International Journal of u- and e- Service, Science and Technology
Vol.8, No.5 (2015), pp.23-36
http://dx.doi.org/10.14257/ijunesst.2015.8.5.03

Applying Overlapping Community Detection Based on Data Field Theory For Twitter Audiences Classification

Liyang Tang[1] and Zhiwei Ni[2]

[1]*Institute of Intelligent Management, School of Management, Hefei University of Technology, Hefei, Anhui 230009, China*
[1] *tangliyang921@gmail.com,* [2] *nzwgd@hfut.edu.cn*

Abstract

Recent years have witnessed the increasing development of Twitter. Given a specific target, all the followers and other participants who have interactions such as mentions with target, are called audiences of the target. As one of the significant applications in social media marketing, classifying Twitter audiences is inherently equivalent to potential customer classification. In this paper, we propose to employ community discovering methods as a solution. On one hand, the significance and influence range of nodes are different, and the influence can be radiated among non-connected nodes as well. To this end, we propose to introduce data field theory to measure the strength between nodes. On the other hand, to minimize parameter selection and settings, we propose a multi-step method by first determining initial clusters based on potentials, and then applying a rough clustering algorithm to detect overlapping communities. Moreover, we collect real world Twitter dataset for the evaluation of proposed method.

Keywords: *Overlapping community detection; Data field; Twitter*

1. Introduction

Recent years have witnessed the increasing development of social media services, which have dramatically changed people's daily lives. For example, in Twitter, user can *follow*, receive updates, and initiate interactions with others. However, even a user is not a *follower*, he/she can still trigger interactions through *mention* mechanism [1]. For ease of description, throughout this paper, the whole set including followers and other participants who have interactions with the target user is called *audiences* of target. In this study, we focus on the classification problem of Twitter audiences for target.

Twitter audiences classification can be significant for Twitter based marketing. For example, by considering not only the fans but also the participants, enterprises can attract potential non-followers and try to convert them into customers. In this case, Twitter audiences classification is equivalent to potential customer classification [2]. Another example is to quantify the proper population of Twitter users to be involved if some marketing activities are launched. Technically, we employ community discovering [3] to cope with Twitter audiences classification.

However, there are several issues along this line. First, in complex networks, the significance of nodes is different, and therefore the influence range of nodes is various as well. For example, some important nodes may have greater radiation intensity upon the surrounding neighborhood nodes. Moreover, in typical social network models, only direct connections are considered. However, uses might be affected by some significant nodes, even though there is no connections between them. To this end, we propose to apply the idea of field theory [4] to model the social network. Instead of only considering direct connections, we also measure the strength of nodes within the range of radiation by calculating the potential value of nodes.

ISSN: 2005-4246 IJUNESST

Second, the accuracy of typical community discovering algorithms relies on the selection of the initial conditions. To minimize the parameter selection and settings, in this paper, we propose to initialize primary clusters by a potential based binary partitioning method.

Third, traditional community discovering algorithms divide the network into separate communities (or clusters or groups), and each member is uniquely belong to one community. However, in real word social networks, there are typically overlapping between communities. For example, user u could be a member of community A, and meanwhile also a member of community B. To capture the overlapping communities, we propose to leverage rough clustering algorithms [5] based on field theory.

The remainder of this paper is organized as follows. In Section 2 we provide some related work. Section 3 presents some preliminaries on data field theory and the overview of proposed method. Section 4 describes the details of four steps for overlapping community detection based on field theory. Then experiments on a real world Twitter dataset are conducted in Section 5. Finally, the paper is concluded in Section 6.

2. Related Work

As the development of microblogs, more and more researchers have noticed the massive potentials and value, and have studied on microblogs from different perspectives. For example, Bernard *et al*. [6] analyzed the ratings of microblog users upon products, and studied the effect of word-of-mouth. Chien-Lung Hsu *et al* [7] investigated the relationship marketing through microblogs based on Commitment-Trust theory. Garcia *et al.* [8] designed a strategy to improve user search by recommending tags to users. Zhao *et al.* [9] studied on why people are using microblogs, and the probability of microblogs to be information exchange medium. Brasoveanu *et al.* [10] discussed how people try to shape and use their avatars in social networks.

Community detection has been a significant social network analysis technique. Basically, there are partition-based methods, modularity optimization based methods, dynamic modeling methods, and spectral analysis based methods [11]. For example, Ning *et al.* [12] proposed an incremental spectral clustering algorithm for discovering communities, to dynamically deal with the complex structure of social networks. Liu *et al.* [13] proposed PCM clustering algorithm for blog community detection, by distinguishing between edge nodes and core nodes. Chen *et al*. [14] proposed a heuristic algorithm to discover overlapping communities.

