.10	2.79
Q1	2	2	2
Q2/Median	4	3	2
Q3	8	6	3
Min	1	1	2
Max	328	177	93

Fig. 7 displays the total number of all the conversations. A comment without a relation with others, has the highest frequency. Conversations composed of 2 nodes are the most prevalent ones. The frequency decreases gradually as the size of the conversation increases. Most of the long conversations only occur once.

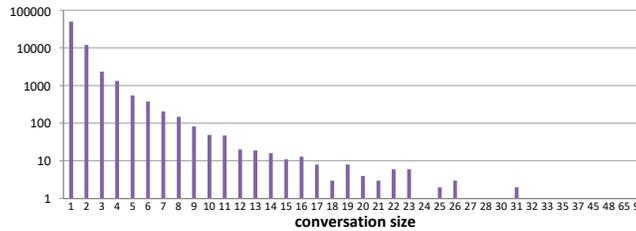


Fig. 7: Frequency for each size of conversation.

Comment Category Distribution Fig. 8 describes the spread of intent categories in the post having at least 30 comments. It shows that *positive* and *thank* comments dominate all conversations. Two other classes that appear almost in all variations of conversation size are *greeting* and *question* types. Comments with invitation and agreement intention are slightly expressed in most conversations, whereas *congratulations* are only mentioned in some discussions.

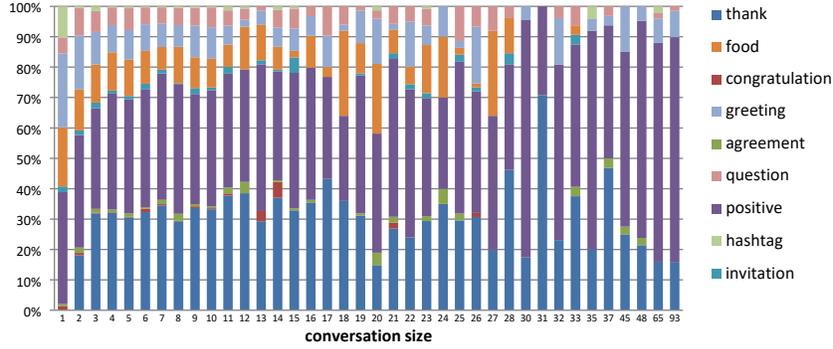


Fig. 8: Distribution of categories on conversations having minimum 30 comments.

As expected, *thank* is not stated in solo conversations, which is most likely in a real discussion. Additionally, *hashtag* comments generally appear in a single comment. In more extended discussions, participants generally talk about *compliments*, *gratitude*, and *salutation*. Considering such online conversations, by investigating the figure, one might conclude that by increasing the conversation size, most of the categories will be dominated by fewer categories. *Food* is the 3rd significant topic; however, it is barely mentioned in extensive conversations. Thus, we investigated photos with 7 to 29 comments.

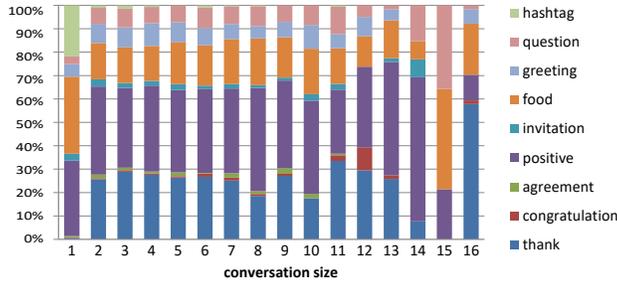


Fig. 9: Distribution of categories on conversations having 7 to 29 comments.

Fig. 9 accounts that the fewer comments, the smaller length of conversation is. *Thank*, *positive*, and *food* categories dominated the overall conversations. Similar to the previous analysis, *agreement*, *congratulation*, and *invitation* categories have low frequency confirming that *hashtag* comments are only written in a single comment. Oppositely, *gratitude* expression is not mentioned in self conversation.

Time Space Analysis The variety in the number of comments for each conversation drives another idea in the time-space analysis. We would like to know if the time and length of conversation are correlated or not.

