Total Influence and Hybrid Simulation of Independent Cascade Model using Rough Knowledge Granules

Suman Kundu
Department of Computer Science and Engineering, Indian Institute of Technology Jodhpur, India
suman@iitj.ac.in

***Abstract*—The paper defines a new theoretical measure *Total Influence*, of a node as well as a set of nodes in the social network. Total influence uses probabilistic theory to obtain the expected size of the information spreading in the social network under the independent cascade model of diffusion. In order to quantify the size of the spreading practically, the paper proposes a new *hybrid simulation* methodology for the independent cascade model. The hybrid method uses rough set theory and defines rough knowledge agents around all the seed nodes from which the information is propagating. The lower approximation is calculated using the probabilistic approach, while the size of influence in the boundary region is quantified by Monte-Carlo simulation on a reduced network. The reduce network is formed by compacting all the nodes in the lower approximate region as a super-node. Experimental results on two synthetically generated directed network show that the hybrid method runs magnitude faster than its counterpart with a similar accuracy of the spreading size.**

***Index Terms*—Influence Maximization, Rough Granule, Rough Set, Social Network, Spreding Size**

I. Introduction

Independent Cascade (IC) model is one of the two (other being Linear Threshold model) widely used information diffusion models available to computer scientists. IC model is a stochastic method in which information flows from a node to another neighboring node based on probabilistic rules [1]. Given a set of starting nodes (referred as seed nodes) one needs to run simulations for a sufficiently large number of times to estimate the overall size of an information spread. Usually, an average of 10000 or 20000 independent Monte-Carlo simulations is used for a reasonable estimation of the cascade size. Ten thousand simulations, even with the help of modern technologies, is time-consuming, especially when the social network is Big-data in size. Unfortunately, no deterministic approach is available to date to measure the spreading size.

The goal of the current research work is to provide an alternative to these tedious simulations of the IC model. This new alternative, in turn, helps in the development of algorithms on information diffusion in Big-data social networks. In this study, a network is modeled with a collection of rough granules, called knowledge agent, around each seed node. A knowledge agent spreads information in the network, and the granule defines the region of the knowledge spread in the network. A measure, namely total influence, is proposed to mathematically determine the region of information spread by a single knowledge agent. Besides, a measure to mathematically determine the total information spread by a set of knowledge agents is also proposed. In order to find out the total information spread practically, a hybrid simulation mechanism which uses the rough set theory has been developed. The algorithm runs in two steps; first, it approximates the lower bounds of the rough knowledge agents; and in the next step, a Monte-Carlo simulation is run on a reduced graph to approximate the "upper − lower" region of the agents. Experimental results with two synthetically generated directed network show that the proposed hybrid simulation is faster than the conventional approach of Monte-Carlo simulation.

A. Related Work

Diffusion of innovation first studied by the sociologist [2], [3] more than half a century back. With the digital presence of the human being, the very topic is being explored by different field of studies, including computer science for the last few decades. One of the pioneering work [4], [5] on viral marketing triggers a series of research studies on diffusion models [1], [2], [6], [7] and several application of information diffusion, especially in the online social networks [1], [5], [8]–[13]. The main objective of these applications was to use the information diffusion for viral marketing, diagnosis, and controlling the epidemic spread and identifying threats, among others. However, these algorithms are highly dependent on the underlying information spreading model. Independent Cascade (IC) model is one of the most popular information diffusion model first used by [1] in order to find a solution for influence maximization.

B. Independent Cascade Model

The diffusion of information in the IC model works as follows. The nodes in the network can be in one of the two states, namely, active and inactive. A node is active when the node accepted the information being circulated in the network, and a node is considered inactive when it does not have the information. In each time step, one of the active nodes tries to activate one of its inactive neighbors whom he never

978-1-7281-5688-0/20/$31.00

tried before. That is a node never tries more than once to activate the same neighbor. The inactive node gets activated with a probability, namely, propagation/diffusion probability. In a real scenario, propagation probability is the property of individual relation, i.e., each relationship will have a different probability of getting activated. However, most of the work considers static propagation probabilities, that is the same diffusion probability for all the edges.

