

# DBSCAN 을 사용한 효과적인 모바일 데이터 오프로딩

김성근\*, 양성봉\*

\*연세대학교 컴퓨터과학과

e-mail : [seungkim1313@yonsei.ac.kr](mailto:seungkim1313@yonsei.ac.kr),

교신저자: [sbyang@cs.yonsei.ac.kr](mailto:sbyang@cs.yonsei.ac.kr)

## Effective Mobile Data Offloading using DBSCAN

SeungKeun Kim\*, Sung-Bong Yang\*

\*Dept. of Computer Science, Yonsei University

### ABSTRACT

Recently, many researchers claim that mobile data offloading is a key solution to alleviating overloaded cellular traffic by dividing the overloaded traffic with femtocells, WiFi networks or users. In this paper, we propose an idea to select a group of users, known as VIPs, that is able to effectively transfer the data to others using Density-Based Spatial Clustering of Application with Noise, also known as DBSCAN algorithm. We conducted our experiments using NCCU real trace dataset. The results show that our proposed idea offload about 70~77% of the network with VIP set size of four, which is better than the compared methods.

### 1. Introduction

Mobile data offloading using opportunistic communication is one of the new network management techniques that able to alleviate overloaded cellular traffic by splitting the work with other users. With this technique, the cellular station is able to lighten the burden by transferring the data to the selected group of users, known as VIPs, and allowing them to directly communicate with others. However, without having a proper way to select VIPs, the result will fluctuate and not be reliable. Thus, define and select the correct group VIPs are the biggest problems that need to be solved.

People often have specific locations where they meet other people with similar interests or characteristics (i.e. co-workers at work, a book-reading club in downtown café). Thus, finding the locations where people with similar interests often meet will greatly help us to predict the future movement of users. Therefore, we first collect contact location of all users and exploit DBSCAN algorithm [1] to classify the data. Since we are clustering the location data, we assume that each classified dataset can form an area, called the hotspot. Then, for each hotspot found, visualize node relationship by constructing a social graph and obtain the most influential user based on the

sum of neighbors' degree.

We applied our proposed solution with real-world mobility trace model called NCCU dataset [2] and evaluate the performance by comparing it with two other VIP selection based on contact frequency and degree centrality. The simulation proves that the performance of our solution is better than two other compared solutions.

### 2. Related Works

While we focus on mobile offloading using opportunistic communication in this paper, there are two other techniques that help to alleviate overloaded cellular traffic: femtocell and WiFi. Femtocell is a pocket-sized base station that is mainly designed for indoor usage [3]. By installing a femtocell, the base station and its service provider can extend their coverage inside the building. Offloading using WiFi network is also an optimal solution for limited mobile network capacity [3]. Many people are already using WiFi network for better internet access. Due to its popularity, there are many WiFi-friendly applications that support users with only WiFi access to send the message and even support voice and video call with other with the cellular network.

While both mobile offloading using femtocell and

WiFi network are great solutions, they both require internet access. Which means, there is no possible way to reduce the network traffic without having every user connected to the internet. To solve this problem, mobile data offloading with opportunistic communication concept is proposed. By using the nature of delay-tolerant network cellular station would directly transfer the data to the small amount of users. Then, those users who received the data would take a role as a bridge and propagate the data to others. However, there are few problems that need to be concerned: selection and required number of users. Han et al. [4] provide algorithms that select k-users for overloaded mobile data traffic, and Barbera et al. [5] proposed a VIP delegation based on popular attributes in graph theory. All of these works predicts user' future movement based on their analyzed mobility pattern obtained by contact history. However, for this paper, we predict users' future movement by investigating contact location and divide the given environment into several areas.

### 3. Proposed Scheme

In this paper, we proposed a VIP selection scheme based on analyzed user movement pattern for efficient mobile data offloading. Normally, the society is constructed with multiple communities. Thus, for our proposed scheme, instead of finding the most influential node in the network as a whole, we divide the network into several areas and find the most influential node in each area.

In our study, we assume users are able to communicate once they are in-range. When the offloading process begins, we first exploit selected VIPs as initial nodes and start to offload the data to others. Note that we use selected VIPs from past week's dataset during the same time period because people tend to have a weekly-based life pattern [6]. Then, once the users who have not received the data encounter VIPs, their status will change to "received" mode, which means they now received the data and allowed to propagate data.

Table 1 introduces the process of our proposed VIP selection scheme. There are overall three steps in our scheme. Followings are the detailed information of each step in detail.

