IMaxer: A Unified System for evaluating Influence Maximization Mechanisms in Location-based Social Networks

Muhammad Aamir Saleem
Aalborg University
Universite Libre de Bruxelles

Rohit Kumar
Universite Libre de Bruxelles
Universitat Politécnica de Catalunya

Toon Calders
Universite Libre de Bruxelles
University of Antwerp

Xike Xie
University of Science and Technology of China

Torben Bach Pedersen
Aalborg University

ABSTRACT

Due to the popularity of social networks with geo-tagged activities, so-called location-based social networks (LBSN), a number of methods have been proposed for influence maximization for applications such as word-of-mouth marketing (WOMM), and out-of-home marketing (OOH). It is thus important to analyze and compare these different approaches. In this demonstration, we present a unified system IMaxer that both provides a complete pipeline of state-of-the-art and novel models and algorithms for influence maximization (IM) as well as allows to evaluate and compare IM techniques for a particular scenario. IMaxer allows to select and transform the required data from raw LBSN datasets. It further provides a unified model that utilizes interactions of nodes in an LBSN, i.e., users and locations, for capturing diverse types of information propagations. On the basis of these interactions, influential nodes can be found and their potential influence can be simulated and visualized using Google Maps and graph visualization APIs. Thus, IMaxer allows users to compare and pick the most suitable IM method in terms of effectiveness and cost.

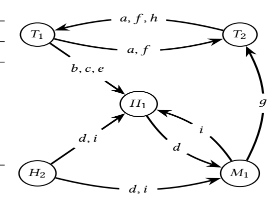
ACM Reference format:
Muhammad Aamir Saleem, Rohit Kumar, Toon Calders, Xike Xie, and Torben Bach Pedersen. 2017. IMaxer: A Unified System for evaluating Influence Maximization Mechanisms in Location-based Social Networks. In *Proceedings of International Conference on Information and Knowledge Management, Singapore, November 6 - 10, 2017 (CIKM'17),* 4 pages.
DOI: 10.1145/nnnnnnn.nnnnnnn

1 INTRODUCTION

In this demonstration, we present *IMaxer*, a unified system that supports a number of different models and algorithms for analyzing and comparing information propagation and influence maximization techniques in LBSNs. In an LBSN, the nodes, i.e., users and locations, form two types of edges: 1) friendships that represent relationships; 2) activities that represent check-ins of users at locations. The propagation of information in LBSNs is studied based on friendships and activities. Let A and B be two nodes representing users who are friends. We say that A *influences* B, or A spreads information to B, if B follows an activity of A. If there are many such following actions by B, A is considered to have a strong influence on B. The process of finding the most influential nodes is called *Influence Maximization*. Applications like viral marketing utilize these influential nodes to maximize the information spread for advertising purposes.

Check-ins

loc	Users t=1	t=2	t=3
T_1	b, c, e, f	a, h	f
T_2	a, h	f, g	a
M_1	g	i	d
H_1	–	b, c, d, e	i
H_2	d, i	–	–



Friendships

User	Friends
a	c, h, i
b	d, f, i
c	a, f
d	b, e
e	d, g
f	c, b
g	e
h	a
i	a, b

Figure 1: Toy example: the tables "Check-ins" and "Friendships" shows the activities of users, and their social friends in an LBSN, respectively. The graph in the center, extracted from the LBSN shows the influence among locations based on visitors that are spread from one location to another [7].

A number of techniques has been proposed in this area. For instance, in [4, 6], authors propose methods for maximizing the number of influenced users. In [7–9] methods for maximizing the number of influenced locations for geographical spread of a message and location promotion are proposed. Moreover, several information propagation models have also been proposed, such as the *Linear Threshold*, the *Information cascade models* [5], the *Time Constraint* [6] and the *Absolute and Relative Influence based information cascade models* [7]. On the other hand, industrial applications [1, 3], focus on management of social media accounts to maximize the spread such as determining the best moment to upload content to maximize viewers. However, to the best of our knowledge, there does not exist any system which provides a complete pipeline of algorithms and models required for all steps of influence maximization and which supports a range of influence maximization techniques for diverse contexts and allows comparison among them. Such a system is useful for selecting an appropriate information propagation mechanism in terms of effectiveness and cost. In this demo, we present a system to handle all such problems in influence maximization.

