Performance Evaluation of Referral Selection Algorithms in Human Centric Communication Networks

Yu-Feng Chou, Ray-Guang Cheng, Senior Member, IEEE, and Hsin-Heng Huang
Department of Electronic Engineering, National Taiwan University of Science and Technology, Taipei, Taiwan, R. O. C.
Email: erg@mail.ntust.edu.tw

Abstract— Human centric communication network (HCCN) is a communication network offering human centric mobile Internet services. In such a network, the relationship among users depends on their social behavior and the available network infrastructure. Advertisement and narrowcasting are two main important services in human centric communication networks. In [8], a viral-marketing-based approach was proposed to disseminate advertisement information in HCCN. Heuristic degree-based referral selection algorithms were presented to selected referrals under the given budget and time constraints. In this paper, key factors that may affect the referral selection algorithms are addressed and two referral selection algorithms are presented. Simulation results showed the proposed referral selection algorithms can achieve a better influence rate in different application scenarios.

Keywords - broadcasting; viral-marketing; social networks; referral selection

I. INTRODUCTION

A human centric communication network (HCCN) is a communication network offering human centric (or, social-network-based) mobile Internet services. HCCN offers flexible and intuitive support in accessing situation dependent multi-media data as well as in exchanging multi-media data with others. Human in a HCCN are normally interconnected by their social relationships such as friendship, kinship, financial exchange, or relationships of beliefs, knowledge or prestige. Human in a HCCN are normally equipped with communication devices, which connect them anywhere and anytime through various communication networks.

Advertisement is one of the most important services in HCCN. Many broadcasting techniques have been proposed for traditional communication networks or social networks. These techniques normally assumed that the social relationship between any two users is identical [1,2,3,4,5] and the communication link between any two users is always available [6,7]. In reality, users are not always connected and the relationship between any two users is not identical. Hence, existing techniques cannot be directly applied to HCCN. In [8], a viral-marketing-based broadcasting technique was proposed for HCCN. The advertiser aims to select a number of referrals to disseminate the advertisement information under given budget and time constraints. The referrals are granted for sharing information with their friends. The viral marketing tries to choose the few key “influential” members (i.e., it is also referred as helping-user in [8] or referral in this paper) to trigger a cascade of influence by which friends will recommend the product to other friends [9]. Referral selection is one of the key issues for viral marketing [10]. Normally, referrals are chosen from high-influence users to maximize the infection rate. In the sociology literature, the degree and centrality-based heuristics are commonly used as estimates of a node’s influence [9]. In [8], several degree-based heuristics referral selection algorithms were presented. Different to existing sociology-based approaches, a joint degree of both social and communication layers were considered in [8]. In this paper, key factors that impact the performance of the referral selection algorithms will be discussed. Before going into details, the conceptual model of HCCN proposed in [8] adopted in this paper is briefly introduced.

Figure 1 shows a three-layer conceptual model of HCCN, which comprises of I users, J social groups (SGs), K radio access technologies (RATs) (i.e., GSM/GPRS, WiFi, 3G, Bluetooth or WiMAX) [8]. In this network, each user is equipped with a communication radio communication device such that each user can simultaneously connect to one or more RATs. Each user is associated with the other users by joining a common social group or using the same network service. The HCCN User i may join the jth SG with social strength pij, which depends on user’s social interest and the frequency of interactions. User i may connect to the kth RAT with connection probability qik, which depends on user’s behavior and the availability of the RAT.

The rest of this paper is organized as follows. Section II summarizes the problem and the design considerations. In Sec. III, the proposed referral selection algorithms are presented. Simulations results were shown in Sec. IV. Finally, conclusions were summarized in Sec. V.
II. SYSTEM MODEL

This paper considers a broadcasting problem in HCCN with I users, J SGs, and K RATs. An advertiser selects some users based on the users’ statistics (i.e., \( p_{ij} \) and \( q_{ik} \)) obtained from the user database provider (e.g., Facebook, network operator). Note that the user database provider does not disclose users’ names to ensure confidentiality. The advertiser has to pay \( C_H \) for encouraging each of the selected users to become a referral (e.g., pay the user database provider for acquiring the contact information; give the users free samples of the product, etc.). The advertiser further offers incentive \( \alpha \) for each user attracted by the referral. The incentive may cover the cost of, for example, discount for the new user, reward for the referral, and the processing fee for the user database provider.

The occurrence and the quality of the connections among users in HCCN can be derived from \( p_{ij} \) and \( q_{ik} \) using the affiliation network model. The strength of the friendship the communication connectivity is symmetric because the influence between two users may not be equal. However, the communication connectivity is symmetric because two users have the identical probability to connect to each other.

