

People Opinion Topic Model: Opinion based User Clustering in Social Networks

Hongxu Chen
The University of Queensland
Brisbane, Australia
hongxu.chen@uq.edu.au

Hongzhi Yin
The University of Queensland
Brisbane, Australia
h.yin1@uq.edu.au

Xue Li
The University of Queensland
Brisbane, Australia
xueli@itee.uq.edu.au

Meng Wang
Xi'an Jiaotong University
China
wangmengsd@stu.xjtu.edu.cn

Weitong Chen
The University of Queensland
Brisbane, Australia
w.chen9@uq.edu.au

Tong Chen
The University of Queensland
Brisbane, Australia
tong.chen@uq.edu.au

ABSTRACT

Mining various hot discussed topics and corresponding opinions from different groups of people in social media (e.g., Twitter) is very useful. For example, a decision maker in a company wants to know how different groups of people (customers, staff, competitors, etc.) think about their services, facilities, and things happened around. In this paper, we are focusing on the problem of finding opinion variations based on different groups of people and introducing the concept of opinion based community detection. Further, we also introduce a generative graphic model, namely People Opinion Topic (POT) model, which detects social communities, associated hot discussed topics, and perform sentiment analysis simultaneously by modelling user's social connections, common interests, and opinions in a unified way. This paper is the first attempt to study community and opinion mining together. Compared with traditional social communities detection, the detected communities by POT model are more interpretable and meaningful. In addition, we further analyse how diverse opinions distributed and propagated among various social communities. Experiments on real twitter dataset indicate our model is effective.

Keywords

Topic model; Opinion; Community detection; Social network

1. INTRODUCTION

Studying opinion variations among different groups of people is an interesting and significant problem. People recently are inclined to discuss everything and express their opinions on social media platform. They talk about their real life, hot topics, objects in the real world, interested events happened nearby, various services they had experienced. Reported

by Twitter¹ company, monthly there are over 310 million active users using its micro-blogging systems to post and share tweets (texts with 140 words limitation). The huge data generated from social media brings us opportunities to sense individual and collective thoughts towards real world objects.

Therefore, it is significant and a new attempt to study community detection along with opinion mining. In the theory of Homophily [6], individuals in homophilic relationships share common characteristics (beliefs, values, education, etc.) that make communication and relationship formation easier. According to this theory, we come up with an idea that *people hold similar opinions toward common topics are likely to form an opinion based community*. The opinion based community is a group of people who are densely connected and highly consistent in their opinions toward common topics.

For example, during US election, there are two major topics related to Hillary and Trump respectively. Figure 1 illustrates the scenario of US election discussions on social media. Let's assume "topic 1" stands for topics related to Hillary, while "topic 2" represents discussions on Trump. Social media users are connected via links (e.g., followers, friends, replies etc.) and forms a network. In the meantime, there are both agreements and oppositions under each topic - users marked as shaded hold the positive opinion and unshaded marks represent users who hold negative opinions. In many scenarios, we need to perform community detection to find out different user clusters for further analysis.

Figure 1(a) shows the community detection results based on traditional link analysis algorithms [7], [8], [9]. The detected communities are purely based on network structure, neglecting the semantic information. Although users in each community are closely connected, their interested topics are different and associated opinions vary - some people are big fans of Trump, while others are keen to talk about Hillary, and support her.

Figure 1(b) indicates the user clustering results are better and more reasonable when taking topic distribution into account [11], [16]. In this case, topics in each community are more consistent. The left community has common topics on Hillary, the middle and right communities are all interested in topics related to Trump. However, there are over-

©2017 International World Wide Web Conference Committee (IW3C2), published under Creative Commons CC BY 4.0 License.
WWW 2017 Companion, April 3–7, 2017, Perth, Australia.
ACM 978-1-4503-4913-7/17/04.
<http://dx.doi.org/10.1145/3041021.3051159>



