Social Correlation Group Generation Mechanism in Social IoT Environment

Deok-Hee Kang, Hoan-Suk Choi, Woo-Seop Rhee
Department of Multimedia Engineering
Hanbat National University, 34158 Daejeon, Korea
Email: { chatqueue, hkrock7904 }@gmail.com, wsrhee@hanbat.ac.kr

***Abstract*— The Social Internet of Things enables cooperation with devices in various domain networks to provide Internet of Things applications. But, as the number of relationships between humans or devices will increase dramatically, device generates a lot of messages when requesting data. Also, in the centralized manner, the server sends data individually to each device even if they need same data. It causes increasing network traffic. Therefore, the way to minimize discovery area and traffic costs is needed. To solve this problem, we propose the Social Correlation Group. It leads to find required data cost-efficiently by minimizing the discovery area.**

***Keywords*— Internet of things, social networks, correlation value, resource discovery.**

I. INTRODUCTION

The Internet of Things (IoT) integrates a large number of heterogeneous devices generating massive data of the real world. To provide IoT applications, we need various kinds of resources generated and provided by IoT devices (e.g., movie, music, text, sensing data, context information and social information). Also, IoT devices should interact and cooperate each other to share these resources. So, it is important to find proper resources cost-efficiently in the dynamic and decentralized IoT environment [1].

However, the current IoT environment can cooperate only among some of devices in the same domain network [2]. But, IoT devices need to cooperate with other devices in other domain networks according to the requirement of various resources in the future IoT. Therefore, a new paradigm of the Social IoT, assigning logical relationships among devices on the existing physical network, has been introduced [3]. The Social IoT paradigm proposes each social relationship among entities including humans, devices, and resources based on social network of humans. It leads to find required resources based on proposed relationships. So, the Social IoT enhances navigability among devices to find and cooperate each other [4]. But, according to increasing the number of relationships, devices may generate a lot of request messages, and it will cause increasing the discovery costs.

Recently, in the Social Network Service (SNS), humans tend to consume similar resources with friends. For example, a timeline of the Facebook user shows some same resources with a friend if they have many common friends. Likewise, there is high possibility that friends having similar friendship have resources required by the user. However, in the centralized manner, the server sends the same resource individually to many users and it causes increasing network traffic. Also, there exists the redundancy problem by sending common resources to both user and user's friends though they are in same domain network. If friends already have a certain resource, the user can request the resource to own friends directly instead of the server. In the Social IoT, devices also share resources between the relevant devices to minimize the traffic costs. So, we need the way to minimize the discovery area and share resources independently.

Therefore, we propose the Social Correlation Group (SCG) to find target resource in Social IoT Environment. The SCG consists of highly relevant neighbors by comparing social correlation and gives high possibility to exist required resources. Also, it can minimize the discovery area. This paper proposes a definition of the SCG and its generation mechanism.

The rest of the paper is organized as follows. Section II presents some raising research issues and Section III presents our proposed definition and mechanisms about the SCG. In the Section IV, we provide simulation results about resource hit ratio of the SCG and message costs of the proposed mechanism. Finally, we conclude with remarks on future work in Section V.

II. RELATED WORKS

A. Social IoT

Recently, many researches defined the social relationship to manage the Social IoT environment efficiently. The [5] proposed an ontology-based platform for the Social IoT, called Lilliput. It conducted a research on the social relationships among humans, devices, and places and represents them in the form of N-Triples. It leads to infer more complex contexts and authentication of control access using the proposed relationships.

The [6], our previous study, proposed the social relationship model which consists of three entities (i.e., human, device, and service). It generated relations among entities by using devices' profile information. So, these relationships can provide IoT services by cooperating devices without human's intervention. Consequently, we can access to devices with these proposed relationships in this paper.

B. Social interest similarity

In view of the SNS, it is important to figure common interests out among users. It facilitates the interest-based resource recommendation to users by exposing their friends to interesting information. The [7] has studied what kind of features correlates to users' interest similarity to enhance the social-based services. It proved that users who have similar demographic characteristics exhibit more similarity depending on interest domains (i.e., movie, music, TV). Therefore, it showed that the interest similarity of users correlates with many social profiles (e.g., age, gender and geographic location). Also, it revealed that friends are more likely to have same interests on any interest domains. Consequently, it proved that various social features are related to users' interest similarity with disparate effects respectively. By using the interest similarity, we can classify similar neighbors with the requested user to find a resource and access to neighbors for discovering the target resource.