There also some existing efforts on overlapping communities detection. One category is core based method. For example, Fang *et al*.[15] proposed a core based method, which first finds the seed set using spectral clustering, and then applies random walk technique for extension. Zhang *et al*. [16] proposed another core based method with a novel modularity evaluation function and rough C-means algorithm. The second category is node splitting method. For example, CONGA (Cluster-Overlap Newman Girvan Algorithm) [17] extends GN (Girvan Newman) algorithm [18] for overlapping community detection. Peacock [19] discovered overlapping communities by first transforming the network to a new one by splitting vertices based on split betweenness, and then applying a disjoint community detection algorithm. The last category is optimization based method. For example, Andrea *et al*. [20] proposed a LFM algorithm by optimizing a local fitness function. Huang *et al*. [21] proposed a rough spectral clustering (RSC) algorithm by integrating spectral theory and rough set theory, to optimize the modularity function of overlapping communities. Ball *et al*. [22] employed a statistics method for optimization using Expectation-Maximization (EM) algorithm. Non-negative matrix factorization (NMF) method is also used to discover overlapping communities [23,24].

Unlike existing methods for community detection, in this paper, we introduce a data field based method. Indeed, data field based methods have been applied in some areas. For example, Li Xue *et al*. [25] proposed a rough clustering algorithm based on data field. Li Chunfang *et al*. [26] designed a probabilistic neural network algorithm based on data field. Besides, Dai Xiaojun *et al*. [27] introduced data field for information representation.

3. Preliminary and Overview

3.1. Field Theory Basics

The concept of field was first proposed by Michael Faraday [28], which is used to describe the interactions between the particles of substances. Classical field theory refers to a physical theory which describes how one or more physical fields interact with matter, where a physical field is the assignment of a physical quantity with a value for each point in space and time [4]. Typically, there exists a function defined in certain space which satisfies that for each point in the space, the field strength is proportional to some metric of the reference point (such as the amount of charge and particle mass) and inversely proportional to the distance between that point and the reference point.

Similarly, abstracting the field theory as data field to describe the distribution of some variable or function in the space [29], can help to overcome the restrictions of traditional data mining algorithms which only consider the one-to-one mapping between objects. Instead, data field based method regards objects as individuals with interactions and therefore the status of a data point is the result of the joint action of all other objects.

The influence strength between objects is described via potential function, which is also used to reflect the data distribution by combining the potential function of all objects and determine which cluster specific object belongs to. The clustering rule based on field theory is: the potentials within clusters should be approximately the same, and potentials across different clusters should be apparently different.

Formally, suppose a data space $\Omega \in R^n$ is constructed by $\{x_1, x_2, ..., x_n\}$, where x_i is named as field source. For each data $d_i \in \Omega$, there exist a function f such that data field Φ is generated in Ω:

$$f(d_i, \Omega) \to \Phi . \tag{1}$$

The potential strength of d_i generated at place $d_j \in \Omega$ is calculated as:

$$p(d_i, d_j) = \exp(-\frac{\left\| d_i - d_j \right\|^2}{2\sigma^2}), \tag{2}$$

where σ is an impact factor which is used to control the influence between objects, and $\left\| d_i - d_j \right\|$ is the distance between objects.

Suppose the radiation range around d_i is $U(i)$. Then potential function of d_i is defined as the summation of Equation (2):

$$p(d_i, U(i)) = \sum_{d_j \in U(i)} \exp\left(-\frac{\left\|d_i - d_j\right\|^2}{2\sigma^2}\right). \quad (3)$$