C. Influence Maximization and Greedy Solution

In information diffusion literature, the problem of influence maximization is to find a set of influential nodes in terms of spreading size in the network. The natural solution to the problem will be to select those persons having higher numbers of neighbors. Domingos and Richardson were the first to study the problem [4], [5] in the algorithmic aspect and proposed probabilistic methods to solve it. Later, Kempe et al. formulated it as a discrete optimization problem [1] and showed that the problem is NP hard. They proposed a greedy hill climbing algorithm shown in Algorithm 1. In each iteration of the algorithm, the marginal contribution of every non-seed node to the information diffusion is separately estimated, and the highest contributor is selected as the next seed. Thus, the algorithm maximizes the influence contribution during the selection of seeds. It is the best-known solution of the influence maximization problem to this day. The main drawback of the algorithm comes from the marginal contribution estimation.

Algorithm 1 Greedy Hill Climbing Algorithm

input : A Social Network $G(V, E)$ and k
output : Set $S \in 2^V$ having cardinality k

initialization: $S := \emptyset$
while $|S| \neq k$ **do**
  $v^* \leftarrow \arg\max_{v \in V \setminus S} \hat{\sigma}(S \cup \{v\})$
  ; /* $\hat{\sigma}$(.) returns the estimated influence */
  $S \leftarrow S \cup \{v^*\}$
end

There is no deterministic method available to get the marginal contribution of a node. In their paper, Kempe et al. [1] used Monte-Carlo simulation for the estimation of such contribution. As the process of information diffusion is highly stochastic, the simulation needs to be performed for a sufficiently large number of times to obtain more accurate results. It may take days to identify the top 50 seeds even on a graph of the moderate size of 30K nodes [14]. The current study tries to remove this bottleneck using a *hybrid simulation* method described in the following sections. Very recently, Zhou et al. [15] proposed a method to find the upper bound of the influence and use it to determine the influential nodes in a social network. This algorithm is further improved in [16]. Both of these algorithms tried to reduce the number of runs of Monte-Carlo simulations based on the upper bound and the idea of the lazy forward method. On the contrary to their work proposed work in this paper tries to reduce the size of the network with the help of Rough Set theory in order to gain execution time benefit in Monte-Carlo simulation. Rough set had been successfully used to model the overlapping neighborhoods [17], [18] in social network analysis. The current paper uses rough set to model the region of information spread in the network.

The paper is organized as follows. Section II defines the total influence measure for both single and multiple nodes. The methodology of hybrid simulation is presented in Section III. Experimental results are reported in Section IV and finally, Section V concludes the contribution and discuss the future works.

II. Total Influence

In the study of influence maximization in social network analysis, a person is considered influential if a piece information given to him/her reaches to the others in the network over a certain period of time. These influencing person is called *information agent* or *knowledge agent*. Size of influence is measured by the number of people (or nodes in the graph representation) having the same information that of the knowledge agent after the diffusion. Propagation of information follows the domino effect. That is neighbors of the knowledge agent (say level 1) get the message first. Then if these nodes get influenced, their neighbors (level 2) are informed and so on. This cascade flows in a stochastic process. Nodes get influenced with a probability referred to as propagation probability or diffusion probability. This section formulates a measure to find the expected size of the influence of knowledge agents. The size we referred to as *total influence* as it comprises of the influences for all the level from the knowledge agent.

When the same information is propagated from multiple knowledge agents, the domain of ones' total influence is affected by the others. Further, the section proposes mathematical formulas to calculate the total influence of a set of information agents.

