Chapter 7

Concluding Remarks and Future Directions

In this final chapter major contributions of the thesis are highlighted along with the discussion of future directions.

7.1 Summary of Contributions

Detection of influential spreaders in the networks has remained utmost important in social network analysis. In this thesis, the role of individual nodes influence to their neighbors along with topical interest has been studied extensively to improve the effectiveness of seed nodes. Multiple featured networks with multiple product adoption are also explored to identify influential users. The information diffusion has been used to develop an application. Main contributions of the thesis are listed below.

- **LAPSO-IM:** A local influence evaluation function is studied and extended for optimizing IM problem. The local influence evaluation function provides a reliable expected diffusion value of influence spread under traditional diffusion models. To optimize local influence evaluation function, a learning automata based discrete particle swarm optimization algorithm is studied and utilized. LAPSO-IM redefines the update rule of particle's velocity based on learning automata action to overcome the weakness of premature convergence. The experimental results on six real-world social networks show that the proposed algorithm is more effective than base algorithm DPSO with same the level of efficiency and more time-efficient than IMLA with approximate influence spread.

- **C2IM:** Developed a community-based context-aware influence maximization algorithm to identify influential users. C2IM uses a community-based framework to improve the time-efficiency that reduces the search space significantly and considers user's interests to address the effectiveness of seed. It extended the traditional information diffusion models to context-aware linear threshold model and context-aware independent cascade model for influence spreading. To identify k most influential users, C2IM first proposed a community detection algorithm to partitions the network into subnetworks. Then it devise a non-desirable nodes finder technique to identify non-desirable nodes. Finally, seed selection algorithm is proposed to compute most influential seed nodes based on diffusion

degree of nodes. Experimental results show that the proposed algorithm performs better than CIM on influence spread and faster than TIM. Therefore, C2IM algorithm is a trade-off between quality and efficiency.

- **MIM2:** Introduced multiple influence maximization across multiple social networks problem. This problem considers the assumption that an influential user can accept multiple products for free and non-influential users have enough purchasing power to adopt multiple promotions from their social interactions. It is also important to consider the role of overlapping users to spread the influence across networks. To address these issues, a unified framework is proposed to analyze and represent the MIM2 problem. More specifically, first, algorithm performed a mapping to couple a set of networks into a multiplex network via direct linkage strategy. Second, it proposed a heuristic method to find the most influential user over multiple product diffusion multiplex networks. Third, proved that MIM2 problem is NP-hard and expected influence spread function is submodular under traditional diffusion models. Finally, the experimental results show that the advantage of proposed IM problem over existing IM problems.

- **CLP-ID:** Developed a community-based link prediction algorithm using information diffusion perspective as an application of link prediction. A community detection algorithm based on label

propagation is proposed that divides the network into clusters. Then a novel algorithm which is based on information diffusion and community structure is proposed to predict target links. Finally, experiments on real-world networks are performed to validate the performance of proposed algorithm and also compare it with the state-of-the-art algorithms. The statistical test justifies that the proposed method is significantly different from the state-of-the-art algorithms.

7.2 Scope for Future Work

In this section, several new avenues of research that have been opened up by this thesis are mentioned below.

1 **Incorporating stability.** The authors of [190] show and diagnose the instability of IM algorithms when influence probabilities $p_{x,y}$ are noisy. Therefore, a slight change in the propagation model may lead to drastic change in optimal seed users. Although there are some works are presented to robust IM problem such as IRIE [80], EASYIM [43], Saturate Greedy [136], and MAX-COVER [137] based on assumption that the structure of network is fixed. However, the structure of the real-world networks are change continuously. Therefore, to identify robust seed set from given limited graph changes is a challenging task.

