Information Propagation Game: a Tool to Acquire Human Playing Data for Multi-Player Influence Maximization on Social Networks

Hung-Hsuan Chen
Computer Science and Engineering
Pennsylvania State University
State College, PA, USA
hhchen@psu.edu

Yan-Bin Ciou
Computer Science and Information Engineering
National Taiwan University
Taipei, Taiwan
b97902012@ntu.edu.tw

Shou-De Lin
Computer Science and Information Engineering
National Taiwan University
Taipei, Taiwan
sdlin@csie.ntu.edu.tw

ABSTRACT

With the popularity of online social network services, influence maximization on social networks has drawn much attention in recent years. Most of these studies approximate a greedy based sub-optimal solution by proving the submodular nature of the utility function. Instead of using the analytical techniques, we are interested in solving the diffusion competition and influence maximization problem by a data-driven approach. We propose Information Propagation Game (IPG), a framework that can collect a large number of seed picking strategies for analysis. Through the IPG framework, human players are not only having fun but also helping contributing the seed picking strategies. Preliminary experiment suggests that centrality based heuristics are too simple for seed selection in a multiple player environment.

Categories and Subject Descriptors

G.2.2 [**Discrete Mathematics**]: Graph Theory—*Graph algorithms*; E.1 [**Data Structures**]: Graphs and networks; H.4 [**Information Systems Applications**]: Miscellaneous

General Terms

Design, Experimentation, Measurement

Keywords

Diffusion Network, Influence Maximization, Information Propagation, Game, Independent Cascade, Linear Threshold

1. INTRODUCTION

Studies have shown that people's decisions on whether to adopt an innovation or product are not purely based on the objective factors of the object, but also the colleagues' or friends' opinions [6, 10]. Therefore, to promote an item, it is essential to identify the key *influencer*s to advertise to, with the hope that these influencers can disseminate the message to a larger group of individuals. Given a network, discovering the k most influential nodes that can eventually propagate certain information to the most nodes is defined as the *influence maximization* problem. Researchers have shown that influence maximization is NP-hardness [8]. Several selection strategies, such as greedy based approach and the centrality based heuristics, were posed to effectively select the initial nodes, called seeds, for influence maximization [8, 5, 9].

In real life, however, it is very likely that more than one mutually exclusive products compete with each other for the market share. For example, two competitive companies usually release similar products within a short range. Due to the exclusiveness natural of their products, each of the company would like to look for the influential personnels on social networks to maximize their benefit in advertisement. This kind of problems, in general called *diffusion competition problem*, have been studied by a number of researchers.

The previous works studied the seed picking strategy of diffusion competition and influence maximization by proposing some intuitive selection strategies [2, 3, 4]. While multiple participants are involved in competing for the seed nodes, the selection policy can become very complicated because every move of a participant is affected not only by its previous moves, but also the previous selections by other camps. Under such circumstance, the strategies designed for single-player influence maximization such as a centrality-based strategy might be too naïve to be effective. Here we argue that when multiple participants were involved, the seed-selection competition can be modeled as a multiple-player game, where each player has some budget to influence some seed nodes for propagation each turn. Thus, it is not hard to imagine that human players can derive better strategies with the accumulation of experiences on playing this game, just like that a novice promotion manager can gradually learn how to outwit their opponents and eventually become master in such advertisement campaign. Similarly, the increasing amount of historical playing record would allow an intelligent program to learn how to play such a game through machine learning (in particular reinforcement learning) models. To effectively collect a large number of seed picking strategies for advanced learning not only by human but also by machines, we design Information Propagation Game (IPG), a chess-like gaming engine in which the chessboard is a set of nodes and edges, and players compete with each other with the goal to maximize the influenced nodes. While the users are playing the game for fun, they are helping collecting data about sophisticated seed picking strategy for learning.

A demonstration page of IPG is available at `http://www.csie.ntu.edu.tw/~hhchen1/ipg/`.

2. IPG RULES

Before diving to the detail, we define some terms used in the paper. Each person involves in IPG is called a *player*. Each player takes turn to perform a legal action, called a *move*. The IPG performs information diffusion in a discrete step, named as a *round*. When the diffusion completes, the *utility* of a player is defined as the number of nodes activated by this player.