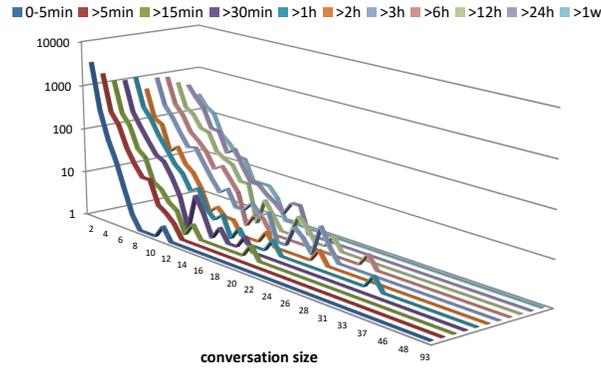


Fig. 10: 3D representation of the conversation size, period, and frequency.

Fig. 10 displays conversation size, period, and frequency. The duration of conversations is calculated by subtracting the time of the latest comment and the first one. Durations range from less than 5 minutes to longer than one week. We were expecting that the smaller conversation takes less time than the longer one. However, the result contradicts our presumption. It visualizes (in logarithmic scale) that generally, conversations occupy a variety of duration. Accordingly, we can conclude that mostly smaller discussions possibly have a longer duration. Conversations with size comments between 2 and 10 have all ranges of duration, while conversations with more than 10 comments tend to narrow the duration.

Fig. 11 shows that long discussions with a conversation size greater than 10 positively do not take a duration of fewer than 15 minutes. It is clearly stated that users involved in the discussion need time to write a comment reply. Another proof states that longer conversations do not take more than 1 day to end the discussion. For instance, a conversation that involves 93 comments takes time between 12 and 24 hours. In conclusion, the small discussions can take a longer time to finish, while more extended ones lean to finish discussion within 24 hours.

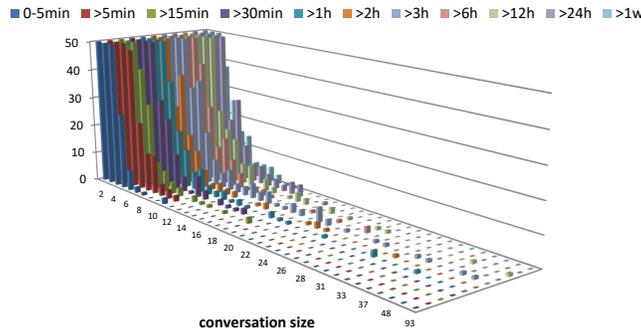


Fig. 11: 3D representation of conversation duration in a smaller frequency range.

5.2 Conversation Patterns Retrieval

The user’s intention in a comment is included in the graphs as a category label. To understand the online communication behavior, we analyze the conversations to retrieve the most frequent patterns generated from intent relationships.

Two-Node Patterns Table 2 illustrates a heat matrix that details the occurrences for each combination of categories. The matrix’s left side represents a comment that replies to a previous comment on the matrix’s top side.

As expected, *thank* → *positive* is the most popular pattern; in other words, a *gratitude* action is generally expressed after a *compliment*. Similar rational behaviors which frequently happened are *thank* → *thank*, *positive* → *positive*, *positive* → *greeting*, *thank* → *invitation*, and so on. These virtual characters imitate real-world communications. It also reveals less popular combinations that most likely do not happen in direct communication, such as agreement after a congratulation or congratulation after someone saying an invitation or even asking a question to someone who gives congratulation. Another less possible pattern is *hashtag* comment used to reply to any other types of comments.

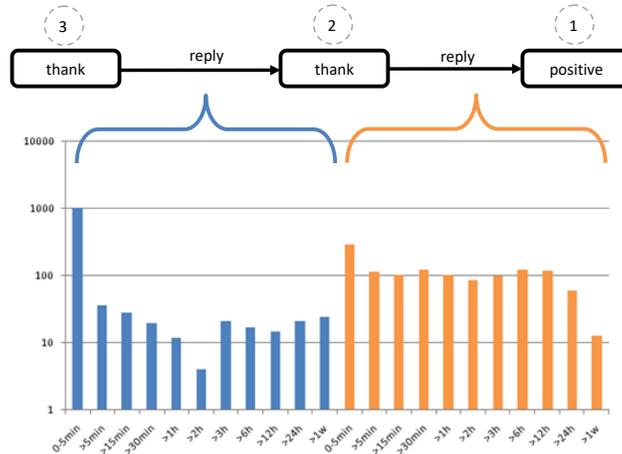
In conclusion, with combinations of all intention labels on the two linked comments, we can obtain digital communication behavior that similarly adopts real-life conversation. Both the most and least popular patterns are likely to happen also in daily communication. Therefore, in the next stage of our analysis, we want to know how far we can expand the length of conversation paths.

