

Finding Full Coverage Agent Group (FCAG) of Information Diffusion in Online Social Networks

Shihu An¹

*School of Computer Science and Technology, Shandong University of Finance and Economics
Jinan, Shandong, China, 250014
E-mail: anshihu126@126.com*

Zhikun Zhao

*School of Computer Science and Technology, Shandong University of Finance and Economics
Jinan, Shandong, China, 250014
E-mail: zhaozk@sdufe.edu.cn*

The situation is considered that the information diffusion is limited by the full coverage goal in online social networks. The existence of full coverage agent group of information diffusion is discussed together with the finding algorithm under the community information sharing mode. First, the channel network model of information diffusion is constructed using network model and information diffusion mechanism of community information spread. Second, it is proved that the minimum dominating set is a satisfactory solution to the full coverage problem of information diffusion by means of mathematical reasoning. Finally, a heuristic search method is used to design an algorithm to find the full coverage agent group of information diffusion and the effectiveness of the algorithm is also analysed. Compared with the traditional two step information diffusion mechanism, the effect of the agent group sharing model proposed in this paper is more stable and predictable.

*CENet2017
July 22-23, 2017
Shanghai, China*

¹Speaker

© Copyright owned by the author(s) under the terms of the Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License (CC BY-NC-ND 4.0).

<http://pos.sissa.it/>

1. Introduction

With the rapid development of network technology and the rapid popularization of mobile terminals in recent years, Online Social Network (OSN) has become an important way of information sharing and rapid spread. The relationship structure of OSNs is usually an extension of such two-way intimacy as friends, colleagues, classmates and relatives in the real world, which shows a character of strong relationship and has explicit structural morphology as community or circle of friends [1]. These two circumstances are referred to as community in this paper collectively. Community is the basic place for information diffusion between users in OSN. OSN users can own or belong to a number of communities. They will select a particular community to diffuse information according to their self-calculation [2].

The information diffusion model is divided into one step sharing mode and two step forwarding mode. One step sharing means that a user shares information with other members directly in the community. If a receiver forwards the information he received, two step forwarding takes place [3, 4]. OSNs have the characteristics of community oriented privacy protection for information diffusion and equal exposure opportunity for community member's information diffusion. In other words, in a community, only members can release information and the information sharing is visible to all members.

Graph is generally used to model OSNs, where nodes represent users and edges represents relationships between them. The degree of a node is a basic parameter to describe the local features of the network, which reflects the individual's ability to communicate and possess to network resources [5]. The centrality of a node is the distance between the node and all other nodes, which is used to measure the node's importance in the network structure [6]. Key nodes refer to such nodes that have enlargement effect in the spread of information, ideas and influence in OSNs. The most influential key nodes have the greatest effect on the information dissemination. The information put on these key nodes will be spread quickly [7]. However, finding the k most influential key nodes in an OSN is a NP problem. Kempe and Kleinberg et al. formalized the problem of finding the most influential member set in a social network, and summarized the independent cascade model, the linear threshold model and their generalizations [8]. In the search process for the most influential members, in order to improve the final range of effect and reduce the running time, the characteristics of the linear threshold model should be considered severely, such as influence accumulation, network feature and transmission feature [9, 10].

Existing methods for most influential node set discovery provides valuable reference for OSN information diffusion agent group discovery. However, because the model implies one step sharing mode and two step forward mode, using these methods to choose node set will result to unstable spread and no guarantee on all users seeing specific information [11]. Therefore, the key problem is how to find a least number of information diffusion agents in an OSN to make the information spread from these agents to all members only through one step sharing (no forwarding). This problem is named as the full coverage of information diffusion in OSNs. The corresponding research results can be applied to advertising, government regulation of public opinion, public health, expert discovery and many other applications.

This paper is organized as follows: The full coverage problem of information diffusion in OSNs will be discussed in section 2. The algorithm that finds the full coverage agent group of information diffusion in OSNs is proposed and analyzed in section 3. A case study is presented in section 4. Section 5 provides summary and future works.

The main contribution of this paper is to prove the existence of the full coverage agent group of information diffusion in OSNs and give the corresponding finding algorithm, dealing with the one step sharing mode. Compared with the two step forward mode, the diffusion effect of the one step sharing mode of the full coverage agent group is more stable.