¹<https://about.twitter.com/company>

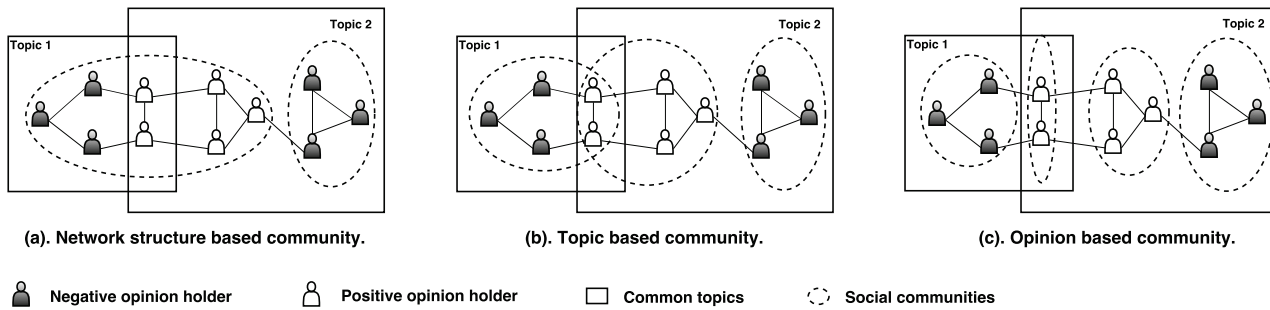


Figure 1: Social communities detected by different approaches.

laps in left two communities, some people are interested in Trump and Hillary at the same time. Moreover, although people in the same community have the common topic, the atmosphere within the community are different because of opposite opinions.

In Figure 1(c), the result we expected based on our opinion based community goes one step further. Users in each detected community are not only closely connected but also share common topics and opinions. For example, users in the most left community are talking about Hillary and support her. By contrast, users in the second left community are all talking about Hillary but with negative opinions. This kind of community results are more valuable and useful when investigating social opinion or monitoring social media.

In this paper, we are focusing on the problem - how to discover different groups of people who have similar opinions towards common topics in social media? Since individual opinion is not sufficient for actions, most opinion mining tasks require studying different opinions from a huge amount of opinion holders. However, opinions are subjective, dynamic, and the vast of opinions make it very difficult for observers to capture. Thus, we have to tackle following challenges:

Sparsity. People’s topics, and opinions on topics are very sparse, so as the interactions among users. This may cause difficulties in clustering similar opinions and tight user relationships, thus leading to deviations in the experiment results. For example, as indicated in Figure 2, the chart shows the topic distribution in a twitter dataset, which contains 41241 users and more than 50 million unique tweets. We extracted hashtags from each tweet and then calculated the frequency for every hashtag. The statistics show that there are about 1.3×10^6 hashtags appear only once in the dataset, while only few topics appear in high frequency.

Dynamic. Individuals’ topics and opinions are drifting over time. When ideas are often changed dynamically in the real world, it will possibly add more uncertainty to the judgement of the polarity of one’ s opinion.

Latency. In twitter, topics and opinions are expressed implicitly. With the latent semantics in short text which are very hard to uncover, another major challenge occurs: how to develop a strong topic model that is capable of capturing identical features for classification.

To tackle above challenges, we propose a generative graphic model based on LDA (Latent Dirichlet Allocation) [1], namely

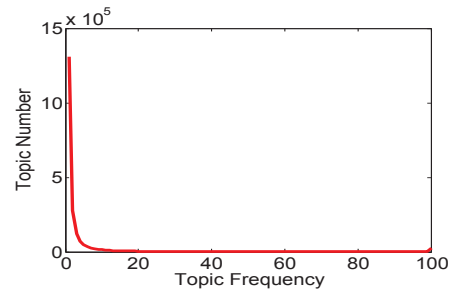


Figure 2: Topic frequency distribution in a twitter dataset.

People Opinion Topic (POT) model. The POT model simultaneously models people’s network structure, topic distribution, and sentiment in a unified way. Further, we also introduce an opinion summary framework that provides concise opinion summary based on community structure, common topics, as well as sentiment polarities. To best of our knowledge, we are the first to study opinion based community detection and propose such an opinion summary framework that enables both quantification analysis and community based opinion analysis on targeted objects in social media.

2. RELATED WORK

The following research directions are related to our problem: community detection or user clustering, topic modelling, sentiment analysis, and opinion summary.

2.1 Community Detection on Social Network

In literature, community detection has been studied from two angles: network structural communities and semantic communities.