3.2. Method Overview

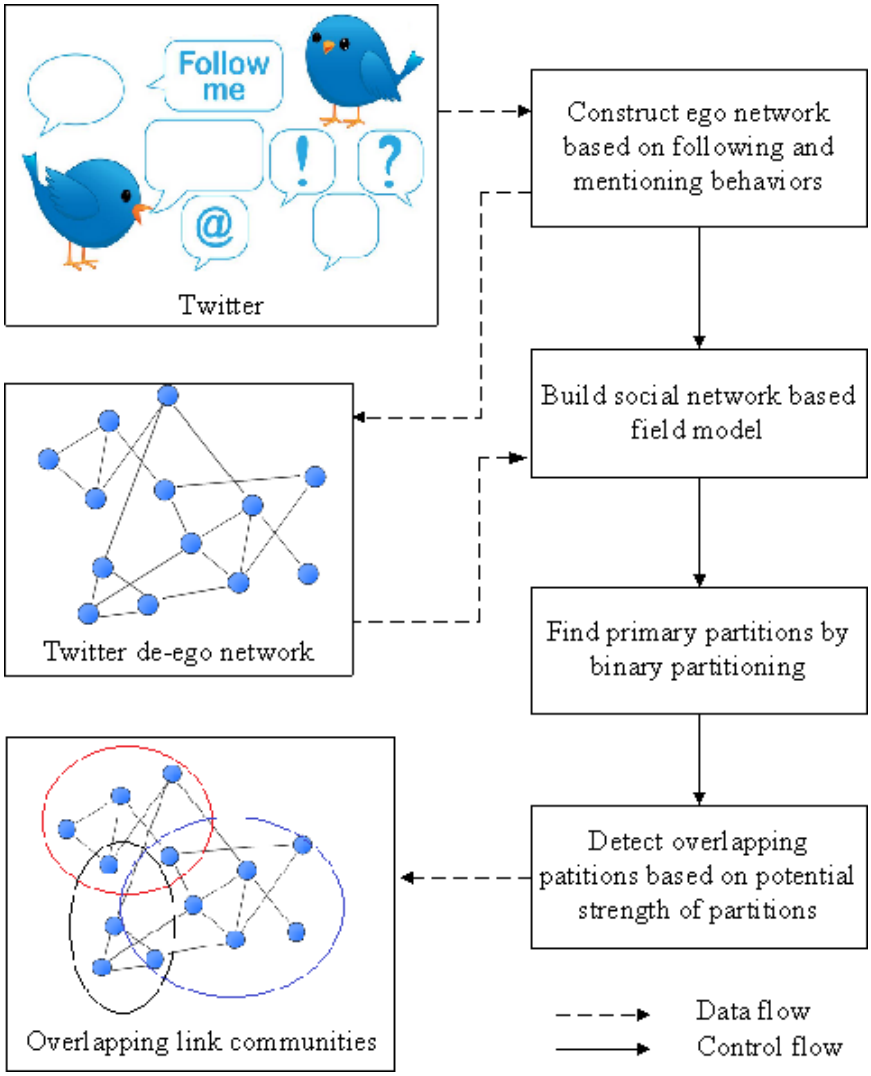

Figure 1. Overview of Proposed Method

In this paper, we are motivated by the problem of Twitter audiences classification for specific target user, and propose a clustering method based on field theory as the solution. The overall process of our proposed method for detecting overlapping link communities in Twitter is shown in Figure 1. First, we construct a network for Twitter audiences by removing the ego node from ego network, which will be discussed in Section 4.1. Second, as discussed in Section 4.2, we model the distance between nodes with consideration of *follow*, *unfollow* and *mention* behaviors, and then apply the field theory by computing the potential for all nodes. Third, we find the primary partitions the in a binary partitioning way based on Kirchhoff laws [30], as presented in Section 4.3. Last, we further revise the previous results by modifying rough clustering algorithm with field theory, to discover

overlapping communities, which is illustrated in Section 4.4.

4. Proposed Method

4.1. Constructing Twitter De-ego Network

Generally, in Twitter, there exist two types of links between Twitter users: (1) follow; (2) reply, retweet, direct message or mention via @ symbol, which are collectively referred to as mention. We use *ego networks* to describe all Twitter audiences and their links.

Definition 1 (Ego network). An ego network $G_e = (V, E, e)$ consists of: (1) an *ego* node e, (2) nodes V (called *alters*) that are directly connected to the ego, and (3) the links E between all nodes.

Figure 2 illustrates a directed ego network, where nodes denote Twitter users, and directed edges denote links between users. In this study, the *ego* node is a specific target such as a brand account, and the links can be following or mentioning behavior. For example, nodes B, D, G are following or mentioning *ego*, and nodes A, C, E, F, H are followed or mentioned by *ego*. Besides, there also exist connections between alter nodes, which means interactions between Twitter audiences are included as well.