Table 2: The frequency of the comment-reply relationship for categories.

		Previous comment								
		thank	positive	food	greeting	question	congrats	agreement	invitation	hashtag
Subsequent comment	replies to ↗									
	thank	1830	9299	1783	1150	397	149	88	790	143
	positive	632	2158	997	439	581	27	73	98	95
	food	247	924	738	203	546	5	24	36	34
	greeting	109	625	180	644	136	8	12	15	13
	question	154	409	279	109	182	1	14	49	26
	congrats	14	37	11	16	7	19	1	1	2
	agreement	21	128	57	37	92	1	10	6	5
	invitation	40	82	54	18	114	1	7	31	6
	hashtag	2	3	0	1	6	0	0	0	0

Three- and Four-Nodes Patterns We take further the analysis patterns into 3 and 4 nodes, and we select the most popular patterns. In this case, we select intent combinations that have more than 1K occurrences. They include *thank* → *positive*, *positive* → *positive*, *thank* → *thank*, *thank* → *food*, *thank* → *greeting*.

The next step is to find the pattern in our conversation graphs for all possible combinations of those 5 schemas by adding another comment category before and after the patterns. The results show that the top pattern is *thank* \rightarrow *thank* \rightarrow *positive*. It replicates direct communication when a person says a complimentary comment; then, the partner replies to express their gratefulness. Afterward, most likely, the first person replies with another gratitude comment. Other popular patterns are reasonable as well. However, the number of occurrences decreases significantly from the most popular one. From the retrieved patterns, we select top ones composed of 3 and 4 nodes to perform temporal analysis and analyze the number of users involved in the discussions. The first analysis seeks to find how long a user takes time to write a reply comment. We pick *thank* \rightarrow *thank* \rightarrow *positive* pattern that has 1,254 occurrences in the conversation graphs. Fig. 12 shows the diversity of reply times. The first part of the chart shows the time needed for the last comment to reply to the previous comment, and the second part is the duration of 2nd comment to reply to the 1st comment.



generally less than 5 minutes. On average, the 2nd comment takes 6 to 12 hours to respond to the 1st one. The 3rd comment requires 30 to 60 minutes to answer the 2nd one, and the 4th comment needs 3 to 6 hours to react to the 3rd one.

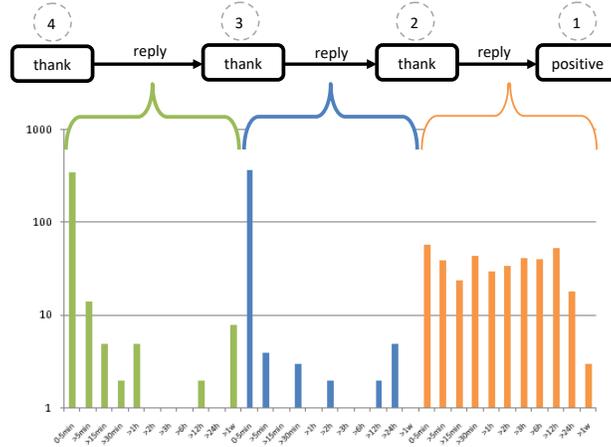


Fig. 13: Reply time in *thank* → *thank* → *thank* → *positive* conversation pattern.

Another thing that interests us is how many users are involved in the conversations. We analyze the top patterns with 3 and 4 nodes. We sum up the number of users that join the discussion. Overall, two users participate in the conversations, and in some cases, 3 and 4 users have taken part in the discussions.