A. The Model

Let the social network is represented with a graph $G(V, E)$ where V is the set of nodes (denoting a person in a social network) and $E \in V \times V$ is the set of edges (denoting the relationships in the people). Each edge $e \in E$ is associated with a probability of activation/propagation λ_e. A node u is influenced by node v via the edge $(v, u) \in E$ with a probability $\lambda_{(v,u)}$. The function $\Gamma(v)$ returns the set of neighbor of node v.

B. Total Influence of a Node

Definition 1 (Expected Influence at Level 1): For a given information agent v the expected number of nodes influenced at level 1 is defined by the expectation $\mathbb{E}^{(1)}$ as:

$$\mathbb{E}^{(1)}(v) = \sum_{u \in \Gamma(v)} \lambda_{v,u} \tag{1}$$

Definition 2 (Expected Influence at Level 2): If a node $u \in \Gamma(v)$ gets influenced at level 1 then only it can try to influence

its neighbors. Thus, the expected number of nodes influenced at level 2 by node v is defined by the expectation $\mathbb{E}^{(2)}$ as:

$$\mathbb{E}^{(2)}(v) = \sum_{u_1 \in \Gamma(v)} (\lambda_{v,u_1} \times \sum_{u_2 \in \Gamma(u_1)} \lambda_{u_1,u_2}) \tag{2}$$

Definition 3 (Expected Influence at Level j): The expected number of nodes influenced at level j by node v is defined by the expectation $\mathbb{E}^{(j)}$ as:

$$\begin{aligned}\mathbb{E}^{(j)}(v) = \textstyle\sum_{u_1 \in \Gamma(v)} (\lambda_{v,u_1} \quad &\times \quad \textstyle\sum_{u_2 \in \Gamma(u_1)} (\lambda_{u_1,u_2} \times ... \\ &\times \quad \textstyle\sum_{u_j \in \Gamma(u_{j-1})} \lambda_{u_{j-1},u_j})\end{aligned} \tag{3}$$

Definition 4 (Total Influence of a Node): Total influence of a knowledge agent in the network is the sum of the number of nodes influenced in each level distance from the node. Here the distance referred to the hop distance in the graph. Thus, the total influence $\mathcal{T}(v)$ of an information agent v is

$$\begin{aligned}\mathcal{T}(v) \quad &= \quad \text{Expected nodes influenced at level 1} \\ &+ \quad \text{Expected nodes influenced at level 2} \\ &+ \quad ...\end{aligned} \tag{4}$$

$$\begin{aligned}\mathcal{T}(v) \quad &= \quad \mathbb{E}^{(1)}(v) + \mathbb{E}^{(2)}(v) + ... \qquad (5) \\ &= \quad \sum_{u \in \Gamma(v)} \lambda_{u,v} + \sum_{u_1 \in \Gamma(v)} (\lambda_{u_1,v} \\ &\times \quad \sum_{u_2 \in \Gamma(u_1)} \lambda_{u_2,u_1}) + ... \qquad (6)\end{aligned}$$

This equation can be rewritten using recursive function as

$$\mathcal{T}(v) = \sum_{u \in \Gamma(v)} (\lambda_{u,v} \times (1 + \mathcal{T}(u))) \tag{7}$$

A point to mention here is that, for practical scenario the recursion continues up to the diameter of the network. Because, by definition that covers all the nodes in the network.

C. Total Influence of a Set of Nodes

Total influence of a set of nodes is the expected number of nodes influenced when all the nodes in the set acts as an information agent of the same information. So, here we need to calculate the expected total node influenced by all the nodes in the set. Simply summing up the total influence of each node will not provide with the desire result because information agents may have overlapping neighbors. One node influenced by more than one information agents should be consider in the calculation of expected values.