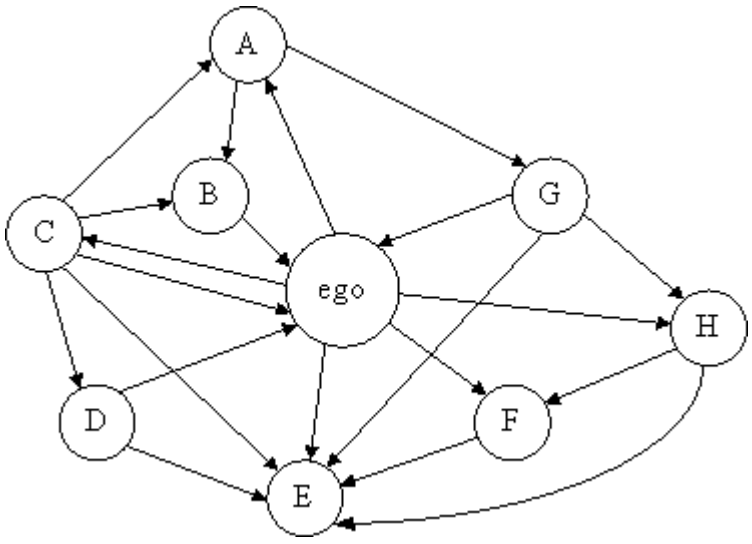

Figure 2. Illustration of a Directed Ego Network G_e

Definition 2 (De-ego network). A de-ego network $G = (V, E)$ is derived from G_e by removing the ego node e and its associated links.

Figure 3 illustrates a de-ego network derived from Figure 2, where only Twitter audiences nodes and the connections between each other are preserved. Indeed, de-ego

network G represents all Twitter audiences for specific target. Therefore, the subsequent steps are performed on G.

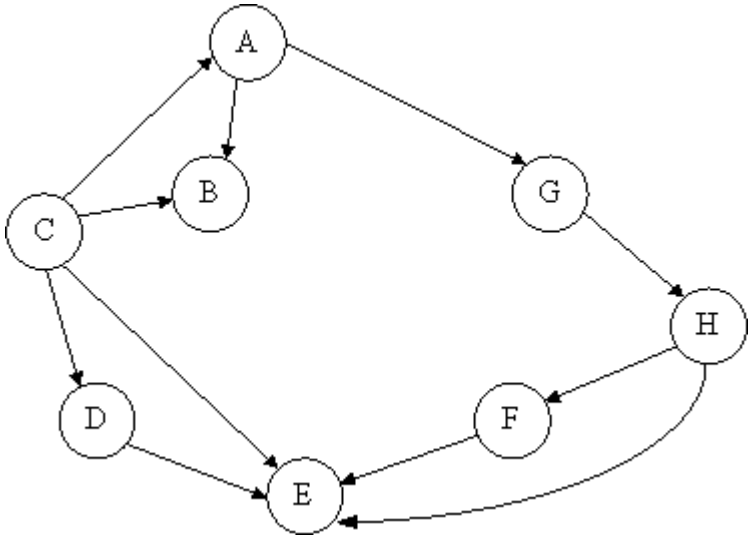

Figure 3. Illustration of De-ego Network G

4.2. Modeling with Field Theory

We employ field theory to model network $G=(V,E)$. Given node $v_i \in V$, and the set of connected neighbors of v_i is $N(i)$. Therefore, based on Equation (3), the potential strength of v_i is calculated as follows:

$$p_i = p(v_i, N(i)) = \sum_{v_j \in N(i)} \exp(-\frac{\|a_{ij}\|^2}{2\sigma^2}) \quad (4)$$

where a_{ij} is the adjacent edge from v_i to v_j, which is defined as:

$$a_{ij} = \begin{cases} a_{ij}+1, & if\ v_i\ follows\ v_j; \\ a_{ij}-1, & if\ v_i\ unfollows\ v_j; \\ a_{ij}+1, & if\ v_i\ mentions\ v_j. \end{cases} \quad (5)$$

Equation (5) shows that we consider three different behaviors in Twitter: follow, unfollow and mention. Follow enables someone to be a fan, while unfollow is the reverse operation. However, unfollowed user can follow again in future if interested. Therefore we simply use +1 and -1 to capture them. Mention behavior includes all kinds of interactions unified via @ symbol, such as replies, retweets, and direct messages. Whenever an interaction happens between v_i and v_j, a_{ij} is increased by 1.

$A = | a_{ij} |_{v_i, v_j \in V}$ is the adjacency matrix of all nodes.

4.3. Finding Primary Partitions using Binary Partitioning

Now we try to find the primary partitions as the initial stage of discovering overlapping communities, which would be discussed in the next section. The basic idea is to first select any random two nodes as poles, and divide G into two parts iteratively.

Before performing binary partitioning, we define *major nodes* as follows.

Definition 3 (Major node). Given a set of potentials of nodes $V = \{v_1, v_2, ..., v_n\}$ and potential strength calculated as Equation (4) $P = \{p_1, p_2, ..., p_n\}$. Sort P in a descending order, denoted as $S = \{p_{(1)}, p_{(2)}, ..., p_{(n)}\}$, where $p_{(i)}$ means the node with i-th higher potential strength. Let the average of S is $\overline{p}$. For each node v_i in S, v_i is *major node* if $p_i \geq \overline{p}$; otherwise, the major node identification process ends.