2.1.1 Network structural community detection

Community detection based on network structure treats social network as a graph, in which nodes represent the users and edges represent relations (friendship relation, mention relations, etc.) among users. The methods for network structural community detection are based on graph partitioning algorithms, which intends to optimize specific quality metrics (Normalized CUT [9], Modularity [7], etc.). The representative approaches include: KernighanLin partition (Kernighan & Lin, 1970), the spectral bisection method (Pothen, Simon, & Liou, 1990), max-flow min-cut theory (Ford & Fulkerson, 1956), and minimizing conductance cut (Leskovec, Kleinberg, & Faloutsos, 2005). More details about

network structure based community discovery can refer to survey article [8].

2.1.2 Semantic community detection

Different from traditional community detection, another line of this research area takes both network structure and semantic attributes of nodes into consideration to improve community detection performance [4], [17]. The semantic attributes of nodes can be user’s interests, interested topics, common hobbies, occupations and other personal attributes. In [16], the author incorporates community discovery into the topic analysis in text-associated graphs to guarantee the topical coherence in the communities so that users in the same community not only have close ties with each other but also share common topics. An LDA based topic model, Group-Topic (GT) model [11] (Wang, et. al., 2005) was introduced to simultaneously cluster users into groups based on multiple modalities at once. Temporal information is also considered in [3] during detecting communities in social media. Further, another Bayesian model UCGT [14] (Yin, et. al., 2016) was proposed to discover interpretable Geo-Social communities in social media, which simulate the generative process of communities as a result of network proximities, spatiotemporal co-occurrences, and semantic similarity. Also, various topic models are developed and has been well studied [12], [13], [15].

However, both network structure based methods and topic based methods still have limitations in interpretability, robustness. This work is the first to integrate people’s opinion into semantic community detection so that make detected communities more coherence and meaningful and capable of detecting opinion based communities.

2.2 Studying community detection and opinion mining together

The related work [2] that studies community detection and sentiment analysis together on twitter data. The intuition is that simultaneously studying community detection and sentiment analysis could mutually enhance each other. In this paper, social community detection was performed on the friend/follower network of four Microsoft accounts using the Speaker-Listener Label Propagation Algorithm (SLPA) and Infomap algorithm. The initially detected community structure was then enhanced by adding weights on links according to additional features (replies, mentions, retweets, hashtags and corresponding sentiment polarities), which shows a significant result that modularity values were increased for community detection, and more granular, community-level sentiment analysis is enabled by combining these two techniques simultaneously. Moreover, based on statistics on topics (hashtags), this study was able to quantitatively analyse the collective opinion on hotly discussed topics. The results exhibited a significant difference on peoples’ sentiment towards an object between detected community and overall network. However, this research did not resolve the problem of effectively discovering the detailed opinions (aspect-level sentiment analysis) towards particular objects.

Wang et. al. 2016 [10] proposed the concept of sentiment community detection. The proposed sentiment community takes into account both connections and sentiments to discover users who are closely connected and highly consistent in their sentiments about one specific product or service. They adopted the optimization models of semi-definite pro-

gramming (SDP). However, the detected social community results are not dynamic, and only under one specific domain (eg., the social network constructed from one particular movie reviews). Also, in their work, only sentiment labels rather than opinions are considered.

3. OPINION BASED COMMUNITY DISCOVERY

3.1 Problem definition

An opinion is opinion holders’ (can be anyone) sentiment, attitude, emotion about an aspect of an object [5]. We use the term object here to denote opinion target (e.g., a product, service, event, topic, individual, an organization). The sentiment, attitude, emotion typically can be classified into three types: positive, negative and neutral.

Thus, we have the following formal definitions regarding general sentiment analysis:

object. An object o_i in social networks is anything people are talking about. It can be a product, service, person, organization, event or topic. It is the root a hierarchical or a tree structure. An example of an object is shown in Figure 3.

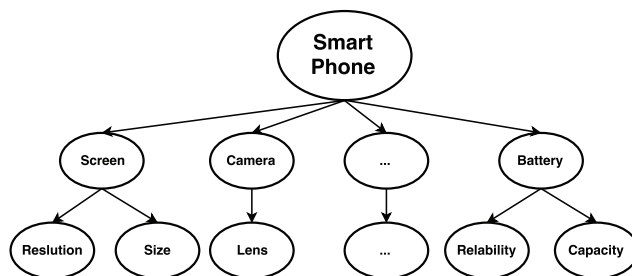


Figure 3: The object and its attributes: A smart phone is an object, and has several attributes like screen, camera, battery, etc..

attribute. An attribute a_{ij} of an object o_i is any non-root node in the tree (tree structure of an object). For example, see Figure 2.1, the screen, camera, and battery are aspects of a smart phone.

opinion. An opinion is a quintuple (p, t, o_i, a_{ij}, s) , where p denotes a person (opinion holder), t is the time when person p express the opinion, o_i is the targeted object, a_{ij} is the j^{th} aspect of o_i , s is the sentiment about aspect a_{ij} of object o_i .