2. The Full Coverage Problem Theory of Information Diffusion

In the full coverage problem theory of information diffusion, the structural data as well as the goal and condition are described, and the existence of feasible solution is proved. The theory provides theoretical support and methodological guidance for the design of the algorithm that finds the full coverage agent group of information diffusion in OSNs

2.1 The OSNs Framework

The framework records the elements and structure of OSNs, which provides the basic data model for the network construction and the agent group finding of information diffusion.

Definition 1. The OSNs Framework

The OSNs framework is defined as the capability of users to communicate with others by their communities, given by a tuple $F = (U, G, h)$, where

U is the user set;

G is the community set;

$h: U \rightarrow 2^G$ is the user-community membership function.

Example 1. The framework of a simple OSN with 5 users and 5 communities is described as (U, G, h) , where

$U = \{u_1, u_2, u_3, u_4, u_5\}$ $G = \{G_1, G_2, G_3, G_4, G_5\}$

$h: h(u_1) = \{G_1\}, h(u_2) = \{G_1, G_2\}, h(u_3) = \{G_1, G_2, G_3\}, h(u_4) = \{G_2, G_4\}, h(u_5) = \{G_5\}$

2.2 The Channel Network of Online Social Information Diffusion

Definition 2. The Channel Network of Online Social Information Diffusion is defined as a relationship network of users that have common community, which is named F-Channel Network in short. Based on the OSN Framework definition F , a F-Channel Network is a labelled undirected graph, given by $G_F = (U, E, L)$, where:

U is the user set;

E is the edge set $\{(u_i, u_j) \mid u_i \text{ and } u_j \text{ have common communities}\}$;

L is the set of edge labels $\{(u_i, u_j) \mid h(u_i) \cap h(u_j)\}$.

Example 2. The corresponding F-Channel Network graph of the OSN in example 1 is shown in Figure 1.

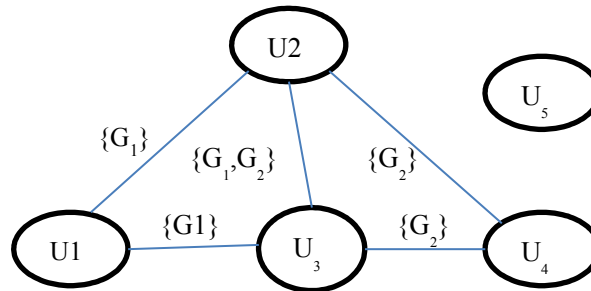


Figure 1 : A F-Channel Network Graph

2.3 The Dominating Set of Online Social Information Diffusion

Definition 3. Dominating Set [12]. Given an undirected graph $G = \langle V, E \rangle$, where V is a set of vertexes and E is a set of edges, S is a subset of V , S is a dominating set if and only if for every vertex v in $V - S$ there is a vertex $u \in S$ making $(u, v) \in E$.

S is a smallest dominating set if it is not a dominating set after removing any element. Among all the dominating sets of graph G , the dominating set that has the minimum number of vertexes is the minimum dominating set. The number of vertexes in the minimum dominating set is the dominating number of graph G .

Example 3. The dominating sets of the graph in Figure 1 are: $\{u_2, u_5\}$, $\{u_3, u_5\}$, $\{u_2, u_3, u_5\}$, $\{u_1, u_4, u_5\}$. Obviously a graph may have multiple dominating sets.

Definition 4. Coverage of Dominating Sets. Given an undirected graph G , the set of all dominating sets is called the coverage of dominating sets of graph G , denoted as $COV(G)$.

Example 4. The Coverage of dominating sets of the graph in Figure1 is $COV(G_F) = \{ \{u_2, u_5\}, \{u_3, u_5\}, \{u_2, u_3, u_5\}, \{u_1, u_4, u_5\} \}$.

2.4 Satisfactory Solution of the Full Coverage Problem of Information Diffusion

Definition 5. Full Coverage Problem of Information Diffusion in OSNs. Given the F-Channel Network $G = \langle U, E, L \rangle$, suppose I is the initial vertex set, $\sigma(I)$ is the reachable vertex set after direct information diffusion of vertexes in I ($I \subseteq \sigma(I)$ because a vertex belongs to its reachable vertex set), the full coverage problem of information diffusion is to find a set $S \subseteq U$ making $\sigma(S) = U$. S is a feasible solution of the problem.