In order to binary partition G, we randomly select two nodes as poles, and the potentials are initialized as $p_h = 1, p_l = 0$. Then based on Kirchhoff laws [30], the potentials for remaining nodes are calculated as:

$$p_i = \frac{1}{de_i} \sum_{v_j \in N(i)} p_j = \frac{1}{de_i} \sum_{j=3}^{n} p_j a_{ij} + \frac{1}{de_i} a_{i1} \quad (6)$$

where de_i is the degree of node v_i, and a_{ij} is the value from adjacency matrix.

Solve Equation (6), we can get potentials for all nodes. Since the potential decreases rapidly in the junction of two clusters, we use the maximum potential interval in the middle position as the dividing of two clusters:

$$p_d = \max_{p_i, p_j \in (0.5 \pm \alpha)} \{p_i + 1/2 \mid p_j - p_i \mid\} \quad (7)$$

where α is the area selection coefficient, denoting the range of selecting two nodes. Therefore, G is divided into two parts by p_d.

The primary partitions are found by iteratively performing the above operations, as shown in Algorithm 1. After identifying major nodes in Line 1, two nodes are randomly selected to determine the dividing point, as implemented in Lines 2-7. If there is only one major node remained, a partition is discovered, as shown in Lines 8-16. Otherwise, repeat until all nodes are processed.

Algorithm 1 Binary partition based on potential

Input: $G(V, E)$;
Output: divided partitions $D = \{C_1, C_2, ...\}$.

Find all major nodes M, and initiate the set of nodes to be processed as $C = V$;
while S is NULL **do**
  Randomly select nodes v_i, v_j from M, and set potential $p_i = 1, p_j = 0$;
  for each node v_k in $S - \{v_i, v_j\}$: **do**
    Compute potential p_k as Equation (6);
  end for
  Determine the dividing point based on Equation (7), and get D_1, D_2;
  Count the remaining major nodes in C_1, C_2, notated as cnt_1, cnt_2 respectively;
  if $cnt_1 = 1$: **then**
    Add C_1 to D;
    Remove nodes in C_1 from C.
  end if
  if $cnt_2 = 1$: **then**
    Add C_2 to D;
    Remove nodes in C_2 from C.
  end if
end while
return D

Figure 4. Algorithm Description of Binary Partitioning

4.4. Detecting Overlapping Communities

First, we need to calculate the centroids of primary partitions. Suppose we have K initial clusters from previous step, notated as $\{C_1, C_2, ..., C_K\}$. We propose to compute centroids of clusters based on field theory. The basic idea is: for each node in each cluster, calculate the potential within the cluster; and then set the position with average potential as the centroid of that cluster. Indeed, the potential of each node within its belonging cluster represents the strength of that position within the cluster, and the average position is the centroid.

Given node v_i, and its belonging cluster C_i. Based on Equation (3), the potential of v_i within C_i is calculated as follows:

$$p_i = \sum_{v_j \in C_i \wedge j \neq i} \exp(-\frac{\|a_{ij}\|^2}{2\sigma^2}) \quad (8)$$

where a_{ij} is the value from adjacency matrix.

Therefore, the centroid of C_i is calculated as the position with the average potential:

$$\mu_i = \frac{1}{m_i} \sum_{j \in C_i} p_j \quad , \tag{9}$$

where m_i is the number of nodes in C_i.

Next, we employ a rough clustering based method for re-clustering, in order to merge duplicate clusters and discover overlapping clusters. The basic idea is to assign objects to the upper and lower approximation of clusters, and then re-calculate the centroids of clusters, and repeat until the algorithm converges.

The clustering rules with rough theories are as follows:

(1) Each object at most belongs to one lower approximation of cluster;

(2) If an object belongs to a lower approximation of cluster, it also belongs to its upper approximation as well;

(3) If an object doesn't belong to any lower approximation, it must belong to at least two upper approximations.