More formally, given a twitter dataset D about targeted objects (such as an organisation, product, service, person, event etc.), a social network can be extracted as a graph $G = (U, E)$, in which U represents users in the social network, E is a set of edges in the network, represents the social relationships between any two users u and w . The textual corpus also can be extracted from individual twitter as $T = \{t_1, t_2, t_3, \dots, t_d, \dots, t_n\}$. Each t_d in T is a set of words, $t_d = \{w_1^d, w_2^d, \dots, w_n^d, \dots, w_{140}^d\}$. Also, each t_d in T has a topic z_d . Assuming that each t_d in T holds a sentiment orientation $s_d, s_d \in \{Positive, Negative, Neutral\}$ towards different topics.

community. A community c is a group of people who have more dense connections within the group than to the rest of the people. For example, in Figure 4, the network forms two communities (shaded nodes 1, 2, 3, 4, 5 and unshaded nodes 6, 7, 8, 9) because the inter-connections within each community are more dense than to the intra-connections between two communities.

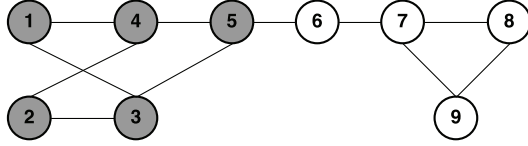


Figure 4: A social network with two communities.

The objective is to discover social communities $C = \{c_1, c_2, c_3, \dots, c_k, \dots, c_m\}$ based on both social network structure and opinion variations posted by each node u in the social graph, such that every node u in each detected community c_k not only closely connects to others, but also shares similar opinions on similar topics.

As social communities are generally more than one and are sometimes overlapping, we treat each community as a multinomial distribution over users. Thus, for each user, the conditional probability $P(u | c)$ measures how likely a user u belongs to the community c . The goal is, therefore, to find out the conditional probability of a user given each community.

Also, for each c_k in C , we give a concise description (community profile) with regard network structure, social properties, interested topics, and opinion summaries. In other words, the problem is how to produce an opinion summary, such that the summary could show what kind of group of social media users (common interests, interested topics) have what attitudes toward different topics about hotly talked objects on the social network.

3.2 POT: People Opinion Topic Model

To discover social communities based on users' opinion, we propose a graphic probabilistic generative model, People-Opinion-Topic (POT), which jointly models users' community, topic, and associated sentiment in a unified manner. It considers the formation of communities as a result of semantic similarity, opinion consistency, and network proximity among social media users.

Figure 5 shows the model structure of POT, and the relevant notations are listed in Table 1. Generally, users have multiple affiliations in the real world. Correspondingly, users in a social network also have multiple community memberships. We associate each user u with a community probability vector θ_c . A community c is assigned to a user u when u expresses opinion o on topic z .

Users within the same communities tend to have the same interests (topics) and share the same opinion orientation. Therefore, we associate each user a latent variable z generated from their interest distribution ϕ_u to indicate one's interested topic. Similarly, when expressing an interested topic z , a user is expressing his/her opinion o towards the topic z . Thus, we also associate each user's an opinion variable o from opinion distribution π_u . Note that in the traditional topic models such as LDA, a document contains a mixture of topics, and each word has a hidden topic label.

NOTATIONS	DESCRIPTION
D	The number of tweets
N	The number of tokens in a tweet T_d
M	The number of social communities
V	The number of users
u_v	v^{th} user
o_d	opinion orientation of a tweet T_d
z_d	topic of a tweet T_d
w_n^d	n^{th} word in tweet T_d
c_k^v	k^{th} community assigned to user u_v

Table 1: Notation of parameters

This is reasonable for long documents. However, the document D in twitter (a short text within 140 words limitation) is usually very short and is most likely to be about a single topic. Thus, in TOP model, all the words in D are assigned with a single topic z , and they are generated from the same word distribution ψ_z .