Let \( n \) and \( N_i(T) \) be the total number of referrals and the estimated number of users attracted by the \( j \)th referral during time \( T \), respectively. Assume that the advertiser must pay \( C_H \) to the company that provides the user database (e.g., Facebook) for each of the referrals. Also, advertiser will offer a reward \( \alpha \) to the referral for each of the attracted user. The referral selection algorithm aims to find

\[
\max_n \sum_{i=1}^{n} N_i(T),
\]

subject to the budget constraint [8]

\[
\sum_{i=1}^{n} (\alpha N_i(T) + C_H) \leq B.
\]

In the implementation, an advertiser only counts the number of users who access the advertisement information through a given referrals. Therefore, some users may be counted multiple times by different referrals for the grant.

The performance of the advertisement is evaluated in terms of an influence rate (IR) and a per-user-reward (PUR). IR is given by

\[
IR(T) = \frac{\sum_{i=1}^{n} N_i(T) - o(T)}{I},
\]

where \( o(T) \) is the number of users who are counted more than once by different referrals during time \( T \). A higher value of IR implies the broadcasting information is received by more users. PUR is the average cost paid by the advertiser to each attracted user, which is obtained from

\[
PUR(T) = \frac{B - B(T)}{\sum_{i=1}^{n} N_i(T) - o(T)}.
\]

where \( B(T) \) is the remaining budget at time \( T \). A lower value of \( PUR \) is also preferred since less cost is required to disseminate information to a user.

III. REFERRAL SELECTION ALGORITHMS

In the sociology literature, referrals are normally chosen from high-influence users [9]. The degree and centrality-based heuristics are commonly used to estimate a node’s influence. In HCCN, a user \( i \)’s influence, \( ID_i \), is proportional to the social relationship, the communication connectivity, and the time constraint. Replace \( N_i(T) \) by \( ID_i \). Eq. (5) can be rewritten as

\[
\sum_{i=1}^{n} (\alpha ID_i + C_H) \leq B.
\]

The referral selection algorithm needs to maximize the IR based on the given budget and time constraints. Normally, the referrals are selected from those with high \( ID_i \). The total number of referrals depends on \( \alpha \), \( C_H \) and \( B \).

Several issues need to be considered in the design of the referral selection algorithm. The first issue is to estimate the influence, \( ID_i \), for each user. From Eq. (8), it can be found that the estimation error affects the number of selected referrals. An overestimation of \( ID_i \) results in a small value of \( n \) and thus, decreases the IR. In contrast, an underestimation of \( ID_i \) may increase the risk of budget deficit for the advertiser. The second issue is to identify the users who may be counted multiple times by different referrals, \( o(T) \), in order to minimize \( PUR \). That is, the referrals should be selected such that \( o(T) \) can be minimized. In the following, impact and solutions for the two issues are addressed.

A. Estimation of Users’ Influence

In [8], a simple estimation of \( ID_i \) is proposed, which is given by

\[
ID_i = \sum_{j=1}^{K} sgn(s_{ij}, T).
\]


In Eq. (9), $ID_i$ is increased by one if none of $s_{j,i}$, $c_{j,i}$, and $T$ is zero. In other words, user $i$ can always influence user $j$ if they join the same social group and access to a common RAT. In real life, however, user $i$ may have less chance to influence user $j$ if $s_{j,i}$, $c_{j,i}$, or $T$ is small. Note that user $i$ can always have the chance to influence user $j$ as long as $T$ goes to infinity. Hence, the $ID_i$ derived in Eq. (8) can be used as an upper bound for estimating user $i$’s influence (i.e., user $i$’s maximum number of friends).

The advertiser selects some referrals and reserves the budget for them based on Eq. (8). However, the estimation error of $ID_i$ in Eq. (9) may become very large if $T$ is small (i.e., for the case of urgent broadcasting). In such situations, most of budget will never be used after the expiry of the time constraint and thus, it may decrease the $IR$. The impact of the estimation error to the performance of the referral selection algorithm will be discussed in Sec. IV.

B. Referrals Selection

In [8], a heuristic maximum in-degree value (MAI) referral selection algorithms in HCCN is proposed. MAI selects the referrals who have the highest $ID_i$ in a descending order until the budget is not enough to support a new referral. However, it is noticed in [8] that different referrals may have some friends in common, which results in a non-zero $\alpha(T)$ and thus, decreases $IR$ and increases $PUR$. Hence, it is preferable to choose uncorrelated users as referrals in order to minimize $\alpha(T)$.