Although IPG is designed as a platform for multiple players, it also provides a single player mode. We introduce the rules for both scenarios in this section.

2.1 Rules of a Single Player Game

Given a social network and a known diffusion strategy such as Independent Cascade (IC) or Linear Threshold (LT) model, the player is asked to select k seeds to maximize the influence. Next, the IPG displays the diffusion process and returns the number of eventually activated nodes. To make the game more interesting, the IPG also shows the current best records given this network topology and diffusion strategy. This setup challenges the players to break the current record. The data collected can be used to analyze the influence maximization problem and facilitate the policy learning from machines.

2.2 Rules of a Multiple Player Game

In multi-player IPG game, each player is asked to choose same amount of seeds for propagation. The IPG performs diffusion until no more activations are possible. The side that occupies the most seeds after propagation wins the game. Unlike the single-player influence maximization problem, here several players may compete for activating an inactive node. Since most of the state-of-the-art diffusion models are designed for the non-competition scenario, we design two diffusion models for multi-player diffusion competition: Competitive Independent Cascade (CIC) model and Competitive Linear Threshold (CLT) model, which are natural extensions of the famous IC model and LT model, to deal with the competition scenario.

In IC model, each active node belong to only one person and has a single chance to activate each of its neighbors. In CIC model, each active node has to belong to one player (i.e. considered as activated by that player). Similar to IC model, in CIC when all active neighbors fail to activate a node, this node would remain inactive. If multiple nodes belonging to multiple players have successfully activated the same node, then the chance that this node is activated by a player will be proportional to the number of neighbor nodes belonging to this player that successfully activated this node. For example, if there are two players each has 2 and 3 nodes that successfully activated a target node, then there are $2/5$ and $3/5$ respectively chance this target node belongs to the first and second player.

For LT model, an inactive node is activated if the summation of the influential power of its active neighbors exceeds the threshold. In a diffusion competing scenario, if an inactive node can be activated by more than one player, the chances of the would-be-activated node belonging to a certain player is proportional to the accumulated influential power of that player. For example, if the threshold of an inactive node is $1/4$ and there are three players attempt to influence the node with the summation influential power be $1/5$, $1/3$, and $1/2$ respectively, the inactive node will choose the second player with probability $(1/3)/((1/3)+(1/2)) = 2/5$ and the third player with probability $(1/2)/((1/3)+(1/2)) = 3/5$. The first player cannot activate the node because its influential power does not exceed the threshold.

2.3 Extension Rules for Multiple Player Game

Assuming that each player can only select the seed nodes at the beginning might not be the case in many real world scenarios. In marketing, companies usually have advertising budgets allocated every month or every season; thus they are able to affect their customers in multiple rounds. In the case of epidemic disease prevention, the disease control center usually is aware of only a portion of the patients in the beginning. As the controlling process goes on, the center usually prevents the spread of infection in one area, and in the meanwhile can discover more patients in other areas. This can be viewed as one player (the disease) selects the nodes for propagation and the other player (the disease control center) adjust the inoculation strategy to minimize the spread in multiple rounds.

Motivated by this, the extension rules allow the players to select the seeds for multiple rounds. For example, one possible variation is each player can select k' nodes every round. Another possible variation is to provide each player k selection quota initially, and the players can decide when to use them during the play.

3. DESIGN NOVELTIES

The IPG is implemented using Java Universal Network Framework (JUNG)[1]. As shown in Figure 1, the system is highly modular so that the important components can be updated or enriched with little effort. We introduce the design novelties of IPG in this section.

3.1 AI Strategy

The player can choose to play with an AI agent in the multi-player game. The current AI strategies are based on three heuristics: the centrality based strategy, the blocking strategy, and the maximum neighbor coverage strategy.

Centrality measures are usually used to quantify the importance of a node within a network. The centrality based AI first chooses one centrality measure from the available options. It then selects the k inactive nodes with the highest centrality as the seeds. Currently, three important centrality measures are calculated: degree centrality, betweenness centrality, and closeness centrality. Other centrality measures, such as Katz centrality and Eigenvector centrality, can be easily integrated into the framework.