Note: function $\sigma(\cdot)$ satisfies the monotonicity, i.e. for any vertex v and vertex set S , $\sigma(S \cup \{v\}) \supseteq \sigma(S)$. In practice, it means the more users involved in the information diffusion, the more users will receive the information. The number of users who received the information will not reduce at least.

Theorem 1. A dominating set is a feasible solution to the full coverage problem of information diffusion.

Proof:

According to the definition of dominating set (refer to Definition 4), for $\forall v \in U - S$, $\exists u \in S$ satisfies $(u, v) \in E$ if $S \in COV(G)$, i.e. u and v have common communities. If u publish information in the communities, v will receive, namely $v \in \sigma(\{u\})$. Therefore $U - S \subseteq \sigma(S)$, and $(U - S) \cup S \subseteq \sigma(S) \cup S$. The left side of the equation equals U , the right side equals $\sigma(S)$ cause $S \subseteq \sigma(S)$, then the equation becomes $U \subseteq \sigma(S)$. On the other side, $\sigma(S) \subseteq U$ because U is the universal set, therefore $\sigma(S) = U$, i.e. S is a feasible solution of the full coverage problem of information diffusion. Prove finished.

There are multiple feasible solutions to the full coverage problem of information diffusion; the one with the least number of elements has the highest reference value for the decision maker to choose the information diffusion agent group under the same circumstance due to the large scale of OSNs in practice.

Definition 6. The Satisfactory Solution of the Full Coverage Problem of Information Diffusion in OSNs. Given the F-Channel Network $G = \langle U, E, L \rangle$, suppose I is the initial vertex set, $\sigma(I)$ is the reachable vertex set after direct information diffusion, set S is a satisfactory solution of the full coverage problem of information diffusion if $S \subseteq U$ and $\sigma(S) = U$ and $|S|$ is the minimum.

Theorem 2. A minimum dominating set is a satisfactory solution of the full coverage problem of information diffusion in OSNs.

Proof:

Firstly, it is to prove that S is a dominating set if $\sigma(S) = U$.

Suppose $S \subset U$, for any $v \in U - S$, there is $v \in U$ and $v \notin S$. Because $\sigma(S) = U$, $v \in \sigma(S)$, i.e. $\exists u \in S$, $(u, v) \in E$, therefore S is a dominating set.

Secondly, it is to prove that S can only be the minimal dominating set when $|S|$ is the minimum.

According to the definition, S is a minimum dominating set if $\sigma(S) = U$ and $|S|$ is the minimum.

Therefore, a minimum dominating set is a satisfactory solution of the full coverage problem of information diffusion in OSNs. Prove finished.

3. The Algorithm to Find the Full Coverage Agent Group of Information Diffusion

3.1 The Algorithm

The pseudo code is shown in Algorithm 1.

Algorithm 1 (FCAG) . $S = \text{MIN_MAXCOVER}(F)$

Input: the OSN framework F .

Output: Agent Group S .

1. ARRAY Degrees ($|V|$)
2. SET Maxnodes, S
3. $S := \emptyset$; GraphG := $G(F)$;
4. For each $s_i \in \text{GraphG}$
 - If the degree of s_i is 0, then remove s_i from GraphG and insert it into S ;
5. Degrees (i) := the degree of s_i in GraphG;
6. Max0 := the max value in array Degrees;
7. Maxnodes := the vertexes with degree Max0;
8. While GraphG has edges Do {
9. Randomly choose a vertex s_i from Maxnodes;
10. Remove all vertexes connecting with s_i together with their attachment edges from GraphG;
11. Remove s_i from GraphG and insert it into S ;
12. Update Degrees; /*recalculate the degree of each node in GraphG */
13. For each $s_i \in \text{GraphG}$
 - If degrees(s_i)=0, then remove s_i from GraphG and insert it into S ;

3.2 Validity Analysis

In order to prove that the algorithm outputs the full coverage agent group S of information diffusion in OSNs, it is necessary to prove that S has the following properties:

(1) Independence, which means for any $s_i, s_j \in S$, there is $(s_i, s_j) \notin E$.

Use method of reduction to absurdity. Suppose $s_i, s_j \in S$ and $(s_i, s_j) \in E$, according to the algorithm, s_j will be removed from GraphG, therefore $s_j \notin S$. The result is contradicted to the assumption.