Example: Consider a marketeer interested in spreading her promotional message to the most regions in New York City (NYC). The information can be propagated using locations or users to

CIKM'17, Singapore

DOI: 10.1145/nnnnnnn.nnnnnnn

achieve this goal as given in the following two use cases. Consider the example LBSN in Figure 1 where for each user, her friends and check-ins in LBSN are given. The graph represents the influence of locations among each other where nodes, labeled with capital letters represent locations and lowercase letters on edges represent visitors as carriers of influence.

Use Case 1: Find locations where she should display the advertisements such that people who see them visit the maximum number of other locations in NYC and spread the message indirectly by talking to their friends and relatives.

Use Case 2: Find users who visit together the most regions of NYC. Give them promotional gifts such as a t-shirt so that they indirectly promote the message to the people of regions they visit.

For aforementioned use cases, a system is required that can be used to select activities in NYC, find influential locations and users for both use cases and simulate their corresponding potential influence. Thus, the system should allow to analyze and compare different methods and find a suitable way to propagate information. The architecture of IMaxer achieves this by providing the following four components.

- *Data selection:* select a desired slice from a dataset in order to target a particular audience.
- *Data preprocessing and transformation:* performs the preprocessing of data to improve the data quality for better accuracy.
- *Influential nodes mining:* capture information propagation mechanisms of diverse contexts for applications such as WOMM or OOH, and find the corresponding influential nodes.
- *Influence spread simulation:* simulate, visualize and analyze the influence of influential nodes. This component can be used to compare information propagation methods and select the most suitable one for a particular use-case.

In short, IMaxer makes following contributions:

- a novel system that provides a complete pipeline of models and algorithms for evaluating and comparing influence maximization mechanisms
- a unified model for capturing diverse information propagation and influence maximization algorithms
- an extensible plugin architecture that allows to integrate new algorithms.

2 IMAXER SYSTEM OVERVIEW

In this section, we map the modules of IMaxer to traditional data mining steps and explain their working:

Data Selection: Usually, marketing approaches target the audience with specific attributes. The data selection module allows users to upload the LBSN dataset and select the desired slice of data based on several given criteria, such as constraints on attributes of users and locations, time, region, combination of friendships and activities. The dataset is filtered on the basis of the given constraints and persisted on disk for subsequent processing.

Example: For both the use cases, in this module, the dataset is filtered to fetch the activities in the region of interest, i.e., NYC.

Data Preprocessing and Transformation: Due to privacy concerns, inaccuracy of GPS devices, and different business models of service providers, there exist anomalies in LBSN datasets, e.g., multiple location IDs are assigned to the same GPS coordinate. Further, in LBSNs different applications may require different location granularities e.g., from a location to a region. Thus, in order to solve data anomaly issues as well as to provide different granularities of locations, we cluster the locations such that each cluster represents a POI. IMaxer provides two algorithms for clustering locations: *grid-based clustering*, and *density-based spatial clustering*.

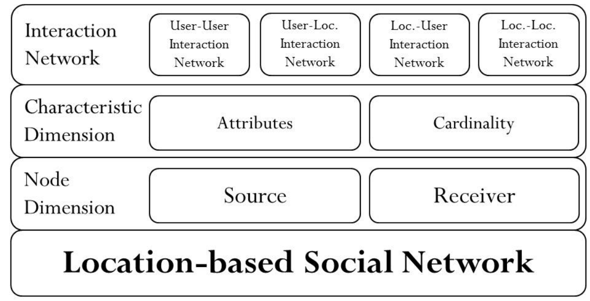


Figure 2: Unified model for influence maximization

Example: In this module, for both use cases, we transform the location IDs in NYC such that each location represents a POI.

Influential Nodes Mining using Unified IM Model: IMaxer provides a unified model composed of three layers as shown in Figure 2 to capture diverse information propagation scenarios. First, the node dimension is responsible for identifying the influential and influenced nodes called source node and receiver node, respectively, which can either be users or locations. Next, the characteristics dimension filters the source and receiver nodes on the required attributes such as age, and gender of users. Moreover, it defines the cardinality for source and receiver nodes, i.e., the number of nodes that should be selected as source to spread the information and as receiver that should be influenced, respectively. Finally, the interaction network layer on the basis of possible interactions among nodes of LBSNs provides four types of interaction networks, i.e., user-user, user-location, location-user, location-location interaction networks. Based on the selection of types of source and receiver nodes the corresponding type of influential activity is fetched as shown in Table 1 and the corresponding interactions network is constructed. Influence scores of source nodes using IM algorithms are computed on the basis of these interactions and finally, the nodes capable of influencing maximum nodes are selected as top-k influential nodes.