C. Social P2P

The SNSs are faced on rising costs by managing and storing a large amount of resources. To solve this problem, the SocialTube [8], the peer-assisted video sharing service, is proposed. It defined the interest cluster groups called swarm, and shared the chunks of a video in each swarm that has a high probability of owning the target video. Therefore, it can prevent the increasing costs of server storage and large amounts of query flooding. Also, they found that most of users have similar viewing patterns depending on their interests. So, the SocialTube gave the navigability and cost-efficiency in order that they find the target resource easily within the swarm. Likewise, if each entity of Social IoT has similar pattern to require resources, we can set the types of each entity and classify them to provide target resource repeatedly.

III. Social Correlation Group Generation Mechanism

A. Social Correlation Group

In the real-world, humans are interacting with own social friends in SNSs. Also, they tend to have own devices and be surrounded by other devices. Devices can provide required resources to humans. Like the feature of real-world, the Social IoT environment consists of entities, that is a human, a device, and a resource. As shown in the Figure 1, we show the Social IoT graph to represent the Social IoT environment. Each entity in the Social IoT is placed as a node in the graph, and each node is connected by edges. The edge means a social relationship between two nodes. Human nodes are linked by the "friends" relation based on relationships of SNSs. Device nodes are connected with the owner. According to computing abilities or operational purposes, device nodes are classified into a representative device, a sensor and an actuator. The representative device called r-device should have enough computing ability to manage owner's information and current status such as smartphone. And r-device also manages sensors or actuators to coordinate required devices for resources and provide the resources to owner. Therefore, in the social graph, the r-device called r-node is connected through the "own" relation with its owner and the "manage" relation with owner's sensors and actuators. The r-node can represent its owner when requesting some resources.

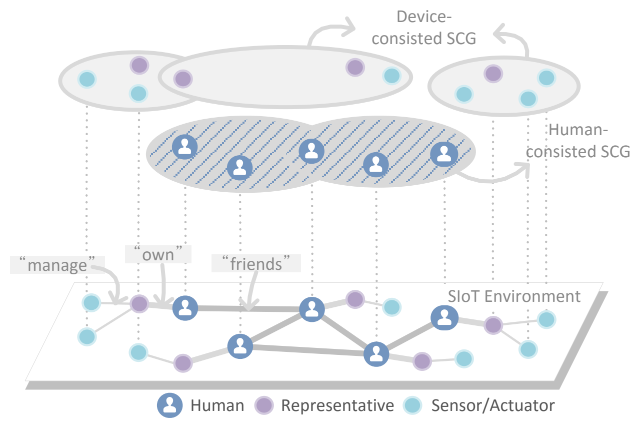


Figure 1. Social Correlation Group in Social IoT Environment

To find the required resource in the SIoT environment, devices are used to generate a lot of messages by broadcasting requests and it causes the flooding problem. It consumes too much of discovery time and costs. Therefore, we propose the SCG as shown in the Figure 1. The SCG consists of highly related nodes by calculating the social correlation between the user and each neighbor. The correlation value reveals how two nodes are related to the required resource based on social information. So, it reveals that highly correlative neighbors are used to have the required resource. Also, as humans and devices have different purposes to use the SCG, we classify the SCG into the Human-consisted SCG and the Device-consisted SCG. The Human-consisted SCG can be generated based on the interest similarity. Humans tend to consume resources such as media, social data in accordance with their interests. But, the interest information is hard to aggregate and compare it because human's interests are changed unexpectedly. According to the [6], the profile similarity is relative to the interest similarity. For example, two humans who are the same age have a high probability that they like same TV programs. So, the Human-consisted SCG is generated with the profile similarity which can estimate the interest similarity. The Device-consisted SCG can be calculated based on each device's ability and condition. Devices tend to find available devices to cooperate with them and provide the requested resource. Therefore, the Device-consisted SCG is classified according to the required functions. For example, when the human wants to light up around his place, the r-device finds available brightness-related devices using the brightness-related SCG. Likewise, various SCGs can be generated depending on each resource domain and facilitate resource discovery to divide each resource area.