Equation 7 express the total number of nodes influenced by an agent v. Let us consider the neighborhood of v is denoted by $\eta(v)$. The neighborhood $\eta(v)$ can be expressed by the following recursive function.

$$\eta(v) = \Gamma(v) \cup (\bigcup_{u \in \Gamma(v)} \eta(u)) \tag{8}$$

Probability that a node get influenced by node v is

$$\Pr(v) = \frac{\mathcal{T}(v)}{|\eta(v)|} \tag{9}$$

Given a set of knowledge agents S, probability that a node in the network get influenced by any node of the set S is

$$\Pr(S) = \Pr(\bigcup_{v \in S} v) \tag{10}$$

Applying Inclusion-Exclusion theory on Equation 10 we get,

$$\Pr(S) = \sum_{q=1}^{n=|S|} ((-1)^{q-1} (\sum_{I \subseteq \{1,2,...,n\}, |I|=q} \Pr(\bigcap_{i \in I} v_i))) \tag{11}$$

Based on this probability we can now calculate the value of expected nodes influenced by a set of knowledge agents.

Definition 5 (Total Influence of a Set of Nodes): Total influence of a set of knowledge agent S is defined as:

$$\mathcal{T}(S) = \Pr(S) \times |\eta(S)| \tag{12}$$

where $\eta(S) = \bigcup_{v \in S} \eta(v)$.

By Bonferroni inequality, upper bound of $\Pr(S)$ is $\sum_{v \in S} \Pr(v)$. So the upper bound of total influence of S is

$$\sup \mathcal{T}(S) = |\eta(S)| \times \sum_{v \in S} \Pr(v) \tag{13}$$

III. Hybrid Simulation of Independent Cascade Model using Rough Knowledge Granule

The previous section reports a measure of total influenced by a set of information agents. This measure theoretically provides the expected number of nodes influenced by a set of nodes. However, deriving the value of $\mathcal{T}(S)$ as per the Equation 12 is not practically feasible because the joint probabilities for all the combinations are not available. It is also very hard to measure these joint probabilities in real scenarios. Researchers usually use Monte-Carlo simulation to determine the value of $\mathcal{T}(S)$. As already mentioned in Section I that the process is stochastic and one needs a sufficiently large number of trials to get the reasonable approximation. For a complex network, it is time-consuming, especially for the greedy algorithm, which is known as the best solution for the influence maximization problem. This section proposes a novel simulation methodology based on the concept of rough set theory. This new methodology reduces the stochastic process into two distinct parts. One is the lower approximation of the influenced set, and another one is the upper approximation of the influenced set. The lower approximation of the set is computed from the expectation, and the boundary region is determined with the existing Monte-Carlo simulation. As the problem space for the Monte-Carlo simulation reduced by the number of edges in the lower approximation region, this provides the opportunity to lower the running time sharply.

A. Rough Granules and Knowledge Agents

Rough granule, defined based on the rough set theory, is constructed around all the knowledge agents in the seed nodes. For a social graph $G(V, E)$ let V be the universe of discourse and S be the set of seed nodes. Information given to all the nodes in S will be propagated to the network, and our goal would be to measure the size of this information spread. The rough granule around an information agent is defined as:

$$\underline{V}(S, v) = \{u \in V : u \in \eta(v) \wedge u \notin \bigcup_{w \in S \setminus \{v\}} \eta(w)\} \quad (14)$$

$$\overline{V}(S, v) = \{u \in V : u \in \eta(v)\} \quad (15)$$

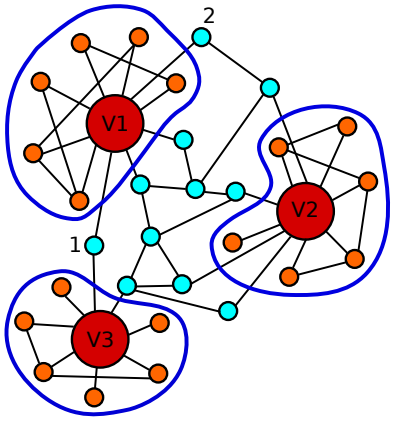

Fig. 1: Knowledge agents and the granules. The lower approximate region is shown in blue circles. Nodes with aqua colors are in the overlapping region and belongs to the boundary region of any two or more knowledge agents. E.g., node 1 belong to the upper approximate region of both V1 and V3, whereas, node 2 is in the boundary region of V1 and V2.