Example: In use case 1, we are interested in maximizing geographical spread using locations, the influential activity is spreading of visitors from a location to another location as shown in Figure 1. Thus location is considered as influential as well as influenced node, and interaction network is location-location interaction network. On the basis of it top-2 influential locations in the figure are T_1 and H_2 as they spread their visitors to the maximum number of other locations. For the use case 2, maximizing geographical spread using users is intended. Thus, visit of a location by a user is considered an influential activity. In this case, user and location are considered influential and influenced nodes, respectively, and the interaction network is user-location interaction network. Here, the top-2 influential users are d and a as they together visit most of the locations.

Influence Spread Simulation: In this module, IMaxer takes the influential nodes as input and outputs the potentially influenced nodes. The influence of top-k nodes are simulated and propagated through the interaction network and all the nodes that are influenced are fetched as potentially influenced nodes. In order to do that IMaxer provides a number of spread simulation algorithms (discussed in Section 3). The potentially influenced nodes are further displayed using Google Maps and Gephi to visualize the spread of the influential nodes, for location and user nodes respectively. Since this module allows users to provide influential seeds directly as an input to find their influenced nodes, it can be used to compare and analyze the results of multiple influence maximization methods and choose the best one for a particular scenario.

Example: In use case 1, in the Figure 1, T_1 and H_1 are influencing all other locations. Because, users a, f visits T_2 and b, c, e visits H_1 after visiting T_1, and users d, i visits H_2 after visiting H_1 and M_1. Similarly, for use case 2, the potentially influenced locations of the top-2 influential users d and a are all the locations. Because, d visits M_1, H_1, and H_2, and a visits T_1 and T_2. In this case, as influence of both approaches is same, decision can be made based on cost of considering a user vs location for spreading message. On the other hand, if we consider $k = 1$, then the the top influential user d propagate message to one more location that the top influential location T_1, thus, in this case it can be choosen for more influence.

3 IMAXER'S TECHNICAL OVERVIEW

In this section, we first model four different types of information propagation and their corresponding IM mechanisms using the unified IM Model. Then based on this model, we provide abstractions of algorithms for IM and spread simulation. Finally, we present the implementation details of the system and methods to incorporate other existing as well as new algorithms.

3.1 Modeling IM using *Unfied IM Model*

Given a set of users U and a set of locations L, an *activity* is a visit/check-in of a user at a location. It is a triplet (u, l, t), where $u \in U$ is a user, $l \in L$ a location and t is time of the visit of u at l. The set of all activities over U and L is denoted by $\mathcal{A}(U, L)$. The *LBSN* over U and L consists of a graph $G_S(U, F)$, called *social graph*, where $F \subseteq \{\{u, v\} | u, v \in U\}$ represents friendships between users, and a set of activities $A \subseteq \mathcal{A}(U, L)$.

Given $u, u' \in U \cup L$, we measure $u's$ influence on u' by two factors. First, the number of influential activities $\mathcal{I}(u \rightarrow u')$ that implies an impact/influence of u on u'. Second, a time window during which the activities are performed called influence window ω. The influence among two vertices is 1 if the number of influential activities is greater than a threshold τ, otherwise 0. The values of τ and ω may vary on the basis of types of influential activities and users' requirements. However, for simplicity and to capture all activities we take $\tau = 0$ and $\omega = 3$ for all the examples given below. The unified IM Model utilize the interactions of nodes in LBSNs to model four different types of information propagation mechanisms which are further utilized to formulate influential activities. These information propagation mechanisms with corresponding influential activities and potential applications are shown in Table 1. Next, we define influences on the basis of these interactions and influential activities.

Source	Dest.	Influential Activity ($\mathcal{I}$)	Application (Maximizes)
User	User	u follows u'	Followers
Location	Location	l spreads visitors to l'	Geographical spread
User	Location	u visits l'	Unique visited locations
Location	User	l visited by u'	Unique visitors

Table 1: Information propagation mechanisms w.r.t. types of the influential and influenced nodes in LBSNs. Here, u and l are the influential user and location, respectively and u' and l' are the influenced user and location, respectively.

