

# Information Spreading in Social Networks

Student: AHMAD ZAREIE

Supervisor: RIZOS SAKELLARIOU

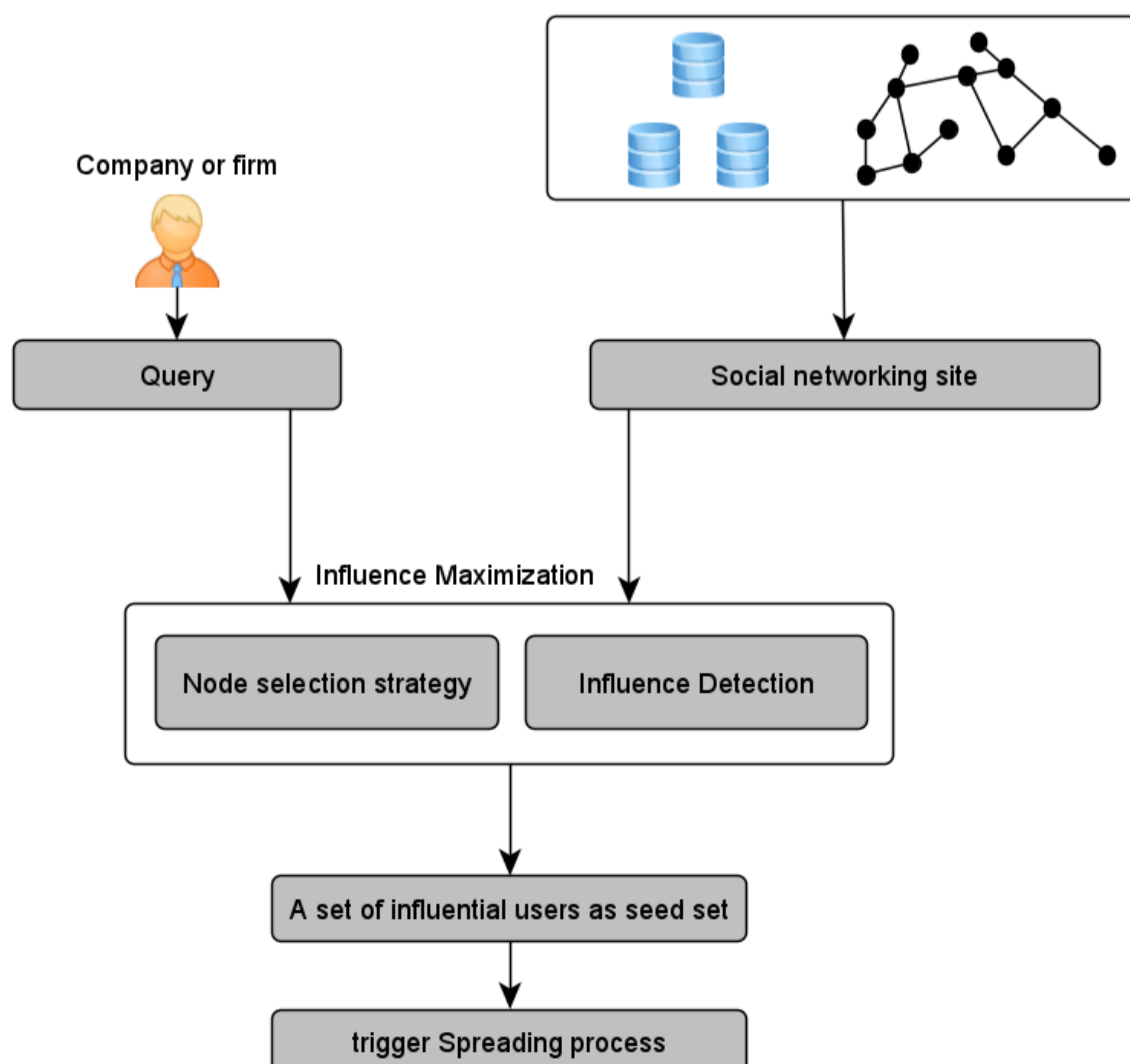
Social media are getting more and more popular, and the number of its users and the amount of exchanged data between them are on increase every day. It has provided an appropriate opportunity for daily activities. Therefore, the analysis of such networks has attracted a lot of attention among which identification of influential users for different applications such as spreading messages, Viral marketing, and preventing rumor propagation is one of the hottest topics. A multitude approaches has been proposed for determining users' influence and identifying a set of influential users in social networks. The links between users and structural information of network are regarded in some research, while users' preferences and behaviours have been taken into account in the other research. In this study, these methods are explained and described.