(2) Minimum

Maxnodes stores the set of vertexes that have the largest degree in the current situation of the algorithm. S can only be selected from it. Therefore, S is ensured to have the minimum elements.

(3) Dominance

Suppose $s_i \in S$, according to Algorithm 1, all vertexes connecting with s_i will be removed. The removed vertexes belong to $V-S$. The vertexes with zero degree are also inserted into S finally. That is to say, every vertex in $V-S$ is connected with one vertex in S at least. Therefore, the final set S is a dominating set.

4. A Case Study

Example 6. Suppose an OSN is composed of five users $U=\{u_1, u_2, u_3, u_4, u_5\}$ and five communities $G=\{G_1, G_2, G_3, G_4, G_5\}$, the user-community membership function is $h(u_1)=\{G_1\}$, $h(u_2)=\{G_1, G_2\}$, $h(u_3)=\{G_1, G_2, G_3\}$, $h(u_4)=\{G_2, G_4\}$, $h(u_5)=\{G_5\}$, the full coverage agent group of information diffusion is calculated as follows:

The final output agent group is set $S=\{u_5, u_3\}$.

The full coverage scheme of information diffusion in the above OSN is: u_5 and u_3 release information to their respective communities, then all OSN users can see the information via direct sharing (no forwarding considered). $\{u_5, u_3\}$ is the smallest agent group that can achieve information full cover.

5. Summary and Prospect

Aiming at the full coverage of information diffusion, a new model of full coverage agent service for OSNs is proposed. The full coverage problem of information diffusion in OSNs is formalized, including the basic data structure as well as the condition and goal of the problem. The existence of feasible solution to the problem is proved. It provides theoretical support and methodological guidance for the design of the algorithm to find the full coverage agent group of information diffusion in OSNs. Based on the idea of least element heuristic search, the algorithm to find the full coverage agent group of information diffusion in OSNs is designed. The effectiveness of the algorithm is analyzed and the process of the algorithm is illustrated by a case study. The proposed FCAG method is mainly for the indirect community sharing model of information diffusion. In the case of no agent fault, it achieves the full coverage of information diffusion using fewer agents theoretically. The effectiveness of FCAG is proved in this paper, while the efficiency of the algorithm needs further analysis. The solution is based on static network structure, and it is necessary to study the full coverage method of information diffusion in OSNs with highly dynamic structure. The data set used in current OSN study is lack of community information; therefore it is very important for the community communication study to establish OSN data set containing the community information.

References

- [1] L. Wang, X.Q. Cheng, *Dynamic Community in Online Social Networks*, *Chinese Journal of Computers*, 2015, 38(2):219-237.
- [2] L.J. Tabitha, et al. *A Dual Privacy Decision Model for Online Social Networks*[J], *Information & Management*, 2015, 52:893-908.
- [3] S. WU, et al. *Who Says What to Whom on Twitter*[C], in proceedings of the 20th international conference on World Wide Web, ACM, 2011:705-714.
- [4] E. KATZ, *The Two-step flow of communication: An up-to-date report on an hypothesis*[J], *Public Opinion Quarterly*, 1957, 21(1):61-78.
- [5] H.F. Du, et al. *Community structure in small-world and scale-free networks*, *ACTA PHYSICA SINICA*, 2007, 56(12):688-689.
- [6] L. Freeman, *A set of measures of centrality based on betweenness*, *Sociometry*, 1977, 40(1):35.

- [7]X.D.Wu, et al. *Influence Analysis of Online Social Networks*, *Chinese Journal of Computers*, 2014, 37(4):735-752.
- [8]D.Kempe, et al. *Maximizing the spread of influence through a social network[C]*, in proceedings of the 9th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, New York: ACM Press, 2003:237-146.
- [9]D.H.Zhou, et al. *A Fine-grained Information Diffusion Model Based on Node Attributes and Content Features*, *Journal of Computer Research and Development*, 2015, 52(1):156-166.
- [10]J.Tian, et al. *A New Hybrid Algorithm for Influence Maximization in Social Networks*, *Chinese Journal of Computers*, 2011, 34(10):1956-1965.
- [11]Y.Z.Wang, et al. *Local Community Discovery Algorithm Based on Bread-first Search*, *Computer Engineering*, 2015, 41(1):37-41.
- [12]T.W.Haynes, et al. *Fundamentals of Domination in Graphs*, New York, USA.:Marcel Dekker, 1998.