In the sociology literature, two users are totally uncorrelated if their social strength is zero. Based on this simple concept, a maximum in-degree value and less correlation (MAILC) algorithm is further proposed to prevent from selecting highly correlated referrals [8]. In MAILC, a user $i$ will be selected as a referral if the user $i$ is uncorrelated with all of the selected referrals. That is,

$$s_{i,m} \times s_{i,n} = 0, \text{ for } i \in S\text{ and all } k \in S,$$

where $S$ be the set of users and $S_H$ is the set of selected referrals.

In this paper, we propose two referral selection algorithms to identify the uncorrelated referrals. The first enhanced MAI (eMAI-I) algorithm will delete the bi-directional social relationship between a selected referral (user $i$) and each of the remaining users. That is,

$$s_{i,m} = s_{i,n} = 0, \text{ for all } m \in S.$$

Hence, existing referrals’ friends will not be duplicated counted in calculating $ID_i$ of the non-referral users. The $ID_i$ for all of the remaining users is then calculated based on the updated social relationship. The user who has the highest $ID_i$ is selected as the new referral. The procedure of eMAI-I algorithm is executed as below:

**Step 1:** Set $S_s = \emptyset$ and $S = \{1, 2, \ldots, 1\}$. Calculate $ID_i$ for all $i \in S$.

**Step 2:** Find the $i$th user such that $ID_i \geq ID_j$, for all $i, j \in S$.

**Step 3:** Add user $i$ to $S_H$ if the remaining budget is large enough to reward user $i$.

That is

$$(B - \sum_{i \in S} (\alpha ID_i + C_m)) \geq (\alpha ID + C_m).$$

**Step 4:** Remove user $i$ from $S$ and set $s_{i,m} = s_{i,n} = 0$ for all $m \in S$.

**Step 5:** Recalculate $ID_i$ for all $j \in S$.

Repeat Steps 2 to 5 until the remaining budget cannot afford any new referral.

Note that the isolated user (i.e., user $i$ whose $p_{ij} = q_{jk} = 0$ for all $j, k$) will be chosen first in Step 2 if $ID_i$ for all of the remaining users in set $S$ are equal to zero.

The second enhanced MAI (eMAI-II) algorithm only removes the social relationship from a selected referral (user $i$) to each of the remaining users. That is,

$$s_{i,m} = 0, \text{ for all } m \in S.$$

The procedure of eMAI-II is very similar to eMAI-I except that only $s_{i,m}$ is set to zero in Step 4 of eMAI-II.

IV. SIMULATION RESULTS

Simulations were conducted on top of C++ simulation platform to investigate the performance of the proposed referral selection algorithms. In the simulations, performance metric of $IR$ and $PUR$ were considered. In the simulation, it was assumed that $I = 200$ users, $J = 20$ SGs, and $K = 5$ RATs. In the simulation, $B = 100$ and $C_H = 1$ were considered. The direct-sale marketing technique, in which the advertiser assumes the users are uncorrelated and transmits the information directly to users, may be used as a reference benchmark herein. In this case, the direct-sale marketing technique can achieve a constant $IR$ of $(B/C_H)/I = 50\%$, which is independent of the social strength and the connectivity probability. $\alpha = 0.1, 0.5$, and 1 were considered in order to investigate how $C_H\alpha$ affect the referral selection algorithms.

In the simulation, $P_S = 0.1, 0.3, 0.5$ and $P_B = 0.5, 0.7, 0.9$ were investigated. A simple method is used herein as an example to generate the social strength $p_{ij}$ and the connection probability $q_{jk}$ in Fig. 1. In the simulation, each user was assumed to join a SG with a probability $P_S$. The social strength $p_{ij}$ between a user $i$ and the joined SG $j$ was assumed to be uniform distributed within 0 to 1. Similarly, it was also assumed that each user subscribed a communication service from a RAT with a probability $P_B$. The connection probability $q_{jk}$ between a user $i$ and the subscribed RAT $k$ was assumed to be uniform distribution within 0 to 1. Note that $p_{ij} = 0$ ($q_{jk} = 0$) if user $i$ does not join SG $j$ (subscribe a communication service from RAT $k$). In the simulation, a user $j$ has a possibility of $s_{ij}$ to be affected by a referral $i$ if they connect to the same RAT. In such an environment, the probability that a user $i$ affects user $j$ in a unit time interval is in proportion to the social relationship and communication connectivity. Hence, the total number of users affected by user $i$ during $T$, $ID_{i,opt}$, can be estimated by
Note that $ID_{opt}$ is equal to $ID_i$ estimated in Eq. (9) as $T$ goes to infinity.