## Introduction

- ▶ Due the extension of internet technology, social networks are getting more important role in people's life.
- ▶ Quick and easy propagation of messages and news between these users is a considerable potential, and can impact on the users' opinions.
- ▶ Users do not have a same role in such networks and some of them are more influential due to their behaviours and friends in the network.
- ▶ Influential users can be employed for different goals such as advertising a product, blocking a rumor and so on.
- ▶ Measuring the influence of users and identifying a set of influential users, which is generally known as influence maximization (IM) problem, is a key problem in this way.
- ▶ IM problem was firstly studied as an algorithmic problem by Domingos et. al, and Kempe et.al are the first to formulate the problem as a discrete optimization problem.
- ▶ A social network is defined as a graph in the literature, which users and the relation between them are modeled by nodes and edges, respectively.
- ▶ Network structure and users' behavioural information are applied to determine users' influence.

## General model

General process of IM problem is presented in figure. Social network sites provide some data including users' information and relationship. A company, firm or campaign which intend to spread a message makes a query. Query may include a message, some features related to the message, some preferences, and constraints. According to the information provided by the social network site and query, influence maximization aims to identify a set of influential users as initial seed set to spread the message and maximize the number of users who are influenced. A set of seeds is identified and propagation is triggered by seeds.



## Problem Definition

IM aims to spread a message as widely as possible, but due to a huge number of users, it is highly time consuming and even impossible to send the message to all users. In addition, the friendship and trust between the users lead to a more effective acceptance of the message, so IM aims to find a set of users that can spread the message more widely and have more influence on the other users. IM problem is generally defined as follow, where function  $\varphi(S)$  is the influence of a seed set  $S$ . Constraints can be the number of seeds or the amount of money paid to seeds.

$$S^* = \arg \max_{S \subseteq V} \varphi(S)$$

s. t. some constraints

## Literature Review

Proposed methods for IM can be classified into three categories:

- ▶ Simulation-based methods: Simulation-based methods simulate spreading in real world to find a set of the most influential users. In the naive greedy method, the seed set  $S$  is initially considered empty. The algorithm is repeated in iterations and in each iteration a node that its union with  $S$  maximizes the spreading ability is added to  $S$  as a new member.
- ▶ centrality-based methods: In these methods vitality of each node is first calculated using a centrality measure such as degree, then top central nodes are selected as seeds.
- ▶ Heuristic methods: These methods are based on the idea that simulation-based methods suffer from high overload due to implementation of diffusion model on one hand, and on the other hand central nodes in the graph may located close to each other and have high overlap. Thus, nodes centrality and distance are both taken into account in these methods to propose an efficient algorithm for IM problem.

## Research Gap

Current methods mainly focus on topological information of users, while beyond topological information can help to improve the identification of influential users, so proposing methods for IM problem based on beyond topological information can be considered in further research.

## References

- ▶ Domingos, Pedro and Richardson, Matt  
*Mining the network value of customers*  
Proceedings of the seventh ACM SIGKDD international conference on Knowledge discovery and data mining
- ▶ Kempe, David and Kleinberg, Jon and Tardos, Éva  
*Maximizing the spread of influence through a social network*  
Proceedings of the ninth ACM SIGKDD international conference on Knowledge discovery and data mining
- ▶ Granovetter, Mark  
*models of collective behavior*  
American journal of sociology
- ▶ Lü, Linyuan and Chen, Duanbing and Ren, Xiao-Long and Zhang, Qian-Ming and Zhang, Yi-Cheng and Zhou, Tao  
*Vital nodes identification in complex networks*  
Physics Reports