Influence among users: The influence of a user $u \in U$ on a user $u' \in U$ is determined by the number of activities performed by u that are followed by u' within the time window ω, such that u and u' are friends. The value of the influence of u is considered 1 if the number of such followed activities by u' is greater than the threshold τ.

Example: Most influential users in this context can be utilized to maximize the number of followers. Such users are exploited in applications like viral marketing. In Figure 1, top-2 such influential users are b and c which influence h and i, and a and f, respectively.

Influence among locations: The influence of a location $l \in L$ on a location $l' \in L$ is determined by the number of users that after visiting l visits l', within time window ω. The value of influence of l on l' is 1, if such visitors are greater than τ.

Example: Use case 1 is the example of this type of influence. In Figure 1, the top-2 influential locations are T_1 and H_2 and their corresponding influenced locations are T_1, T_2, M_1, H_1 and H_2.

Influence of user on location: The influence of a user u on a location l' is determined by the number of visits of u at l' within ω. If such visits are greater than τ then the influence of u on l' is 1.

Example: Use case 2 lies in this category of the influence. In Figure 1, the most influential users for this type of influence are d and a and their influenced locations are T_1, T_2, M_1, H_1, and H_2.

Influence of location on user: The influence of a location l on a user u' is determined by the number of times l is visited by u' within ω. The influence of l is considered 1 if the number of such visits of u' at l is greater than τ.

Example: Such type of influence can be used to find the locations on which advertisements can be displayed such that the message is viewed by the maximum number of users. In Figure 1, the top-2 most influential locations for this type of influence are H_1 and T_2, which collectively attracts all the users.

Influence Maximization: Next, we utilize these interactions and influential activities to create interaction networks. An interaction networks is a directed graph $\mathcal{G}(\mathcal{V}, \mathcal{E})$, where $\mathcal{V} \subseteq U \cup L$ is the set of vertices and $\mathcal{E}$ is the set of edges: $\{(u, u') \mid \mathcal{I}(u \rightarrow u') > \tau\}$. The set of vertices that are influenced by a vertex u is called influence set $\sigma(u)$. Once the interaction network $\mathcal{G}(\mathcal{V}, \mathcal{E})$ is obtained for a use case we can use the corresponding IM algorithms to determine the influence set of individual vertices. IMaxer provides 2 IM algorithms taken from [4, 6] for capturing user-user influence, from [7] for location-location and two novel IM algorithms based on time constraint window and edge degree of interaction network

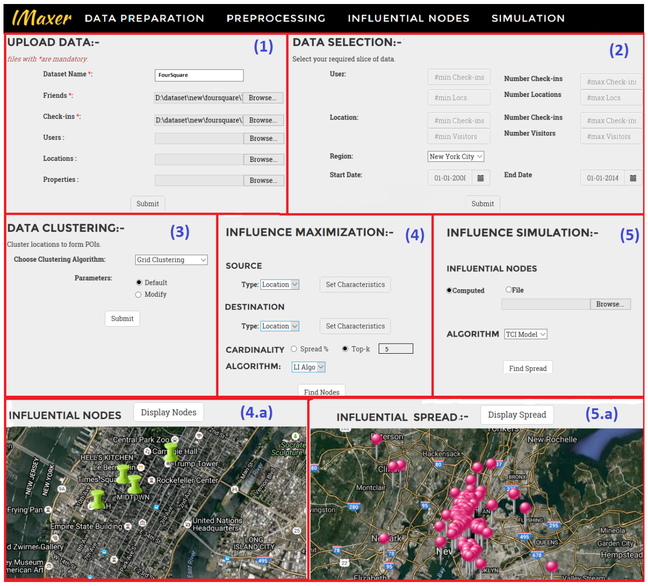


Figure 3: Compact snapshots of usage of IMaxer for maximizing geographical spread with locations in NYC

for user-location and location-user influence. Once the influence sets of individual vertices are known we apply a greedy method based on influence scores of vertices to find top-*k* vertices.