Here, $\underline{V}(S, v)$ is the lower approximation of the knowledge agent v containing all the nodes which can only be influenced by $v \in S$. On the other hand, $\overline{V}(S, v)$ contains all the nodes that can be influenced by any node $u \in S$ as well as by v. Thus, nodes in the boundary region, i.e., $\overline{V}(S, v) \setminus \underline{V}(S, v)$ are in the overlapping influence region. Figure 1 shows a toy example with three knowledge granules. Multiple knowledge agents can influence nodes in aqua color (e.g., 1, 2, etc.). Hence, they belong to the boundary regions of the granule. Nodes with the brown color can only be influenced by the respective knowledge agent. Therefore, they are in the lower approximation of corresponding knowledge granule (marked with blue borderline).

The system of information diffusion will now be represented by the collection of all the rough granules around all the knowledge agents in S. So, for every $v \in S$, there is a rough set defining its granule. The proposed hybrid simulation works on this system of information diffusion in two steps, as explained in the following part of the Section. A block diagram of the whole process is shown in Figure 2.

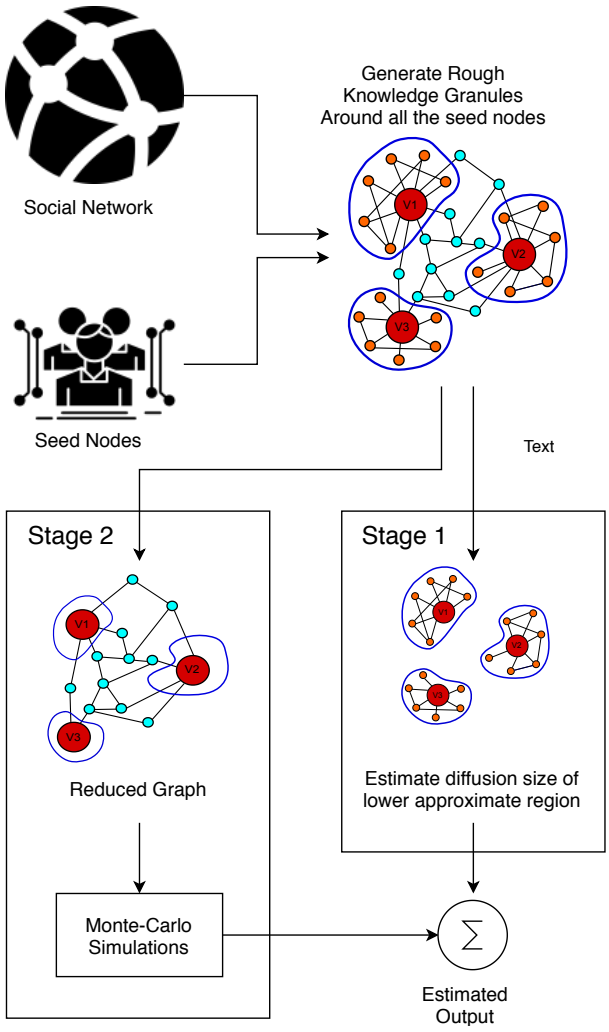

Fig. 2: Block diagram of Hybrid ICM process

B. Stage 1: Expected size in lower approximate region

The first step is to calculate the expectation in the lower approximate region of all the knowledge granules constructed around each $v \in S$. There is no overlapping influence in this region, hence the expected numbers can be calculated by summing up the expectation in each level from the seed nodes. A single run of BFS can be used to calculate the expected size for this case.