While the first player seems to have predominance by selecting the most influential nodes, an effective responding strategy of the second player could even up such advantage. Here, we propose two responding AI strategies: the blocking strategy and the maximum neighbor coverage strategy. The blocking AI examines the inactive neighbors of the seeds selected by the competitors, and among them chooses the ones with higher centrality values. By blocking the diffusion path that are potentially influential, the second player could hopefully decrease the first player's influence. The other responding strategy, maximum neighbor coverage, selects the nodes whose neighbors are largely overlapped with the neighbors of the first player's selected seeds. Maximum neighbor coverage AI aggressively competes the nodes with the first player in the beginning to prevent the first player gaining obvious initial advantage.

3.2 Network Topology Selection

Different seed picking strategies may gain different level of effectiveness on different types of network structures.

[1]http://jung.sourceforge.net/

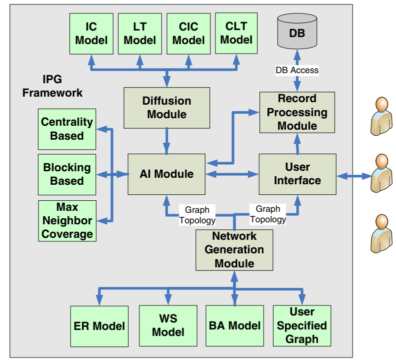

Figure 1: The system components

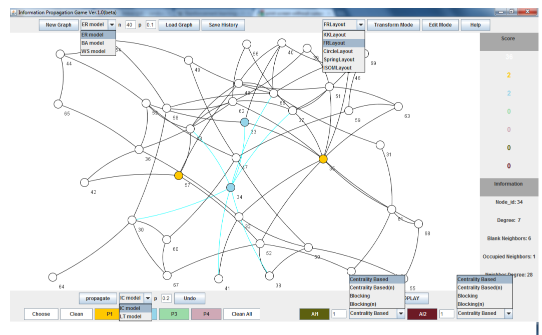

Figure 2: The snapshots of the IPG framework

The IPG system allows the generation of different types of networks which facilitates further study about the relationship between network structure and the seed picking strategy. The current IPG framework can produce graphs of the following types: Erdős-Rényi (ER) network [7], Watts-Strogatz (WS) network [11], and Barabási-Albert (BA) network [1]. Furthermore, IPG also allows users to load an existing social network from file.

3.3 Interface Novelty

Users may find the IPG framework provides a novel interface that facilitates seed selection by human. We introduce several of the characteristics here.

Players can view the nodes' information, including the centrality values, the number of occupied and non-occupied neighbors, and the summation of neighbors' degree. These information can be useful clues for the players to decide which nodes to pick.

IPG allows the players to undo their decisions if they figure out a better move. Since we are interested in collecting effective strategies, we hope the users to play as well as they can.

In addition, the total number of occupied and unoccupied nodes are displayed in real time, as shown in Figure 2. This feature is especially helpful when applying the extension rules in which the players can select the seeds for multiple rounds. They can decide to use the remaining seed picking quota at a more proper moment.

3.4 AI to AI competition

While the human to human and human to AI competition allow us to collect valuable playing records with complicated seed picking strategies, AI to AI competitions enables us to accumulate a large number of playing records with little effort. IPG framework allows users to select the AI strategy from the available strategy list. Through the large number of automatic playing records, an reinforcement learning agent might be able to discover possible response strategy given different types of network topologies, diffusion models, and the competitor's seed picking strategies.