To easily integrate out $\lambda, \epsilon, \alpha, \psi, \beta$, we adopt conjugate priors in our model. Specifically, we place a Dirichlet prior over each multinomial distribution $(\theta, \Gamma, \psi, \phi)$, and a Beta distribution over the Bernoulli distribution π . The following distributions are drawn:

$$\begin{aligned}
 \Omega | \sigma &\sim \text{Dirichlet}(\sigma) \\
 c_k | \Omega_v &\sim \text{Multinomial}(\Omega_v) \\
 \theta_c | e &\sim \text{Dirichlet}(e) \\
 u_k | \theta_c &\sim \text{Multinomial}(\theta_c) \\
 \pi_u | \beta &\sim \text{Beta}(\beta) \\
 o_d | \pi_u &\sim \text{Bernoulli}(\pi_u) \\
 \phi_u | \alpha &\sim \text{Dirichlet}(\alpha) \\
 z_d | \phi_u &\sim \text{Multinomial}(\pi_u) \\
 \psi_t | \lambda &\sim \text{Dirichlet}(\lambda) \\
 w_n^d | \psi_u &\sim \text{Multinomial}(\psi_u)
 \end{aligned}$$

Based on the model, we obtain the joint distribution of the observed and hidden variables as described in Equation 1.

Consider a user u is a member of an opinion based community c and share common interests and opinions with others in the community. When he/she posts tweets on a specific topic z , he/she first selects the community membership c by his/her community's distribution θ_c . After choosing the community, he/she selects a topic z and opinion o , which are consistent with other group members. With the chosen topic z and opinion o , words a set of words is generated from the topic's word distribution. The generative process in POT model is summarized as following steps:

- Since a user may belong to different communities under the multinomial distribution, a user posts tweets following the atmosphere (similar opinion towards common topics) in their communities.
- When posting a tweet, the user firstly pick up a topic from a multinomial distribution and an opinion orientation from the binomial distribution.
- After the topic and opinion are decided, words are chosen from the topic according to a multinomial distribution.

$$P(\theta_c, \Omega_v, \pi_u, \phi_u, \psi_t, u_v, c_k^v, o_d, z_d, w_n^d | \alpha, \beta, \gamma, \epsilon, \sigma)$$

$$= P(u | c, \theta)P(\theta | \epsilon)P(c | \Omega)P(\Omega | \sigma)P(o | u, \pi)P(\pi | \beta)P(z | u, \phi)P(\phi | \alpha)P(w | o, z, \psi)P(\psi | \lambda) \quad (1)$$

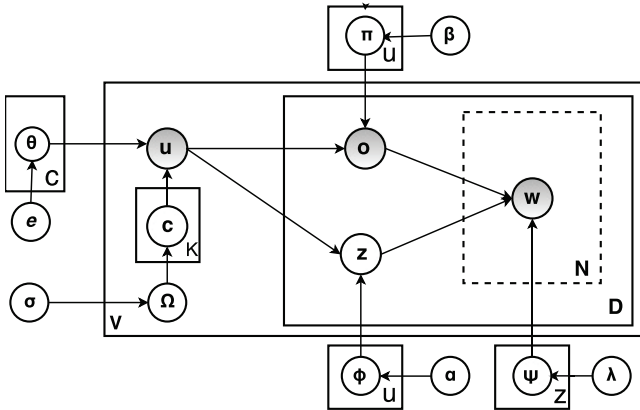


Figure 5: The graphical representation of POT model. Shaded nodes represent observed variables.

4. EXPERIMENTS

The dataset we used for experiments is collected from Twitter by using Twitter API². Based on our problem and application scenario, the data should satisfy the following characteristics: 1), the data (tweets) should be opinionated - tweets should contain sentiment orientations (positive, negative or neutral). 2), the data should have social relations. To this end, we crawled the data starting from some official accounts from a particular organization. From each of these official accounts, we crawled their friends list. In the end, we got a full list of users. After that, we crawled the timeliness (individual collection of tweets with time stamps) for all users in the list. Specifically, we started from 88 official accounts and their 41241 friends. We assume that these 41241 users are mostly related to the same organization or have dense connections, which meet the requirements of our application. Figure 7 shows the social network extracted from our crawled data.