Spread Simulation: The next problem IMaxer addresses is simulation of the influence spread of the selected top-k vertices after influence maximization. In order to do that, IMaxer currently, provides the implementation of Linear Threshold model, Independent Cascade model [5], Time Constrained Information Cascade Model [6], and, Absolute and Relative Influence based information cascade models [7].

3.2 System Implementation

The IMaxer's system implementation is modular as it is based on the factory design pattern. Currently, IMaxer provides three datasets taken from real world LBSNs: Foursquare, BrightKite, and Gowalla. However, it allows to add new data sources. Moreover, IMaxer allows its users to add other influence maximization algorithms by implementing an interface: *BaseAlgo.java*. Similarly, new algorithms for spread simulation can be added by implementing the interface *BaseSimulation.java*. The resulting visualization is not configurable and is fixed based on the receiver node dimension. If the receiver is location, Google Maps are used to visualize the spread and for the cases when the receiver is a user, graph visualization using Gephi is provided.

IMaxer is built using Java for business logic with HTML and Javascript for the front end. We use the Google Maps API to present the location spread and Gephi to present user graph spread. The configuration details are stored as XML files making it easier to configure new data sources and attributes.

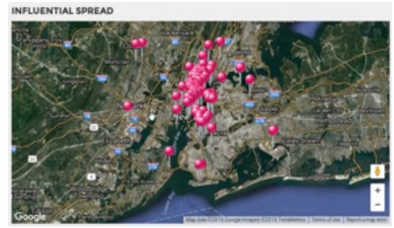


Figure 4: Potential geographical spread with users in NYC

4 DEMONSTRATION

We demonstrate the two use cases given in Section 1 to show and compare the propagation of information using locations and users, respectively, with a real dataset from Foursquare LBSN. The information is spread to 236 locations in use case 1 as compared to 192 locations in use case 2. Due to space limit, we combined the snapshots of usage of different modules of IMaxer for use case 1, as shown in Figure 3. Further, potential influence of users in case 2 is shown in Figure 4. Moreover, we explain the working of IMaxer for both use cases in the video [2] [1].

Demonstration: First, an IMaxer's user uploads the mandatory files containing friendship and check-in information of the data set (Foursquare) as shown in step 1 in Figure 3. Then she filters the data on the basis of user requirements as shown in step 2. By default, she considers all users and locations. However, only activities within NYC and given dates are fetched. In step 3, she transforms the location using grid clustering and maps the activities correspondingly. In step 4, first types of source and destination nodes, and their cardinalities are asked by the user. Then, the influence maximization algorithm is selected, i.e., LI. The nodes marked in green in step 4.*a* were obtained by running the LI algorithm for maximizing geographical spread [7] to find the top-5 most influential locations. The usage of spread simulation is shown in step 5. The locations marked in red in step 5.*a* indicate the locations in NYC that will potentially get influenced according to the selected spread simulation algorithm TCIC. Moreover, IMaxer also provides details of the influenced locations with their corresponding influential locations in a summary table.

REFERENCES

[1] Hootsuite. https://hootsuite.com.
[2] Imaxer's demonstration. https://youtu.be/hkj6aAfhYkk.
[3] sproutsocial. http://sproutsocial.com/.
[4] A. Goyal, F. Bonchi, and L. V. Lakshmanan. Learning influence probabilities in social networks. In *WSDM*, 2010.
[5] D. Kempe, J. Kleinberg, and E. Tardos. Maximizing the spread of influence through a social network. In *KDD*, 2003.
[6] R. Kumar and T. Calders. Information propagation in interaction networks. In *EDBT*, 2017.
[7] M. A. Saleem, R. Kumar, T. Calders, T. B. Pedersen, and X. Xie. Location influence in location-based social networks. In *WSDM*, 2017.
[8] T. Zhou, J. Cao, B. Liu, S. Xu, Z. Zhu, and J. Luo. Location-based influence maximization in social networks. In *CIKM*, 2015.
[9] W.-Y. Zhu, W.-C. Peng, L.-J. Chen, K. Zheng, and X. Zhou. Modeling user mobility for location promotion in location-based social networks. In *KDD*, 2015.

[1] IMaxer will be demonstrated in a similar fashion for other use cases as well, i.e., user-user, using three real-world data sets. The users will be able to interact and find the influential nodes and their potential spread on the basis of their given inputs such as a region of interest, characteristics of nodes, and types of source and destination nodes (influential activities).