Obviously, we can see that leveraging rough clustering can help revise the primary partitions and discover overlapping clusters. Denote $\underline{C_i}$ as the lower approximation of C_i, and $\overline{C_i}$ as the upper approximation of C_i. Then the centroid can be updated by:

$$\mu_i = \begin{cases} w_l \dfrac{\sum_{v_j \in \underline{C_i}} v_j}{|\underline{C_i}|} + w_u \dfrac{\sum_{v_j \in \overline{C_i} - \underline{C_i}} v_j}{|\overline{C_i} - \underline{C_i}|}, & \text{if } \underline{C_i} \neq \Phi, \overline{C_i} - \underline{C_i} \neq \Phi; \\ \dfrac{\sum_{v_j \in \underline{C_i}} v_j}{|\underline{C_i}|}, & \text{if } \underline{C_i} \neq \Phi, \overline{C_i} - \underline{C_i} \neq \Phi; \\ \dfrac{\sum_{v_j \in \overline{C_i} - \underline{C_i}} v_j}{|\overline{C_i} - \underline{C_i}|}, & \text{if } \underline{C_i} = \Phi, \overline{C_i} - \underline{C_i} \neq \Phi. \end{cases} \quad , \tag{10}$$

where w_l, w_u are the weights of lower and upper approximations, $\overline{C_i} - \underline{C_i}$ is the boundary between upper and lower approximations, and $|\cdot|$ is the number of objects within each area.

The overlapping communities discovering is described in Algorithm 2. In Line 1, we have K field sources. Lines 2-7, assign each object v_i to the clusters with strongest potential. Later, as shown in Lines 10-15, for any C_j, if it is close enough to the former assignment, assign the object to the lower approximation; otherwise, assign the object to the intersection of two upper approximations. In Lines 16-20, if object v_i belongs to the

intersection of any two uncertain areas, update the assignments. Then, update the centroids as Equation (10), as shown in Line 23. The termination condition of the algorithm is that the discovered clusters remain stable.

Algorithm 2 Discover overlapping communities with rough clustering

Input: the set of objects V, initial K primary clusters and their centroids $\{\mu_1, \mu_2, ..., \mu_K\}$;
Output: overlapping clusters $\{C_1, C_2, ...\}$.

1: Set $\{\mu_1, \mu_2, ..., \mu_K\}$ as initial sources of field;
2: **for** Each object $v_i \in V$ **do**
3: **for** Each μ_j **do**
4: Calculate $p(v_i, \mu_j)$ as Equation (2);
5: **end for**
6: Set $p(v_i, C_l) = max\{p(v_i, \mu_1), p(v_i, \mu_2), ..., p(v_i, \mu_K)\}$;
7: **end for**
8: **while** C_j is not stable **do**
9: **for** Each object $v_i \in V$ **do**
10: Calculate $\Delta = p(v_i, C_l) - p(v_i, C_j)$;
11: **if** $\Delta \geq \alpha$ **then**
12: $v_i \in \underline{C_l}$;
13: **else**
14: $v_i \in \overline{C_l} \cap \overline{C_j}$;
15: **end if**
16: **for** Each cluster C_t **do**
17: **if** $v_i \in (\overline{C_i} - \underline{C_i}) \cap (\overline{C_t} - \underline{C_t})$: **then**
18: Set $p(v_i, C_l) = max\{p(v_i, C_i), p(v_i, C_t)\}$;
19: Set $p(v_i, C_j) = min\{p(v_i, C_i), p(v_i, C_t)\}$;
20: **end if**
21: **end for**
22: **end for**
23: Update centroid μ_j as Equation (10);
24: **end while**
25: **return** $\{C_1, C_2, ...\}$

Figure 5. Algorithm Description of Overlapping Communities Discovering

5. Experiment

In this section, we demonstrate the experiments conducted on a real world Twitter dataset to validate the effectiveness and efficiency of our approach.

Our dataset was collected from Twitter between 28th June and 6th August 2013 using Twitter API [31]. The dataset includes: (1) user profiles, obtained by /users/show; (2) followers list of each user, obtained by /followers/ids; and (3) mentions of specific account, obtained by /statuses/mentions_timeline. Besides, we encrypt all IDs to protect user privacy. We construct de-ego network for each node, and perform proposed overlapping community detection algorithm on all nodes. The average performance of all nodes are regarded as the results.

We employ modularity [3] for measuring the performance of community detection.

Suppose e_{ij} is the edge from the node in community i and the node in community j, and e_{ii} is the edge is the edge within community i. Therefore modularity Q is defined as:

$$Q = \sum_i (e_{ii} - (\sum_j e_{ij})^2) \quad (11)$$

where $\sum_j e_{ij}$ is all edges connected to nodes in community i. As indicated in [18], the structure of communities is relatively distinct when $Q > 0.3$.

We compare our method with Newman algorithm [18] in Figure 6. We can observe that Newman algorithm is basically unaffected by network size, while our proposed method is different. Specifically, when the number of nodes is relatively small, our method is slightly worse than Newman. However, when the size of network growing, it exhibits obvious improvement. Therefore, we can conclude that our method is more suitable for large scale social networks.