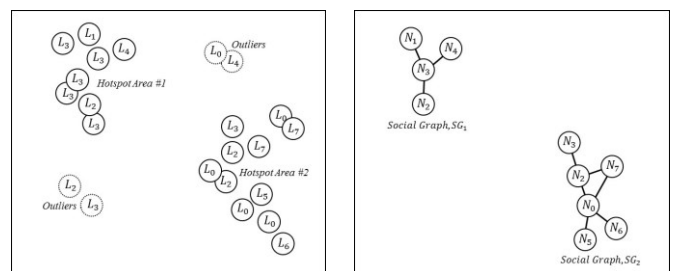
**Hotspot Detection:** First, collect all contact locations during the monitoring period because our focus is to detect frequently encountered areas. Then, cluster collected location points to classify the network into several areas. In this paper, DBSCAN algorithm [1] is used because it is

density-based clustering algorithm that does not require users to insert the number of clusters. Defining the number of clusters is not only a complicated procedure but also a time-consuming task with a large sized dataset. Thus, DBSCAN algorithm is often used to analyze vast amounts of data.

<Table 1> Process of VIP Selection

Step(s)	Description(s)
<b>Step 1 (Hotspot detection)</b>	Use DBSCAN algorithm to classify the contact location dataset.
<b>Step 2 (Node relationship)</b>	Construct undirected social graph for each hotspot area detected.
<b>Step 3 (Scoring)</b>	For each social graph, score each node by its neighbors' degree

**Node Relationship:** As we can see in Fig. 1, collected contact locations are now classified into several hotspot areas. In order to find out the impact that the nodes have in each area, we constructed an undirected social graph for each hotspot detected. Thus, social undirected graph,  $SG_A = (V, E)$ , where  $A$  is a set of hotspot detected,  $V$  is a set of nodes within  $a \in A$ , and  $E$  is a set of edges between nodes based on contact happened in  $a$ , is constructed. Then, multiple contact locations of a single user are united into one single node and form node-relationship that helps to predict VIPs.



(A) Detected Hotspots (B) Social Graphs  
(Fig. 1) Example of visualizing Node Relationship

**Scoring:** In a real-life scenario, when people have a notice that needs to be known by many people within a short duration of time, they normally start with telling a group of friends who are able to contact the people. With this concept, we decide to score each user based on their neighbors' degree. The final step of VIP selection can be divided into four procedures: First, given a

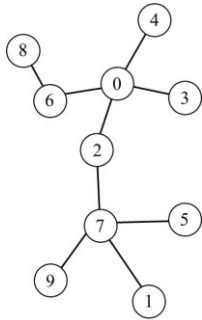
social graph  $SG_a = (V, E)$  from previous step, create  $V \times V$  binary adjacency matrix  $M(a)$ , where  $M_{xy}(a) = 1$  if node  $x$  and  $y$  have a connection, 0 otherwise. The binary matrix is used because we focus on the number of encountered users, not the number of contact with single user. Then, we exploit Equation 1 to calculate each node's degree centrality  $DC_v$ .

$$DC_v = \sum_{j=0} M_{vj} \quad (1)$$

where  $v \in V$  is a number of nodes in a graph. Finally, we define each user's score by solving Equation 2.

$$Score_v = \sum_{j \in F(v)} DC_j \quad (2)$$

where  $F(v)$  represents a list of connected users with user  $v$ . Based on the calculated scores, select the node with the highest score as VIP for each hotspot areas. To visualize these procedures, Fig. 2 illustrates an example of neighbors' degree-based scoring, where node #2 is the most influential user.



(A) Social Graph  $SG_1$

	0	1	2	3	4	5	6	7	8	9	$DC_n$	Score
0	-	0	1	1	1	0	1	0	0	0	4	6
1	0	-	0	0	0	0	0	1	0	0	1	4
2	1	0	-	0	0	0	0	1	0	0	2	8
3	1	0	0	-	0	0	0	0	0	0	1	4
4	1	0	0	0	-	0	0	0	0	0	1	4
5	0	0	0	0	0	-	0	1	0	0	1	4
6	1	0	0	0	0	0	-	0	1	0	2	5
7	0	1	1	0	0	1	0	-	0	1	4	5
8	0	0	0	0	0	0	1	0	-	0	1	2
9	0	0	0	0	0	0	0	1	0	-	1	4

(B) Neighbors' degree-based scoring

(Fig. 2) Example of scoring with given graph  $SG_1$

#### 4. Simulation Environment

For the simulation, we exploit real-world mobility trace, NCCU real-trace dataset [2]. It is one of the recent real-world mobility traces that are available online. The dataset contains 115 students' movement history for two weeks by collecting GPS data from smartphone devices.

For the performance evaluation, our proposed scheme is compared with two other VIP selection schemes – contact frequency and degree centrality. These two schemes select top- $k$  node based on the number of its contact frequency and degree.

Table 2 summarizes the details for simulation environment. The areas that NCCU dataset covers is  $2000 \times 2000$  meters and contains the movement of 115 students. Since our goal is selecting a set of VIPs that is able to effectively offload data in a short duration of time, we set offloading period as 1800 seconds. We also consider monitoring period as 1800 seconds because we are exploiting the VIPs based on past week's dataset during the same time period. For example, if there is a data that needs to be propagated during the period of 8:00 to 8:30 today, the base station will select VIPs based on the dataset from last week's 8:00 to 8:30. We also set  $\epsilon$  and  $minPts$ , two input parameters for DBSCAN algorithm as 30 and 10. Transmission range, the maximum distance of nodes to communicate with others, set to 15 meters.