B. Social Correlation Group Generation Mechanism

In this section, we propose the SCG generation mechanism that includes the SCG creation, align, and update. As shown in the Figure 2, when (i) the user requests a certain resource, the r-node checks whether own SCG repository has the existing SCG information about the resource. If the r-node has the information, it sends queries about the target resource to the existing SCG. When the r-node has not SCG information, (ii) it requests each node information to neighbors: neighbors consist of the r-node's friends and friend's friends. Neighbors can respond with their own SCG information or node information. After aggregating neighbor's information, the r-node decides to starts (iii) the SCG creation or (iv) the SCG align depending on neighbor's responses. If the r-node receives SCG information from neighbors, the SCG align is progressed. But, if the r-node only receives node information, it should start the SCG creation. When the SCG creation or align is finished, the r-node has generated the SCG and then, (v) it can request the target resource to the SCG. If the SCG member nodes receive the request messages, they send response messages whether to provide the target resource. (vi) The SCG information should be stored to the r-node's SCG repository. If some changes occur in the SCG member's information, (vii) the r-node receives changed information from the SCG member and updates it.

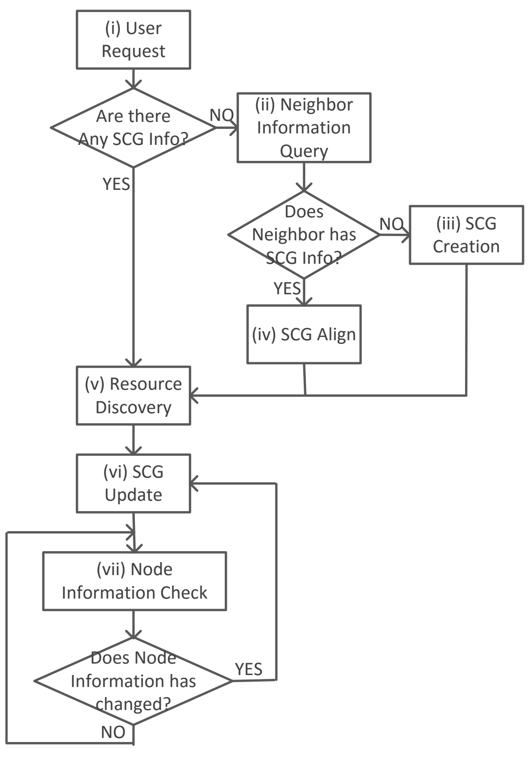


Figure 2. SCG Generation Mechanism

To maintain several required information for SCG generation and reuse the existing SCG, we propose the SCG information format as shown in the Table I. The SCG information is saved in the SCG repository of a node which has generated it. The SCG information format contains 5 types of SCG information: a resource type, a relevant information, a member list, a provider, and a requestor. When the r-node generates the SCG, it stores a target resource type, comparison information types, and member IDs of the SCG. The provider and requestor node information is stored when the r-node requests SCG information or is requested about SCG information: the provider is a node which provides the SCG to the r-node, and the requestor is a node which requests the SCG to the r-node. If the SCG information or provider node information about target resource exists, the r-node can request the target resource repetitively by reusing the existing SCG. Also, with the requestor node information, it can notify the change of own node information to requestors.

TABLE I. SCG INFORMATION FORMAT

Type	Explanation
Resource type	SCG's target resource type
Comparison information type	Relevant node information type to calculate the social correlation
Member list	Node ID lists included in SCG
Provider	Node IDs which have provided SCG to r-node
Requestor	Node IDs which have requested SCG to r-node

1) SCG Creation

Figure 3 shows the SCG creation. The r-node is connected with friend nodes and friends' friend nodes (i.e. neighbor nodes). Friend and friend's friend area consist of each node. When the r-node receives the user's request, (i) it classifies a type of the target resource. (ii) The r-node queries to own neighbors with the resource type and then, neighbors find the relevant information (i.e., profile information or SCG information). (iii) When the r-node has received the responses of neighbors, it aggregates neighbor's profile information and (iv) calculates the social correlation between the user and each neighbor. Finally, the r-node selects some neighbors that have similar interests with the user and (v) creates the SCG area with them.