In conclusion, it is a natural behavior that when a compliment is presented at the beginning of the talk, the following responses are all gratitude, and two people are communicating. This generally means that the 1st expresses a positive opinion, then the 2nd expresses gratitude. Then, the 1st responds, and so on.

6 Conclusion and Future Work

This study aims at understanding communication behavior on SM discussions compared to real-life. Intent analysis using keyword-based classification is proposed on SM comments. For the case study, we use Instagram photos of the *Your-Expo2015* challenge. Initially, the approach classifies comments into 9 categories, *thank*, *congratulation*, *agreement*, *positive*, *invitation*, *food*, *greeting*, *question*, and *hashtag*, on each class’s defined keywords. Comments that do not contain any keywords are assigned to the *other* category. Then, we perform Naïve Bayes and SVM on the uncategorized comments. Finally, we perform human-in-the-loop to improve keywords from misclassified comments with the algorithms. In the end, our performance shows a significant result with an accuracy of 98%, with the dominant categories being *compliment* expression and *food* talk.

We also utilized a directed multigraph composed of more than 450K nodes and 1.4M edges representing the collected SM dataset, including intent analysis on the comments. It contains essential information representing relationships among nodes, together with their attribute information. The list of nodes is composed by *posts*, *comments*, *authors*, *locations*, *comments*, and *hashtags*.

A conversation from a post is constructed by identifying the relationships among all comments on an SM post. A virtual discussion is built from one comment that replies to another and analyzes and checks whether other comments are linked as well. Our proposed approach is also able to recognize comment-reply that does not follow the *reply* feature provided on the SM platform.

The analysis of online discussion is not limited to conversation graph retrieval but also understanding users' intentions. Thus, the study's final stage is mining popular patterns of conversation composed of comments with labels. The most popular obtained patterns resemble real-life conversation, where people tend to say *thank* after others say something positive to them. Another observation about the challenge is that most participants are willing to write compliments in the comment section, even when they talk about food.

Future work concerns a more in-depth analysis of mechanisms, particularly in the intent analysis. Even our proposed intent analysis has high accuracy, we perform a plain way to classify SM comments; thus, in the future, we can perform other text classification methods such as [27] to obtain the ground truth. Since we are involved with SM data in which emoticon symbols mostly appear together with text, another work is the learning of emoticon expressions as studied in [9].

References

1. Mushtak Al-Atabi and Jennifer DeBoer. Teaching entrepreneurship using massive open online course (mooc). *Technovation*, 34(4):261–264, 2014.
2. Pablo Aragón, Vicenç Gómez, and Andreask Kaltenbrunner. To thread or not to thread: The impact of conversation threading on online discussion. In *Proceedings of the International AAAI Conference on Web and Social Media*, volume 11, 2017.
3. Erik Aumayr, Jeffrey Chan, and Conor Hayes. Reconstruction of threaded conversations in online discussion forums. *ICWSM*, 11:26–33, 2011.
4. Marco Balduini, Marco Brambilla, Emanuele Della Valle, Christian Marazzi, Tahereh Arabghalizi, Behnam Rahdari, and Michele Vescovi. Models and practices in urban data science at scale. *Big Data Research*, 17:66–84, 2019.
5. Mathieu Bastian, Sebastien Heymann, and Mathieu Jacomy. Gephi: an open source software for exploring and manipulating networks. In *Proc. of 3rd ICWSM*, 2009.
6. Marco Brambilla, Alireza Javadian Sabet, and Marjan Hosseini. The role of social media in long-running live events: The case of the big four fashion weeks dataset. *Data in Brief*, 35:106840, 2021.
7. Giovanni Brena, Marco Brambilla, Stefano Ceri, Marco Di Giovanni, Francesco Pierri, and Giorgia Ramponi. News sharing user behaviour on twitter: A comprehensive data collection of news articles and social interactions. In *Proc. of the Int. AAAI Conference on Web and Social Media*, volume 13, pages 592–597, 2019.
8. Cody Buntain and Jennifer Golbeck. Identifying social roles in reddit using network structure. In *Proc. of the 23rd int. conf. on world wide web*, pages 615–620, 2014.