<Table 2> Simulation setting

Parameters	Values
Number of nodes	115
Area ( $m \times m$ )	$2000 \times 2000$
Movement model	NCCU Trace data
$\epsilon$ for DBSCAN	30
$minPts$ for DBSCAN	10
Transmission range	15
Monitoring period ( $s$ )	1800
Offloading periods ( $s$ )	1800
Number of VIPs ( $k$ )	1~7

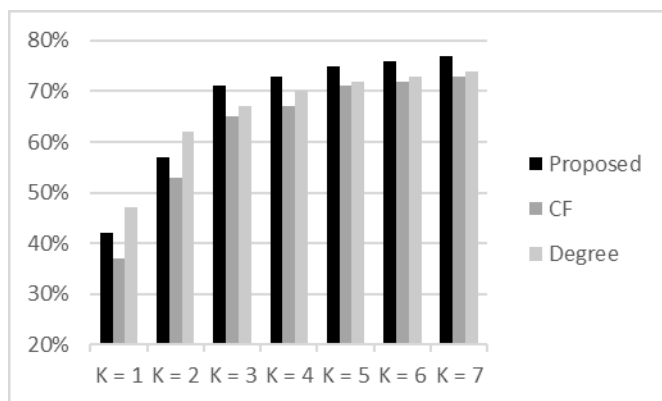
#### 5. Simulation Results

We evaluate the performance of our proposed scheme by comparing the number of received

nodes after the offloading period is finished. Based on the given number of VIPs, we analyze how selected VIPs are performing while assuming the nodes are able to communicate with others within their transmission range. For the comparison, VIP selection based on contact frequency and degree centrality is also implemented and evaluated.

As we can see in Fig. 3, the results show that our proposed scheme performs better than two compared schemes. There are some points where degree-based selection scheme performs better than our proposed scheme. However, as the number of VIP nodes increases, our proposed scheme performs much better than other two compared schemes and offload the data to 70~77% of the network. Note that these results are calculated by averaging the results of 20 evenly divided dataset by time.

The problem of selecting VIPs based on contact frequency and degree centralities occurs when there are multiple small but non-negligible communities with one giant community. In this situation, there are high chances of selected VIPs only being part of one giant community and ignore other small communities, which is the reason why the results of degree-based VIP selection had better performance when  $k$  is small. Therefore, by dividing the network into several areas using a clustering algorithm, we are able to select a set VIPs that can evenly offload the data throughout the entire environment.



(Fig. 3) Percentage of data offloaded

## 6. Conclusion

In this paper, VIP selection using DBSCAN algorithm is proposed for effective mobile data offloading to alleviate overloaded cellular network. We proposed an idea that selects a set of VIPs with size  $k$  by analyzing the past movement patterns. Instead of only considering the contact information, we added the location information

along with contact to detect areas where people frequently encounter, known as hotspots. We also conducted simulations exploiting NCCU real-trace dataset to prove our scheme is effective. Our results show that our scheme performs better than other VIP selection scheme using contact frequency and degree centrality, and offloads about 70~77% of the network with VIP set size of three to seven within 30 minutes. For future work, we would like to extend our simulation with various clustering algorithm such as k-means, spectral and hierarchical clustering as well as the other real-trace dataset for more detailed simulation.

## Reference

- [1] M. Ester, H. Kriegel, J. Sander, X. Xu, "A density-based algorithm for discovering clusters in large spatial databases with noise," *KDD'96 Proceedings, Oregon*, pp. 226-231, Aug. 1996.
- [2] T. Tsai, and H. Chan, "NCCU trace: social-network-aware mobility trace," *IEEE Communications Magazine*, vol. 53, no. 10, pp. 144-149, Oct. 2015.
- [3] A. Aijaz, H. Aghvami, and M. Amani, "A survey on mobile data offloading: technical and business perspectives," *IEEE Wireless Communication*, vol. 20, no. 2, p. 104-112, 2013.
- [4] B. Han, P. Hui, V.S. A. Kumar, M. Marathe, J. Shao, and A. Srinivasan, "Mobile data offloading through opportunistic communications and social participation," *IEEE Transaction on Mobile Computing*, vol. 11 no. 5, p. 821-834, May. 2012.
- [5] M. V. Barbera, A. C. Viana, M. D. D. Amorim, and J. Stefa, "Data offloading in social mobile networks through VIP delegation," *Ad Hoc Network*, vol. 19, pp. 92-110, Aug. 2014.
- [6] J. Su, A. Goel, and E. Lara, "An empirical evaluation of the student-net delay tolerant network," *Proceedings of the 3rd Annual International Conference on Mobile and Ubiquitous Systems*, Jul. 2006.