4.1 Qualitative Analysis of Modelling user’s interests

By using our POT model, we have obtained the results of a user’s interested topics. As we can see from Figure 8, there are 100 topics (set as a parameter of POT model) and has been ordered from 1 to 100. The X-axis represents 100 different topics, and Y-axis stands for how many times a person posts tweets related to corresponding topics. Figure 8 shows the user is much more interested in two topics (topic 20 and topic 97, which appears 678 and 221 times, respectively), while other topics are rarely talked. Then, we are

²<https://dev.twitter.com/rest/public>

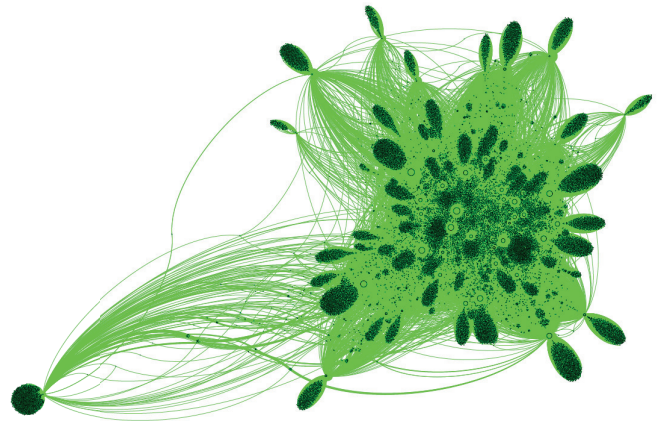


Figure 6: The network constructed from crawled Twitter data.

curious about what topic 20 and 97 is. Then, by looking at the words distribution under each topic in Table 2, which was obtained from results of POT model. Then, we know the user’s interests. Table 2 shows the words distribution of topic 20 and topic 97.

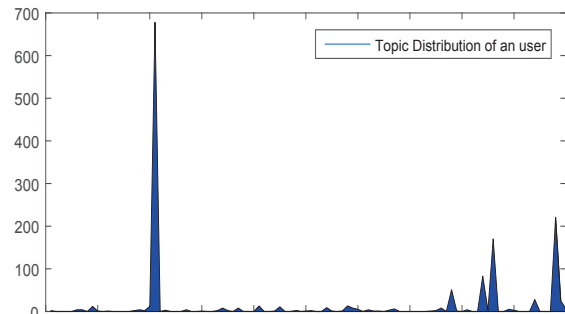


Figure 7: The topic distribution of an user.

To verify the effectiveness of modelling user’s interests, we check the accordance between user’s profile description and calculated frequent topics. we found that the user in Figure 7 is an official account of The University of Queensland. The description in its profile described as “University of Queensland Society of Fine Arts - like Brad Pitt says: ’ Art history. It’s reputable’. ...” and its name is “UQ SoFa”. The profile description is in accordance with the word list under topic 20 and topic 79. For example, words sofa, art, exhibition, gallery, history in topic 20, and words museum, uqartmu-

Topic 20:	Probability	Topic 79:	Probability
sofa	0.028654372	art	0.0313731404
art	0.019463927	museum	0.018811606
cocktail	0.008471429	exhibition	0.008171772
party	0.0072099958	gallery	0.0062718014
exhibition	0.0061287666	uq	0.005511814
sofa's	0.005768357	uqartmuseum	0.005321817
opening	0.0052277427	collection	0.004941823
silent	0.005047538	opens	0.0041818344
auction	0.004867333	prize	0.0041818344
members	0.004687128	bulter	0.0039918376
tickets	0.0041465135	national	0.0038018406
tomorrow	0.0041465135	antiquities	0.0038018406
forget	0.00414665135	artist	0.0036118436
artist	0.0039663087	curator	0.0036118436
gallery	0.0039663087	self-portrait	0.0032318495
night	0.003786104	anthropology	0.0032318495
history	0.0032454894	nspp2015	0.0030418525
happy	0.0028850797	museumweek	0.0030418525
art	0.0028850797	tonight	0.0028518555
semester	0.0027048748	march	0.0026618585
desk	0.0027848748	rex	0.0024718614
lovely	0.0027048748	sally	0.0022818644
everyone	0.00252467	left	0.0022818644

Table 2: Word distribution over topic 20 and 97.

seum, collections, artist, self-portrait, etc. in topic 79 are highly related to arts and uq. These words can very well describe the user’s profile. This result shows that the user’s interests are what he/she talked on twitter, and indicates that our model is capable of discovering user’s interests.