C. Stage 2: Reduce Social Graph and Monte-Carlo Simulation

Multiple knowledge agents may influence a node in the boundary region due to the overlapping influence. A reduced graph is constructed by compacting all the nodes in the lower approximation as a super-node and connecting them with all the other nodes in the boundary region (Figure 3). Edge propagation probabilities are calculated based on the path probabilities. A Monte-Carlo simulation is performed on this reduced graph for a sufficiently large number of times to get an approximate influence in the boundary region of all the knowledge agents.

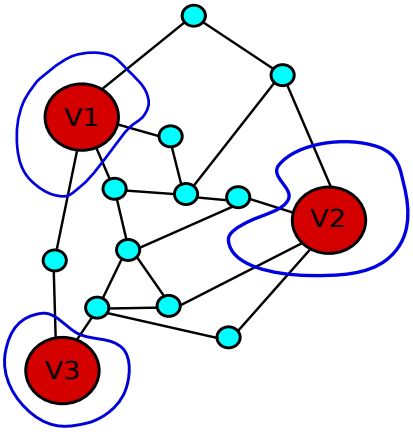

Fig. 3: Reduced graph for staged 2 of Figure 1

TABLE I: Synthetic Data set

Properties/Name	Net200	NetBA
Network Type	Directed	Directed
Nodes	148	500
Edges	604	1984
Avg. Degree	4.0811	3.9680

Thus, one value is obtained from the stage 2 as the size of the boundary/overlapping region of the knowledge agents and $|S|$ of values as the size of lower approximate regions of all the knowledge agents. Finally, the cumulative value of all of these is the total size of spreading in the network when we propagated information via the set S.

IV. Experiments and Results

Experiments have been performed on two synthetic data to compare the between traditional Monte Carlo simulation and the proposed hybrid simulation method. The Greedy algorithm for influence maximization has been executed using both the aforementioned methods, and the comparative results are reported in this section.

A. Description of Data Set

Details of the two synthetically generated networks used in the experiments are shown in the Table I. These networks are referred to as Net200 and NetBA. Net200 is generated using LDBC DATAGEN [19], and the data is based on Facebook degree distribution. NetBA is generated using networkx internal graph generating algorithm which follows Barabási-Albert preferential attachment model [20]. Both the network used in the experiments are considered as a directed network.

B. Results

A version of the greedy algorithm for influence maximization [1] is implemented using Monte-Carlo simulations (about 1000 runs) and using the proposed hybrid simulation. Recorded execution time for extracting different number of seed nodes (i.e., influencing nodes) is reported in Table II.

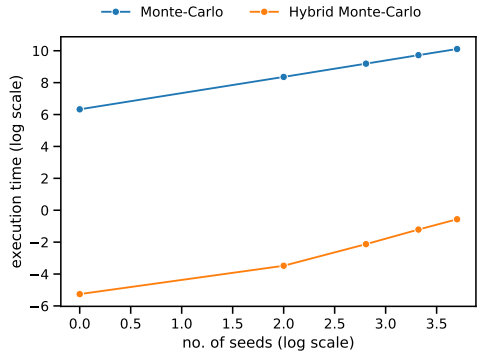

Fig. 4: Execution time for Net200.

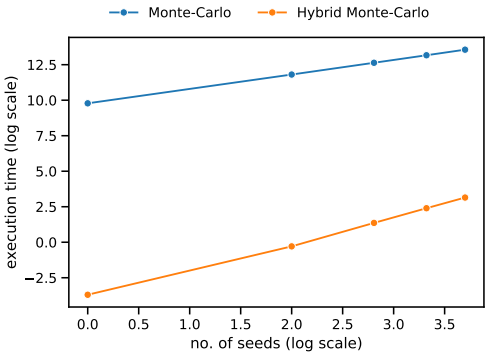

Fig. 5: Execution time for NetBA.