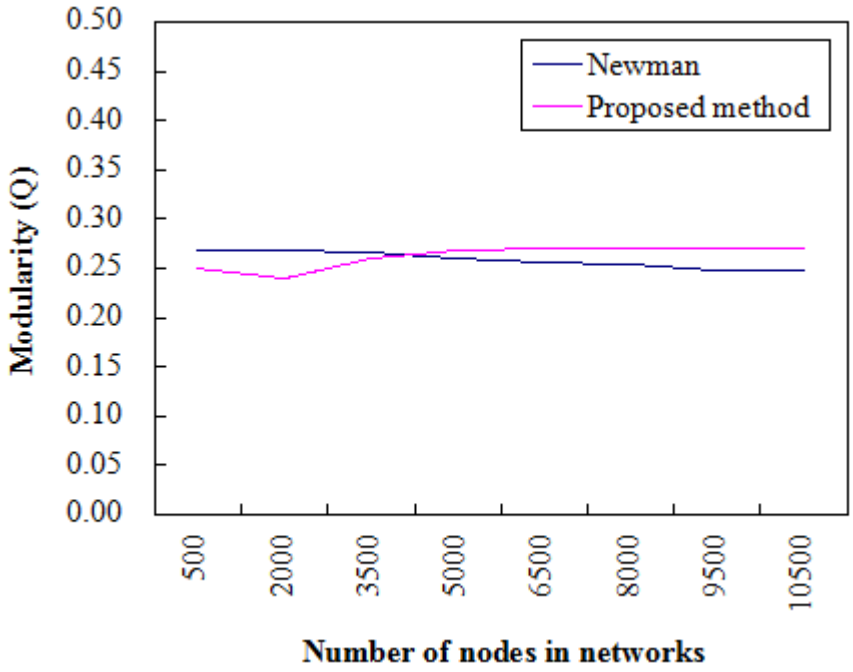

Figure 6. Modularity Comparison with Different Numbers of Nodes

Besides, we also list some of the results in Table 1. We can see that our method is more stable when the size of network increases. For example, the discovered number of communities varies greatly for Newman algorithm, while the number is around 200 for our method. As for time cost, Newman performs better than our method when the network size is small. The reason might be that our method is a multi-step framework, and extra

cost would be spent on details between steps. However, when there are more nodes in network, our method requires less time cost.

Table1. Results of Community Detection of Newman and our Method

Nodes	Algorithm	Communities	Q	Time (s)
2000	Newman	183	0.2673	765
	Our method	162	0.2400	1088
5000	Newman	295	0.2600	3542
	Our method	178	0.2680	3423
8000	Newman	524	0.2545	9881
	Our method	205	0.2711	6335

6. Conclusion

In this paper, we propose to solve the Twitter audience classification problem using community detection, and design an overlapping community detection algorithm by introducing data field theory. Specifically, after constructing de-ego network for each target, primary clusters are found based on nodes potential. Then, based on rough theory, previous results are revised to detect overlapping clusters.

However, there exist some limitations in this study. First, only structural data is used in this work. That is, we use follow, unfollow or mention connections between users, without consideration of context information such as the content of Tweets. Therefore, in this work, we aim to discover link based communities. In future, we will try to integrate context information as well.

Second, in experiments we observe that the modularity is less than 0.3 for both methods, which might be the results of our randomly collected dataset. To further explore the community structure for enterprise accounts to facilitate consumer classification, we would like to monitor specific Twitter data and provide case studies for typical brands in future.

References

[1] L. Y. Tang, "Locating targets through mention in Twitter", World Wide Web, **(2014)**, pp. 1-31.
[2] E. W. T. Ngai, L. Xiu and D. C. K. Chau, "Application of data mining techniques in customer relationship management: A literature review and classification", Expert systems with applications, vol. 36, no. 2, **(2009)**, pp. 2592-2602.
[3] M. E. J. Newman and M. Girvan, "Finding and evaluating community structure in networks", Physical review, vol. E 69, no. 2, **(2004)**, pp. 26-113.
[4] L. D. Landau and E. M. Lifshits, "The classical theory of fields", Butterworth-Heinemann, vol. 2, **(1975)**.
[5] P. Kumar, "Rough clustering of sequential data", Data & Knowledge Engineering, vol. 63, no. 2, **(2007)**, pp. 183-199.
[6] B. J. Jansen, "Micro-blogging as online word of mouth branding", CHI Extended Abstracts on Human Factors in Computing Systems, ACM, **(2009)**.
[7] C. L. Hsu, C. C. Liu and Y. D. Lee, "Effect of commitment and trust towards micro-blogs on consumer behavioral intention: a relationship marketing perspective", International Journal of Electronic Business Management, vol. 8, no. 4, **(2010)**, p. 292.
[8] S. G. Esparza, M. P. O'Mahony and B. Smyth, "Towards tagging and categorization for micro-blogs", **(2010)**.