2 **Switching from strict submodularity to moderate (comparatively weaker) submodularity.** In order to develop theoretical bound IM solutions, submodularity of objective function is important. The submodularity property of objective function is too strict in some scenario such as the opinion-aware IM approaches [43, 191, 192] adopts non-submodularity. This is because of nodes can switch their states between negative and positive opinion. Therefore, the greedy algorithm is not applicable in such circumstances. To tackle such circumstances heuristic solutions can be proposed. In order to obtain a better solution than heuristic solutions, a possible direction is to develop objective function $\sigma(S)$ with more general submodular function. The authors of [193] introduced a more general function, named as weakly submodular function. The weakly submodular function guarantees the theoretical approximation of the problem.

3 **Incorporating conformity.** Most of the existing approaches only focus on social influence between a pair of individuals x and y with an edge (x,y) connecting them. However, in real world, users are influence by their friends or acquaintances as well as group norms such as conformity. Users which have similar background, education, age etc., in the group conform each others using conformity behavior. The authors of [194–197] extract the conformity characteristic from the network and ignores users profiles regarding social group. In general, there are two possible future directions: incorporating users profiles with conformity

characteristic, and consideration of different types of conformity such as identification, compliance, and internalization, etc., in IM problem.

4 **Identification of overlapping users.** Most of users in social networking sites like Facebook, Twitter, etc., have multiple accounts and spread influence to their friends in each network simultaneously. Therefore, it brings opportunity to account the users influence across multiple networks. This leads to generation of more effective seed nodes. To perform information diffusion across multiple networks simultaneously, we need to find users who actively involve in multiple networks. These users are useful in graph coupling and known as overlapping or identical users. There are some efforts are done in this direction [163–165]. The identification of overlapping users is a challenging task because of diversification of users information across networks and also network data is noisy, big, unstructured and incomplete.

5 **Network coupling and incorporating heterogeneous diffusion models.** In order to estimate aggregate influence spread of a user under MIM2 framework, there is a need of a coupling scheme to couple multiple networks into a single multiplex. There are already some work [38–40] propose in this direction. All these works consider an assumption that each network have same diffusion model. However, in real-world, these networks may have its own

diffusion model. Therefore, one possible future direction is to incorporate heterogeneous diffusion models in network coupling to accurate estimation of influence spread.

6 **Budget fixing: number of items per product.** Let consider a scenario that an advertising company wants to promote multiple products simultaneously in the social network. In real-world, distinct users have different influence or interest for distinct product in social networks. Therefore, to maximize the product adoption or profit, the marketing company needs to decide the number of items per product such that $\sum_{i=1}^{i=m} |P_i| = k$, where m and P_i denotes the number of products and product-type. The authors of [37] provides an strategy to fix the budget and number of items per product. However, they ignores the profit of individual product and consider that each product generate same profit. Also, there is a possibility of left-out some products in advertising. One possible future direction is to incorporate the product characteristics in fixing the budget of each product.

7 **Considering contextual features.** There is a need to incorporate some contextual features such as location and topic with classical IM problem for novel applications. Although some works have been done to incorporate contextual features [104, 131]. However, context-aware IM is still a less researched area and many topics need to be explore. For example, time-aware information diffusion models

CTM and TRM are still largely unexplored, absence of users utilities varying with time periods, etc.

8 **Incorporating uncertainty and incompleteness of the diffusion process.** In this framework, there is concern of uncertainty and incompleteness of the diffusion process. There are some effort are already done in this direction [124, 132]. They ignore user's history and involvement to learn influence probability of users and use random variable conforming to certain distribution. One possible future direction is to consider users history to update influence probabilities iteratively.

9 **Ethical use of social media.** Nowadays, social media is increasingly becoming an integral part of modern life. An application of influence maximization problem also need to address the ethical use of social media and its data from both users and service providers perspective. Therefore, the ethical issues regarding the use of social media can also be considered.

10 **Others.** There are some other possible directions like privacy preserving AI, Bayesian and deep learning framework, linear programming, etc., can be considered to find elegant and effective solution to influence maximization problem. For example, privacy leakage is not considered in existing IM algorithms. These algorithms may results in leakage of private and sensitive

information regarding users. This becomes more important in view of global legislation regarding private data protection.