TABLE II: Execution time of Greedy algorithm with different simulation techniques for different values of k. (In Second)

Data set	k	Monte-Carlo Simulation	Hybrid Simulation
Net200	1	80.311	0.0261
	4	328.295	0.0895
	7	584.191	0.2290
	10	845.544	0.4319
	13	1103.660	0.6758
	15	1279.231	1.0737
NetBA	1	879.732	0.0771
	4	3583.240	0.8172
	7	6363.880	2.5676
	10	9183.980	5.2605
	13	12057.800	8.8544
	15	14447.561	11.7210

TABLE III: Mean and variance of different experiments with respect to propagation probabilities (PP) and k.

			Monte-Carlo		Hybrid Simulation	
Data set	PP	k	Mean	Variance	Mean	Variance
Net200	0.1	15	7.866	10.201	8.303	11.175
Net200	0.1	10	5.871	7.91	5.873	7.509
NetBA	0.05	10	2.144	2.225	2.152	2.003
NetBA	0.05	7	1.761	2.12	1.727	1.81

It is clear from the record that the proposed hybrid simulation based greedy algorithm runs magnitude faster than the traditional Monte-Carlo method based greedy algorithm. For better understanding a pictorial plot is provided in the Figures 4 and 5. With both the final 15 seeds resulted by both the greedy algorithms, separate Monte-Carlo simulations (about 1000 runs) were performed to see the comparative size of the spread. It is found that the spread size is comparative for Net200 (28.494 for Monte-Carlo and 28.815 for the hybrid simulation) and very close for NetBA (76.689 for the Monte-Carlo and 80.914 for hybrid simulation).

As the size of the spread is different, close observation was required. It is found that the seed set obtained by the two methods are different, and about 20% seeds in Net200 and about 30% in NetBA are not common. These findings lead to a question, "Will the same seed produces same spread size with both the simulation?" A new experiment is conducted to answer the question in concern. A seed set generated uniformly randomly with a pre-defined size k is used to estimate the spread using both the Monte-Carlo and hybrid simulation. The experiment is repeated for 100 times with separate independent choices of seed. Mean and variance of these experiments are presented in the Table III. It is clear from the data that the mean is very close for most of the experiment demonstrating the validation of the proposed hybrid simulation.

V. Conclusions

The contribution of the paper is two fold. Section II proposed a new measure to quantify the expected spreading size of the information under the independent cascade model of diffusion. The measure is called *total influence*. A *total influence* measure to calculate the spreading size for a set of nodes is also presented in the same Section. Total influence measure is based on the probability theory; Obtaining the joint probabilities therein is difficult to calculate for practical social networks. To overcome this difficulty, Section III outline a novel methodology of *hybrid simulation* which quickly measures the size of information spread under the independent cascade model of information diffusion. The hybrid simulation uses rough set theory to define rough knowledge granules.

Experimental results show that the proposed hybrid simulation can quantify the size of information spreading magnitude faster than Monte-Carlo simulation. Thus, it improves the running time of the greedy algorithm for influence maximization. It is also showed that the cascade size measured by the hybrid method is comparable with more traditional Monte-Carlo simulations. The simulation of IC model needs to handle two possible uncertainties arises in the IC method; one is due to the probability of activation of edges, and the second one is due to the overlapping influence region. The proposed simulation method able to produce comparable results with more traditional Monte-Carlo experiments because the method correctly modeled the uncertainties due to the overlapping influence region with the help of the rough set. Accordingly, hybrid-simulation uses traditional Monte-Carlo methods only for the overlapping region while it uses probability theories to calculate the influence in the non-overlapping region.

All the experiments were conducted on the directed synthetic social graph. A future interest of the work would be to testify the hybrid simulation methodology for undirected as well as real social networks.

References

[1] D. Kempe, J. Kleinberg, and É. Tardos, "Maximizing the spread of influence through a social network," in *Proc. of the 9th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. New York, NY: ACM Press, 2003, p. 137.

[2] M. Granovetter, "Threshold models of collective behavior," *The American Journal of Sociology*, vol. 83, no. 6, pp. 1420–1443, 1978.