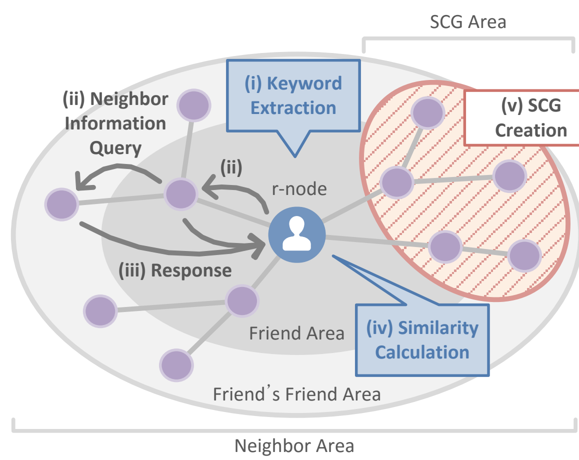


Figure 3. SCG Creation

To compare the social correlation between the user and each neighbor, we conduct the calculation according to the equation (1) [9]. It uses the cosine method to compare social correlation value between two nodes. The information set P_A includes node information of the r-node A and P_B includes node information of r-node's neighbor B. The denominator of Equation (1) means the number of node information of P_A, P_B and numerator means multiplication of the number of same node information.

$$\mathrm{SIM}(A,B) = \frac{\overrightarrow{P_A}\cdot\overrightarrow{P_B}}{|P_A||P_B|} = \frac{\sum P_A \times P_B}{\sqrt{\sum {P_A}^2} \times \sqrt{\sum {P_B}^2}} \qquad (1)$$

Each node information type consists of the relevant profile information as shown in the Table II and III. In the case of human nodes as shown in the Table II, each type is selected and included to the SCG depending on the target resource. For example, the SCG to find one of 'music' resource is calculated by more relevant information (i.e., age, hometown, education degree, work employer). If both information of two users are equal, the selected information is set to 1. So, the correlation of humans can be calculated to apply common information values. The table III shows the device's profile information. The SCG of devices can be calculated depending on the purpose. For example, if the r-device wants to request temperature control, the SCG's purpose is same about the temperature. And other information is set differently depending on each device's condition.

TABLE II. INFORMATION TYPES OF HUMAN NODES

Type	Explanation
Age	Whether two users were born in the same year
Gender	Whether two users have the same gender
Location	Whether two users are close geographically
Language	Whether two users use the same language
Hometown	Whether two users were born in the same place
Education Degree	Whether two users have attended on the same school
Work Employer	Whether two users are working in the same place

TABLE III. INFORMATION TYPES OF DEVICE NODES

Type	Explanation
History	Whether two devices have cooperated for same resource
Type	Whether two devices are the same type
Owner	Whether two devices are belonged to the same owner
Location	Whether two devices are close geographically
Manufacturer	Whether two devices have made by the same manufacturer
Purpose	Whether two devices have the same purpose

2) SCG Align

If some of neighbors already have the SCG information, the r-node aligns existing SCG information to reduce the creation cost by reusing the SCG. When the r-node sends the request messages about neighbor's node information, neighbors check own SCG information whether they already have the SCG information or the provider node information. If a part of neighbors has the existing information, they send the information. Also, if there exists the provider information, each neighbor sends the requests to the provider node for providing the existing SCG. After aggregation of the relevant SCG information, the r-node removes the redundancy by sorting whole information. By using the aligned SCG, the r-node can request the target resource.

Also, as shown in the Figure 4, the r-node can aggregate several SCG information not only the neighbors but also indirect neighbors which are out of range more than 2 hops, therefore, it can extend the discovery area effectively. But, in the case of aligned SCG, it can include less relevant SCG. The provider information makes other nodes even that are less similar provide the requested information. We define this problem as the "unlimited extension problem". It causes that the r-node receives the SCG information created based on provider's similarity. It can be possible that the provider node is less similar with the r-node by aligning without any restriction of the distance. So, some ways to restrict the align range is needed.

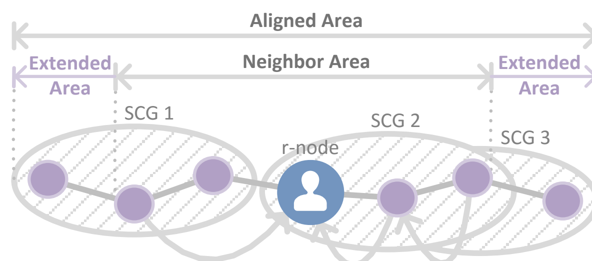


Figure 4. SCG Align

3) SCG Update

Each node updates its existing SCG information whenever node information is changed. If the information is changed, the r-node deletes the existing SCG and creates it newly for the maintenance of correlation. Also, they should notify to the provider or the requestor node to decide whether the node creates the SCG and updates it.