4.2 Experiments on analysing user’s community

Since the created social network (Figure 7) is quite large, and thousands topics and opinions were discussed in this big local community, it is very hard for observers to grab the valuable information about how different group of people thinks one particular thing. Thus, a further experiment was carried out to demonstrate our POT model and opinion based community results.

According to our statistics, see Table 3, it shows the how many times that different topic was discussed in the big community. We only list the most popular topics and corresponding frequency.

Topic	Frequency	Topic	Frequency
auspol	230457	agchatoz	65265
ausvotes	56869	qanda	56480
Australia	51305	climate	49061
science	47984	qldpol	46758
Brisbane	45095	climatechange	44969
jobs	41633	quote	40305
travel	37478	innovation	37472
business	37101	health	36523
job	35292	highered	34558
education	31535	UQ	31494
leadership	31454	art	31402
FF	29379	research	29266
startup	27854	architecture	26879
China	25466	brisbane	25128
Melbourne	25040	tech	24545

Table 3: Topics and frequency in the constructed social network.

As shown in Table 3, we can see the most popular topics are “auspol”, “ausvotes”, “Australia”, “climatechange”, “qldpol”, “science”, “Brisbane”, “UQ”, etc.. To demonstrate the

effectiveness of our proposed opinion summary framework. We select some hot topics, and then use our model and framework to produce opinion based communities. We selected 7 topics to find opinion based communities. These topics are “apple”, “iphone”, “ISIS”, “macbook”, “qldpol”, “qut”, “uq”. The reason why we choose these topics is because these topics are argumentative, hot discussed and carry abundance opinions.

Specifically, we choose two topics “qldpol” and “uq” as examples. “qldpol” is an abbreviation of Queensland Poll, which stands for the topic related to Queensland election. During the election period, people talking about different aspects and policies in the local community. For topic “uq”, it stands for the topic that is talking about a university. When people are talking about a university, various aspects of the university will be discussed.

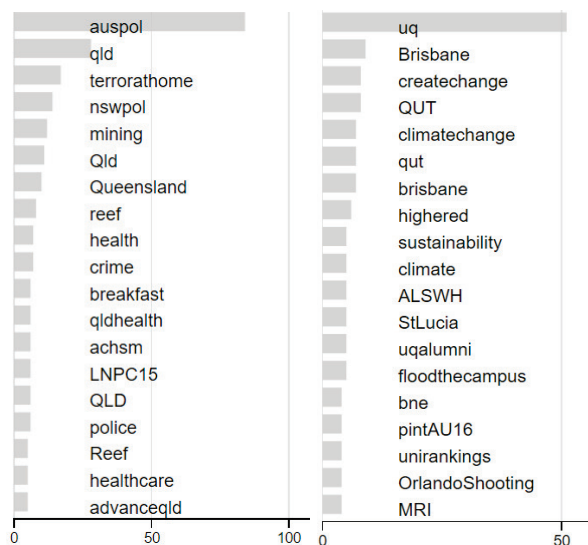


Figure 8: Opinion communities discovered based on two topics.

As shown in Figure 8., two bar charts represent two topics, the top bar in each chart is the topic that was frequently talked in the big communities. The discovered sub-communities are the bars under the top one, and the name of the bar is the sub-topic name (different aspects of the root topic) of each sub-communities. In addition, the number on X-axis indicates the population of each sub-community. Sub-communities represents different group of people talking about different aspects of the root topic. For example, in the big “auspol” community, we want to know what are the most popular topics related to “auspol” (Australia poll), and which group of people talking about it, and what opinion they share in the community. Then, we can see from Figure 9., among the total 100 users who are talking about this big topic, there are 28 people saying “qld” (Queensland), almost 15 people talking about mining, some people talking about health policies, etc.. We do not show the opinion orientation here since the sentiment labels is not accurate at this moment. This is also another problem we will tackle in the future. In the future work, when we obtain good results of sentiment labels, we can finally get the opinion based communities.