4. EXPERIMENTS AND DISCUSSION

Table 1: Comparison of the second player's strategy (BA network with 1000 nodes, CIC diffusion model with activation rate 0.1, the first player tend to pick high degree nodes, U_i: the number of activated nodes of player i)

	Deg.	B2n.	Cls.	Blk-Deg.
U_1	267 (0%)	261 (−2%)	272 (+2%)	259 (−3%)
U_2	152 (0%)	158 (+5%)	163 (+8%)	174 (+15%)

Table 2: Comparison of the second player's strategy (BA network with 1000 nodes, CLT diffusion model with threshold 0.6, influence value 0.3, the first player tend to pick high degree nodes, U_i: the number of activated nodes of player i)

	Deg.	B2n.	Cls.	Blk-Deg.
U_1	886 (0%)	876 (−1%)	860 (−3%)	876 (−1%)
U_2	114 (0%)	120 (+5%)	140 (+23%)	124 (+8%)

For a two player game, we investigate the second player's strategy when the first player tends to select the high degree nodes under BA network through IPG. By varying the second player's strategy be degree centrality (Deg.), betweenness centrality (B2n.), closeness centrality (Cls.), and the degree-based blocking strategy (Blk-Deg.), i.e., the second player selects high degree nodes from the inactive neighbors of the first player's seeds, the utility U_i (number of final influenced nodes) of player i, given the diffusion model be CIC and CLT, are shown in Table 1 and Table 2 respectively. Under CIC diffusion model, the second player's utility improves 15% by applying degree based blocking strategy instead of degree centrality. The result suggests that centrality heuristics are too simple when the competitors exist. Although the improvement under CLT model is not as obvious as the CIC model, degree based blocking strategy still shows the advantage over degree centrality and betweenness centrality strategies.

A questionnaire survey is conducted to the pilot player who played dozens of games in IPG. The report shows that the first player apparently has some advantages given small graphs. However, the second player can apply a more aggressive seed picking strategy, such as the maximum neighbor coverage strategy introduced in Section 3.1, to make up the disadvantage. The experiences shared by the real player also support our claim of simply selecting the high centrality nodes is too naïve in a competing scenario.

For future work, we will study the records retrieved by IPG systematically. In particular we plan to apply reinforcement learning, which uses the reward feedback, i.e., the utility in IPG, to learn a good policy for players in different positions. We are also interested in applying game theory studies to diffusion competition problem. Another interesting research topic is the relationship between rate of spread of diffusion and other player's next move.

5. ACKNOWLEDGMENTS

This work was also supported by National Science Council, National Taiwan University and Intel Corporation under Grants NSC 100-2911-I-002-001, and 101R7501.

6. REFERENCES

[1] A. Barabási and R. Albert. Emergence of scaling in random networks. *Science*, 286(5439):509–512, 1999.

[2] S. Bharathi, D. Kempe, and M. Salek. Competitive influence maximization in social networks. *Internet and Network Economics*, pages 306–311, 2007.

[3] C. Budak, D. Agrawal, and A. El Abbadi. Limiting the spread of misinformation in social networks. In *Proceedings of the 20th international conference on World wide web*, pages 665–674. ACM, 2011.

[4] 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*, pages 351–360. ACM, 2007.

[5] W. Chen, Y. Wang, and S. Yang. Efficient influence maximization in social networks. In *Proceedings of the 15th ACM SIGKDD international conference on knowledge discovery and data mining*, pages 199–208. ACM, 2009.

[6] J. Coleman, E. Katz, and H. Mentzel. Medical innovation: Diffusion of a medical drug among doctors. *Indianapolis: Bobbs-Merrill*, 1966.

[7] P. ERDdS and A. R&WI. On random graphs i. *Publ. Math. Debrecen*, 6:290–297, 1959.

[8] D. Kempe, J. Kleinberg, and É. Tardos. Maximizing the spread of influence through a social network. In *Proceedings of the ninth ACM SIGKDD international conference on knowledge discovery and data mining*, pages 137–146. ACM, 2003.

[9] J. Leskovec, A. Krause, C. Guestrin, C. Faloutsos, J. VanBriesen, and N. Glance. Cost-effective outbreak detection in networks. In *Proceedings of the 13th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 420–429. ACM, 2007.

[10] B. Ryan and N. Gross. The diffusion of hybrid seed corn in two iowa communities. *Rural sociology*, 8(1):15–24, 1943.

[11] D. Watts and S. Strogatz. Collective dynamics of small-worldnetworks. *Nature*, 393(6684):440–442, 1998.