IV. EVALUATION

In this section, we provide two evaluations. First, to reduce discovery costs, we calculate the hit ratio about how many target resources exist in whole area and the SCG area. We compare hit ratio of the SCG to hit ratio of whole area and prove that the node can discover target resource in the SCG more cost-efficiently than whole area. So, we gather friend's profile information (313 humans) and "Like" page information of the one person on the Facebook. With the aggregated profile information, we generate three SCGs about movie, music, and TV domains. And we select one resource respectively in each domain and calculate hit ratio values of whole area and the SCG area: we assume that each "Like" page means a media resource that nodes have stored in their devices. As shown in Figure 5, all resource hit ratios of the SCG on the right sides reveal higher than the whole nodes on the left sides. For example, in the case of the "Jason Mraz", the 6.5% member nodes of the music SCG have the target resource (Jason Mraz). But the 3.8% nodes of whole area have only target resource. Therefore, with same discovery costs, there is higher possibility that node can find target resource in the SCG than whole area.

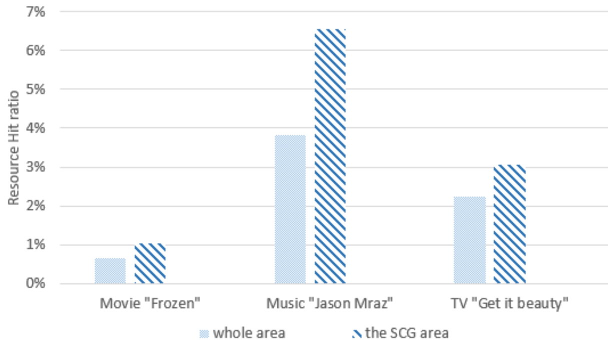
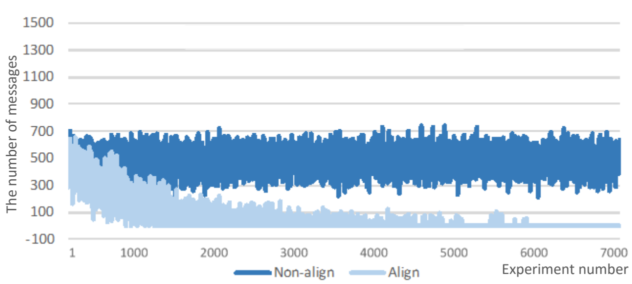


Figure 5. Resource hit ratio with interest domains

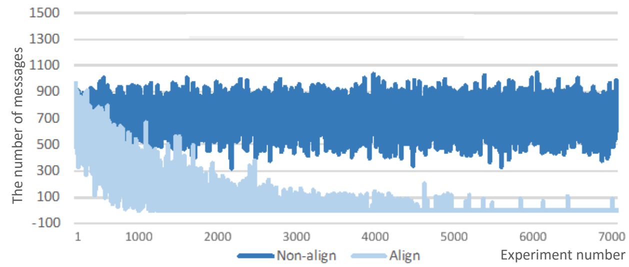
Second, we simulate the proposed SCG align mechanism to reduce SCG creation cost. We use the Facebook dataset provided by the Stanford Network Analysis Platform (SNAP) [10] and select two of them: one of datasets includes information of 768 nodes and 14024 edges and another dataset includes information of 747 nodes and 30025 edges. With two datasets, we compare the creation costs depending on the number of edges. We try the SCG generation of each dataset about 7000 times.

To show the number of request messages for the SCG generation, we select one of nodes randomly and calculate the SCG generation costs of the node as r-node. As shown in the Figure 6-(A) and (B), the number of messages with the align mechanism is approaching to almost zero when the experiment number is over 5000 times. Because, when the r-node has the existing SCG information or provider/requestor node information, it does not need to generate the new SCG and can reuse the aligned SCG. But, without the align mechanism, the number of messages is not reduced because the r-node cannot reuse the existing SCG.

Also, in order that the dataset of the Figure 6-(A) has less number of edges (lack of node connections) than the Figure 6-(B), the result shows maximum number of messages that is 742. But in the Figure 6-(B) contained more edges than the Figure 6-(A), it generates more messages and the result shows maximum number of messages that is 1049. Consequently, the proposed SCG align mechanism is highly cost-efficient if there are a lot of connections between each node in the social network.