[3] E. M. Rogers, *Diffusion of Innovations*, 5th ed. Free Press, 2003.

[4] P. Domingos and M. Richardson, "Mining the network value of customers," in *Proc. of the 7th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. San Francisco, CA: ACM, 2001, pp. 57–66.

[5] M. Richardson and P. Domingos, "Mining knowledge-sharing sites for viral marketing," in *Proc. of the 8th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. Edmonton, Alberta: ACM Press, 2002, pp. 61–70.

[6] D. J. Watts, "A simple model of global cascades on random networks," *Proceedings of the National Academy of Sciences*, vol. 99, no. 9, pp. 5766–5771, 2002.

[7] A. L. Hill, D. G. Rand, M. A. Nowak, and N. A. Christakis, "Infectious Disease Modeling of Social Contagion in Networks," *PLOS Computational Biology*, vol. 6, no. 11, pp. 1–15, 2010.

[8] J. Leskovec, A. Krause, C. Guestrin, C. Faloutsos, J. VanBriesen, and N. Glance, "Cost-effective outbreak detection in networks," in *Proc. of the 13th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. San Jose: ACM Press, 2007, pp. 420–429.

[9] T. Carnes, C. Nagarajan, S. Wild, and A. van Zuylen, "Maximizing influence in a competitive social network: a follower's perspective," in *Proceedings of the ninth international conference on Electronic commerce*. ACM, 2007, pp. 351–360.

[10] W. Chen, Y. Wang, and S. Yang, "Efficient influence maximization in social networks," in *Proc. of the 15th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. Paris: ACM Press, 2009, pp. 199–208.

[11] R. Narayanam and Y. Narahari, "A Shapley value-based approach to discover influential nodes in social networks," *IEEE Transactions on Automation Science and Engineering*, vol. 8, no. 1, pp. 130–147, 2011.

[12] C. Wang, W. Chen, and Y. Wang, "Scalable influence maximization for independent cascade model in large-scale social networks," *Data Mining and Knowledge Discovery*, vol. 25, no. 3, pp. 545–576, apr 2012.

[13] S. K. Pal, S. Kundu, and C. A. Murthy, "Centrality measures, upper bound, and influence maximization in large scale directed social networks," *Fundamenta Informaticae*, vol. 130, no. 3, pp. 317–342, 2014.

[14] W. Chen, Y. Yuan, and L. Zhang, "Scalable influence maximization in social networks under the linear threshold model," in *Proc. of 2010 IEEE International Conference on Data Mining*. Sydney: IEEE, 2010, pp. 88–97.

[15] C. Zhou, P. Zhang, W. Zang, and L. Guo, "On the Upper Bounds of Spread for Greedy Algorithms in Social Network Influence Maximization," *IEEE Transactions on Knowledge and Data Engineering*, vol. 27, no. 10, pp. 2770–2783, oct 2015.

[16] J. Shang, "Dynamic Update Upper Bounds Influence Maximization Algorithm," in *Proc. of the 2018 2Nd International Conference on Computer Science and Artificial Intelligence*. Shenzhen, China: ACM, New York, 2018, pp. 212–217.

[17] S. Kundu and S. K. Pal, "Fuzzy-rough community in social networks," *Pattern Recognition Letters*, vol. 67, pp. 145–152, 2015.

[18] ——, "Double Bounded Rough Set, Tension Measure and Social Link Prediction," *IEEE Transactions on Computational Social Systems*, vol. 5, no. 3, pp. 841–853, 2018.

[19] A. Prat, "DATAGEN: data generation for the Social Network Benchmark," 2015. [Online]. Available: http://ldbcouncil.org/blog/datagen-data-generation-social-network-benchmark

[20] A.-L. Barabási and R. Albert, "Emergence of scaling in random networks," *Science*, vol. 286, no. 5439, pp. 509–512, oct 1999.