5. CONCLUSIONS

In this paper, we solved the problem of finding opinion based community, which is a new concept that requires to study community detection with opinion mining together. We introduced the People-Opinion-Topic (POT) model, which is an LDA based graphical model. We show how POT is useful to detect people's interests, opinion simultaneously. We also introduced a new opinion summary framework that is a community oriented opinion summary framework. The framework is useful when studying public opinions towards certain topics, products, services from different groups of people. However, the opinion based community detection is a new concept in this research area and it is still in its infancy. More work still needs to be done in the future work.

Acknowledgements

This work is partially supported by ARC Discovery Early Career Researcher Award (Grant No. DE160100308) and ARC Discovery Project (Grant No. DP160104075 and Grant No. DP170103954)". The authors would also like to thank the anonymous referees for their valuable comments and helpful suggestions.

6. REFERENCES

- [1] D. M. Blei, A. Y. Ng, and M. I. Jordan. Latent dirichlet allocation. *Journal of machine Learning research*, 3(Jan):993–1022, 2003.
- [2] W. Deitrick and W. Hu. Mutually enhancing community detection and sentiment analysis on twitter networks. *Journal of Data Analysis and Information Processing*, 1:19–29, 2013.
- [3] Z. Hu, C. Wang, J. Yao, E. P. Xing, H. Yin, and B. Cui. Community specific temporal topic discovery from social media. *CoRR*, abs/1312.0860, 2013.
- [4] Z. Hu, J. Yao, B. Cui, and E. Xing. Community level diffusion extraction. In *Proceedings of the 2015 ACM SIGMOD International Conference on Management of Data*, pages 1555–1569. ACM, 2015.
- [5] B. Liu and L. Zhang. A survey of opinion mining and sentiment analysis. In *Mining text data*, pages 415–463. Springer, 2012.
- [6] M. McPherson, L. Smith-Lovin, and J. M. Cook. Birds of a feather: Homophily in social networks. *Annual review of sociology*, pages 415–444, 2001.
- [7] M. E. Newman and M. Girvan. Finding and evaluating community structure in networks. *Physical review E*, 69(2):026113, 2004.
- [8] S. Parthasarathy, Y. Ruan, and V. Satuluri. Community discovery in social networks: Applications, methods and emerging trends. In *Social network data analytics*, pages 79–113. Springer, 2011.
- [9] J. Shi and J. Malik. Normalized cuts and image segmentation. *IEEE Transactions on pattern analysis and machine intelligence*, 22(8):888–905, 2000.
- [10] D. Wang, J. Li, K. Xu, and Y. Wu. Sentiment community detection: exploring sentiments and relationships in social networks. *Electronic Commerce Research*, pages 1–30.
- [11] X. Wang, N. Mohanty, and A. McCallum. Group and topic discovery from relations and text. In *Proceedings of the 3rd international workshop on Link discovery*, pages 28–35. ACM, 2005.
- [12] H. Yin, B. Cui, L. Chen, Z. Hu, and Z. Huang. A temporal context-aware model for user behavior modeling in social media systems. In *Proceedings of the 2014 ACM SIGMOD international conference on Management of data*, pages 1543–1554. ACM, 2014.
- [13] H. Yin, B. Cui, X. Zhou, W. Wang, Z. Huang, and S. Sadiq. Joint modeling of user check-in behaviors for real-time point-of-interest recommendation. *ACM Transactions on Information Systems (TOIS)*, 35(2):11, 2016.
- [14] H. Yin, Z. Hu, X. Zhou, H. Wang, K. Zheng, N. Q. V. Hung, and S. W. Sadiq. Discovering interpretable geo-social communities for user behavior prediction. In *ICDE*, pages 942–953, 2016.
- [15] H. Yin, X. Zhou, B. Cui, H. Wang, K. Zheng, and Q. V. H. Nguyen. Adapting to user interest drift for poi recommendation. *IEEE Transactions on Knowledge and Data Engineering*, 28(10):2566–2581, 2016.
- [16] Z. Yin, L. Cao, Q. Gu, and J. Han. Latent community topic analysis: integration of community discovery with topic modeling. *ACM Transactions on Intelligent Systems and Technology (TIST)*, 3(4):63, 2012.
- [17] Y. Zhu, X. Yan, L. Getoor, and C. Moore. Scalable text and link analysis with mixed-topic link models. In *Proceedings of the 19th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 473–481. ACM, 2013.