The accuracy of the $ID_i$ estimated from Eq. (9) is first investigated. It is found that a higher $PS$ results in a higher $ID_i$ because a referral has a higher chance to connect/afflict his friends. For small $T$, $ID_i$ is much higher than the real value (i.e., simulation) in both topologies. As $T$ is increased, the difference between $ID_i$ and the real value is decreased. It is also found that $ID_{opt}$ is a very good estimation of the total number of affected users. In the following, the performance of the referral selection algorithms is investigated based on the incorrect $ID_i$ estimated from Eq. (9). However, it is found that eMAI-I and eMAI-II can still perform well since they can reduce $o(T)$ effectively. It is also found that eMAI-I and eMAI-II can have a better performance if accurate information of $ID_{opt}$ is used.

The $PUR$ of eMAI-I and eMAI-II is higher than that of MAI. It means that eMAI-I and eMAI-II selects more referrals to achieve a higher $IR$ (i.e., the rewards paid for the referral and the user are $C_{it} = 1$ and $\alpha = 0.1$, respectively). The extra referrals also results in a higher $IR$ at $T = 1$. 

Figure 2. The performance of the four referral selection algorithms ($\alpha = 0.1$, $PS = 0.1$, $PR = 0.5$).

The performance of the proposed referral selection algorithms is then examined. Fig. 2 shows the performance of the four referral selection algorithms for $\alpha = 0.1$, $PS = 0.1$, and $PR = 0.5$. It is used to mimic the behavior of loosely coupled users who are normally connected to the networks. It is shown in Fig. 2(a) that eMAI-I and eMAI-II greatly enhance the $IR$ of MAI and MAILC in a very short time. eMAI-II achieves a slightly higher $IR$ than that of eMAI-I. The eMAI-II increases the $IR$ of MAI by 20% when $T = 100$. The $PUR$s of the four referral selection algorithms are shown in Fig. 2(b). It is found that eMAI-I and eMAI-II spend similar $PUR$ to acquire each user.

Figure 3. The performance of the four referral selection algorithms ($\alpha = 0.1$, $PS = 0.5$, $PR = 0.9$).

Figure 4. The $IR$ of eMAI-II
Figure 3 shows the performance of the four referral selection algorithms for $\alpha = 0.1$, $P_{S} = 0.5$, and $P_{R} = 0.9$ (i.e., users are tightly coupled and strongly connected). It is found in Fig. 3(a) that eMAI-I and eMAI-II outperform MAI and MAILC for small $T$ but the performance gain vanishes when $T$ is increased. However, the PUR of eMAI-I and eMAI-II is almost four times as the PUR of MAI and MAILC since they select more referrals. Although MAI and MAILC select fewer referrals, the selected referrals only need time to affect most of users in HCCN. The advertiser may not have to spend too much money on referrals if he reserves a long enough time for the few referrals to disseminate the information. The results for $\alpha = 0.5$, $P_{S} = 0.5$, and $P_{R} = 0.9$ is not shown herein. In this case, all of the four algorithms have almost the same IR and PRU.

Figures 4 and 5 show the IR and PUR of eMAI-II for $\alpha = 0.1$ and 0.5, respectively. It is found in Fig. 4 that IR increases as the increasing of $P_{S}$ and $P_{R}$ because a referral participated in more social groups and subscribed more communication service from a RAT. That is, a referral can affect more users through different social group or RAT. IR also increases as the decreasing of $\alpha$ since the same budget can be shared by more attracted users. It is found in Fig. 5 that the PUR decreases as the increasing of $T$. It is because that the cost paid for the referrals can be shared by their attracted users. The PUR of eMAI-II in $\alpha = 0.1$ does not have big difference as that in $\alpha = 0.5$ as expected. It is mainly because that many referrals are selected by eMAI-II in $\alpha = 0.1$, which significantly increase the average cost spent on each user.

V. CONCLUSION

This paper presents two referral selection algorithms for HCCN. The proposed algorithms work well even if user’s influence, ID, can only be roughly estimated. Simulation results show that the proposed referral selection algorithm can significantly improve the IR than heuristic degree-based algorithms. It is found that eMAI-I and eMAI-II are suitable for short-term advertisement. The advertiser may adopt eMAI-I or eMAI-II to reach high IR in a very short time for loosely coupled users who are lightly/normally connected. In contrast, MAI is a cost efficient solution if users are tightly coupled and strongly connected.

ACKNOWLEDGMENT

This work is supported in part by National Science Council, Taiwan, under contract No. NSC 100-2219-E-011-001.

REFERENCES