(A) 768 nodes and 14024 edges



(B) 747 nodes and 30025 edges

Figure 6. The number of messages with experiment number

V. CONCLUSION

Currently, there are billions of devices in the IoT environments and it has faced a significant issue about the resource discovery among devices. The Social IoT paradigm has been proposed to assign navigability in the existing IoT network: this paradigm connects humans, devices, and resources based on human's social relationships. But, when devices are consistently generating their relationships, it causes the extension of discovery area according to the increasing number of the relationships. Also, when the number of common friends is increasing between the user and friend, they will consume the same resource. So, the server sends resources redundantly to both. The way to reduce the resource discovery area and to share the resource independently is need.

Therefore, we proposed the SCG. Based on the social correlation, it consisted of the selected nodes which has similar node information with the r-node. Also, we provided the generation mechanism to create, align, and update the SCG. It creates the various SCGs to reduce the discovery area. Also, the created SCGs can be aligned to minimize the SCG creation costs and updated for the maintenance of social correlation.

To evaluate how nodes' profile similarity relates to their correlation values, we aggregated 313 human information from the Facebook. Through this evaluation, we showed the result that the resource hit-ratio follows the profile similarity. It means each node tends to have more similar resource if they have similar profile information (e.g., location, age, and gender). Also, we simulated the proposed generation mechanism with two Facebook datasets of SNAP. When the experiment number is over 5000 times, the number of query messages is approaching to almost zero by removing redundancy of the SCG creation. However, we performed two simulations only in terms of the Human-consisted SCG due to lack of device information.

As a future works, we plan to aggregate device's profile (e.g., type, owner, and location) and analyze the performance of the device-consisted SCG. Moreover, we will discover required resources through the SCG and share resources without the server to reduce traffic costs.

Acknowledgment

This research was supported by Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education, Science and Technology (grant number: NRF-2014R1A1A2059500) and the National Research Council of Science & Technology (NST) grant by the Korea government (MSIP) (No. CRC-15-05-ETRI).

References

[1] L. Atzori, A. Iera, and G. Morabito, "The internet of things: A survey," Computer Networks, vol. 54, no. 15, pp. 2787 – 2805, 2010.

[2] L. Atzori, A. Iera, and G. Morabito, "Making things socialize in the Internet—Does it help our lives?," In Kaleidoscope 2011: The Fully Networked Human?-Innovations for Future Networks and Services (K-2011), Proceedings of ITU. IEEE, pp. 1-8, 2011.

[3] L. Atzori, A. Iera, and G. Morabito, "The social internet of things (siot)– when social networks meet the internet of things: Concept, architecture and network characterization," Computer Networks, vol. 56, no. 16, pp. 3594-3608, 2012.

[4] M. Nitti, A. Luigi, and I. P. Cvijikj, "Network navigability in the social internet of things," In Internet of Things (WF-IoT), 2014 IEEE World Forum on. IEEE, pp. 405-410, 2014.

[5] J. Byun, S. H. Kim, and D. Kim, "Lilliput: Ontology-Based Platform for IoT Social Networks," Services Computing (SCC), 2014 IEEE International Conference on. IEEE, pp. 139-146, 2014.

[6] H. S. Choi, D. H. Kang, and W. S. Rhee, "Social IoT Graph Generation Mechanism to provide Autonomous IoT Services," ICCC 2015, pp. 225-226, 2015.

[7] X. Han, L. Wang,S. Park, A. Cuevas, and N. Crespi, "Alike people, alike interests? a large-scale study on interest similarity in social networks," In Advances in Social Networks Analysis and Mining (ASONAM), 2014 IEEE/ACM International Conference on IEEE, pp. 491-496, August, 2014.

[8] H. Shen, Z. Li, Y. Lin, and J. Li, "SocialTube: P2P-assisted video sharing in online social networks," Parallel and Distributed Systems, IEEE Transactions on, vol. 25, no. 9, pp. 2428-2440, 2014.

[9] Wikipedia, Definition of Cosine similarity. [Online]. Available: https://en.wikipedia.org/wiki/Cosine_similarity

[10] SNAP : Stanford Network Analysis Platform, http://snap.stanford.edu/