[9] D. Zhao and M. B. Rosson, "How and why people Twitter: the role that micro-blogging plays in informal communication at work", Proceedings of the ACM international conference on Supporting group work, ACM, **(2009)**.
[10] A. Brasoveanu, "The Avatar in the Context of Intelligent Social Semantic Web", International Journal of Computers Communications & Control, vol. 5, no. 4, **(2010)**, pp. 477-482.
[11] F. L. Huang and N. F. Xiao, "Research on discovering online communities", Nanjing Information Engineering University: Natural Science Edition, vol. 2, no. 1, **(2010)**, pp. 19-25.
[12] H. Z. Ning, "Incremental Spectral Clustering With Application to Monitoring of Evolving Blog Communities", SDM, **(2007)**.
[13] Z. M. LIU, "Blog Community Discovery Based on PCM Clustering Algorithm", Journal of Xiamen University (Natural Science), vol. 4, **(2009)**, p. 11.
[14] D. B. Chen, "Detecting overlapping communities of weighted networks via a local algorithm", Physica A: Statistical Mechanics and its Applications, vol. 389, no. 19, **(2010)**, pp. 4177-4187.
[15] F. Wei, "Detecting overlapping community structures in networks", World Wide Web vol. 12, no. 2, **(2009)**, pp. 235-261.
[16] S. H. Zhang, R. S. Wang and X. S. Zhang, "Identification of overlapping community structure in complex networks using fuzzy c-means clustering", Physica A: Statistical Mechanics and its Applications, vol. 374, no. 1, **(2007)**, pp. 483-490.
[17] S. Gregory, "An algorithm to find overlapping community structure in networks", Knowledge discovery in databases: PKDD, Springer Berlin Heidelberg, **(2007)**, pp. 91-102.
[18] M. Girvan and M. E. J. Newman, "Community structure in social and biological networks", Proceedings of the National Academy of Sciences, vol. 99, no. 12, **(2002)**, pp. 7821-7826.
[19 S. Gregory, "Finding overlapping communities using disjoint community detection algorithms", Complex Networks, Springer Berlin Heidelberg, **(2009)**, pp. 47-61.
[20] A. Lancichinetti, F. Santo and J. Kertész, "Detecting the overlapping and hierarchical community structure in complex networks", New Journal of Physics, vol. 11, no. 3, **(2009)**, pp. 33-15.
[21] F. L. Huang and N. F. Xiao, “Rough Spectral Clustering Algorithm Applied to Overlapping Network Communities Discovery”, Journal of Chinese Computer Systems, vol. 33, no. 2, **(2012)**, pp. 263-266.
[22] B. Ball, B. Karrer and M. E. J. Newman, "Efficient and principled method for detecting communities in networks", Physical Review, vol. E 84, no. 3, **(2011)**, pp. 36-103.
[23] I. Psorakis, "Overlapping community detection using bayesian non-negative matrix factorization", Physical Review, vol. E 83, no. 6, **(2011)**, pp. 66-114.
[24] Y. Zhang and D. Y. Yeung, "Overlapping community detection via bounded nonnegative matrix tri-factorization", Proceedings of the 18th ACM SIGKDD international conference on Knowledge discovery and data mining, ACM, **(2012)**.
[25] X. Li, D. Q. Miao and Q. R. Feng, "Rough clustering algorithm based on data field", Computer Science, vol. 36, no. 2, **(2009)**, pp. 203-206.
[26] C. F. Li, L. Z. Liu and Z. Lu, “Probabilistic neural network algorithm based on data field", Electronic Journal, vol. 39, no. 8, **(2011)**, pp. 1739-1745.
[27] X. J. Dai, "Data field in information Characterization", Journal of Fudan University (Natural Science), vol. 43, no. 5, **(2004)**.
[28] R. D. TWENEY, "2 Inventing the Field: Michael Faraday and the Creative of Electromagnetic Field Theory", Inventive minds: Creativity in technology, **(1992)**, p. 31.
[29] W. Y. Gan, D. Y. Li and J. M. Wang, "Hierarchical clustering method based on data fields", Dianzi Xuebao (Acta Electronica Sinica), vol. 34, no. 2, **(2006)**, pp. 258-262.
[30] J. Gabelli, "Violation of Kirchhoff's laws for a coherent RC circuit", Science, vol. 313, no. 5786, **(2006)**, pp. 499-502.
[31] Twitter API. https://dev.twitter.com/docs/api/1.1.