9. Yoonjeong Cha, Jongwon Kim, Sangkeun Park, Mun Yong Yi, and Uichin Lee. Complex and ambiguous: Understanding sticker misinterpretations in instant messaging. *Proc. ACM Hum.-Comput. Interact.*, 2(CSCW), November 2018.
10. Koyel Chakraborty, Siddhartha Bhattacharyya, and Rajib Bag. A survey of sentiment analysis from social media data. *IEEE Trans. CSS*, 7(2):450–464, 2020.
11. Peter Cogan, Matthew Andrews, Milan Bradonjic, W Sean Kennedy, Alessandra Sala, and Gabriel Tucci. Reconstruction and analysis of twitter conversation graphs. In *Proc. of the 1st ACM Int. Workshop on HotSocial*, pages 25–31, 2012.
12. Tawanna R Dillahunt and Jennifer Mankoff. Understanding factors of successful engagement around energy consumption between and among households. In *Proceedings of the 17th ACM conference on CSCW*, pages 1246–1257, 2014.
13. John Qi Dong and Weifang Wu. Business value of social media technologies: Evidence from online user innovation communities. *J. Strat. Inf. Sys.*, 24(2), 2015.
14. Rosta Farzan, Laura A Dabbish, Robert E Kraut, and Tom Postmes. Increasing commitment to online communities by designing for social presence. In *Proc. of the ACM 2011 conf. on Computer supported cooperative work*, pages 321–330, 2011.
15. Mattia Gasparini, Giorgia Ramponi, Marco Brambilla, and Stefano Ceri. Assigning users to domains of interest based on content and network similarity with champion instances. In *Proc. of the IEEE/ACM Conf. on ASONAM*, pages 589–592, 2019.
16. Alireza Javadian Sabet. Social media posts popularity prediction during long-running live events. a case study on fashion week. 2019.
17. Ravi Kumar, Mohammad Mahdian, and Mary McGlohon. Dynamics of conversations. In *Proc. of the 16th ACM SIGKDD*, pages 553–562, 2010.
18. Linda SL Lai and Wai Ming To. Content analysis of social media: A grounded theory approach. *Journal of Electronic Commerce Research*, 16(2):138, 2015.
19. Jure Leskovec and Rok Sosič. Snap: A general-purpose network analysis and graph-mining library. *ACM TIST*, 8(1):1–20, 2016.
20. Julian McAuley and Jure Leskovec. Discovering social circles in ego networks. *ACM Transactions on Knowledge Discovery from Data (TKDD)*, 8(1):1–28, 2014.
21. Seth A Myers, Aneesh Sharma, Pankaj Gupta, and Jimmy Lin. Information network or social network? the structure of the twitter follow graph. In *Proceedings of the 23rd International Conference on World Wide Web*, pages 493–498, 2014.
22. Ke Ning, Ning Li, and Liang-Jie Zhang. Using graph analysis approach to support question & answer on enterprise social network. In *IEEE APSCC*, 2012.
23. Erik Qualman. *Socialnomics: How social media transforms the way we live and do business*. Wiley, 2012.
24. Bapuji Rao and Anirban Mitra. A new approach for detection of common communities in a social network using graph mining techniques. In *ICHPCA*, 2014.
25. Tobias Schreck and Daniel Keim. Visual analysis of social media data. *Computer*, 46(5):68–75, 2012.
26. Bogdan Vasilescu, Alexander Serebrenik, Prem Devanbu, and Vladimir Filkov. How social q&a sites are changing knowledge sharing in open source software communities. In *Proc. of the 17th ACM conference on CSCW*, pages 342–354, 2014.
27. Baoxun Xu, Xiufeng Guo, Yunming Ye, and Jiefeng Cheng. An improved random forest classifier for text categorization. *JCP*, 7(12):2913–2920, 2012.
28. Jaewon Yang, Julian McAuley, and Jure Leskovec. Community detection in networks with node attributes. In *2013 IEEE 13th ICDM*, pages 1151–1156, 2013.
29. Victoria Zayats and Mari Ostendorf. Conversation modeling on reddit using a graph-structured lstm. *Transactions of the ACL*, 6:121–132, 2018.
30. Zhou Zhao, Furu Wei, Ming Zhou, and Wilfred Ng. Cold-start expert finding in community question answering via graph regularization. In *DASFAA*, 2015.