Evaluation of Heuristic Marketing Approaches in Social Networks

Bachelor Thesis*

Raphael Mitsch
a1006529@unet.univie.ac.at

Universität Wien, Universitätsring 1, Wien, Österreich

February 18, 2014

Keywords: Pricing, Monetizing Social Networks, Marketing, Uniform Additive Model, Posted Price

* Supervised by: Univ.-Prof. Dr. Monika Henzinger, Dr. Wolfgang Dvořák. Field of study: Informatics / Scientific Computing.

Table of Contents

List of Figures

Abstract. This paper discusses and evaluates influence-and-exploit marketing strategies aiming to maximize revenue in the context of real-world social network datasets with positive network externalities. Preceding work [8] introduces the concept of capizalizing on these externalities by incorporating them into approximation algorithms and proposes offering to one client at a time at individual prices. Subsequent research [4] suggests a limitation of price-discrimination by segregating clients into groups whose members are all offered the same price, moreover a finite number of rounds is specified. These measures are propounded on grounds of an expected decline in processing time (especially in large-scale networks) and to minimize perceived unfairness among clients owing to unequal pricing.
Based on aforementioned sources, mainly [4], we analyze the algorithm as defined there. In doing so we strongly focus on how clients are mapped to groups by reason of certain characteristics inherent to them. Four of these group mapping heuristics are compared quantitatively - by measuring the generated revenues - and qualitatively. The clients' valuation function follows the Uniform Additive Model (UAM).
The most significant findings are that (i) choosing a group's minimal myopic price as group price is not necessarily the most effective option, (ii) the groups at the top and at the bottom, e.g. those containing the clients willing to pay the most respectively the least, contribute very little to the total revenue[1] and (iii) involving connections not only to adjacent clients in the influence set but also to those *not* in the influence set increases revenue under particular circumstances.

1 Introduction

Lately, the rapidly increasing influence and importance of social networks has given rise to new approaches in the research of marketing strategies concerning this domain. The growing amount of time and attention spent by the average user on such platforms has led to various efforts to systematically utilize the inherent properties of social networks to discover new and more efficient ways to promote products. One way to achieve this is by factoring in and capitalizing on so-called network externalities - the impact of buyers on other, adjacent members of the network, who haven't reached a buying decision yet. They ought to reflect well-known concepts describing a group's social dynamics such as word of mouth, peer pressure, networking effects etc. Positive externalities lead to potential clients being more inclined to buy a product, while negative externalities lessen the chance they will do so.

Thus, when pursuing the goal of increasing the earnings gained from selling a product, one should take these externalities - and how to best make use of them - into consideration. A group of marketing strategies named *influence-and-exploit* strategies focuses on this idea: A certain fraction of the member base of the

[1] The former on account of the low number of clients, the latter because the average myopic price is close to zero with the graphs used.

social network (chosen randomly or by the influence they wield over other nodes or based on other characteristics) is given the product for free or discounted, hence assuring they accept the offer and thereby *influence* related clients to be more biased towards buying. In the following *exploit* phase other clients are offered the product for a non-zero amount of money, with the hope of them more than outweighing the losses of potential revenues in the influence round.

One distinctive feature amongst the plethora of varieties of influence-and-exploit strategies [6, 7, 15, 4, 8, 1] is the question of how to handle the process of offering to the clients. While some approaches propose to offer to one client at a time [8], others [4] advocate, by reason of the unfairness possibly perceived by clients due to the nature of fully discriminative pricing, splitting the network's total population into groups and charging all clients of one group the same (posted) price. [4] as well advances the idea of additionally limiting the exploit phase to a fixed number of rounds, as the offering process may require much more time otherwise, especially in large-scale networks.

This paper will take on these so-called *(k,l)*-PP strategies[2] and introduce several heuristics for the assignation of clients to groups. With this we aim at increasing revenues and deepening the understanding of the implications of these mechanisms. Accordingly, aforesaid group mapping heuristics are evaluated using real-world anonymized data extracted from diverse social networks. This is accomplished by means of exploring the effectiveness of different configurations (composed of one of the mentioned heuristics, group price selection procedures, the number of groups and rounds admitted, the probability for any client to be selected as a member of the influence set and a divisor used for the calculation of lower and upper group thresholds, all of which are discussed in detail later). The most promising combinations are examined more thoroughly; their results interpreted in order to find clues helpful for the optimization of this subclass of influence-and-exploit algorithms.

In a nutshell, the investigated situations show that while the strategy put forward by [4] produces high gains and exhibits a very steady performance, there are certain circumstances respectively configurations in which it is does not the best scoring heuristic. Besides, open questions about the development of revenues as a function of the quantity of rounds and groups with one of the heuristics designed for this paper remain, as it's behaviour defies the author's expectations on that score. Furthermore, it is demonstrated that the optimal probability for entries into the influence set for the datasets researched indeed conincides with the hypothetical approximation deduced by [4]. Ultimately, measurements of the distribution of the myopic price sums among groups display a high degree of inequality; the amount of income contributed by members of the lowest ranking groups can essentially be neglected in comparison to those supplied by the groups in between.

[2] The terminology is clarified in detail in the following section.

2 Preliminaries

2.1 Definitions

As we build strongly on the theoretical research of [8] and [4], we quote some of their definitions and explanations regarding the applied terminology we consider necessary for the understanding of our paper. We start with a more precise description of the fundamental setting and context (along the lines of [4]).

Our goal is to maximize our revenue in selling a digital product, whose production costs are zero and which we can produce in unlimited numbers - e.g. a website, software, etc. - in a social network. This network of n clients is defined as $G = (V, E)$, whereas V is the set of clients and $E \subseteq V \times V$ the connections between these clients.

Positive network externalities are considered, i.e. clients valuate the product higher if adjacent (related) clients are in possession of it. Negative externalities are not incorporated into our model. When A is the set of clients who have already acquired the product and N_i the set of neighbours of client i, then the valuation of client is defined as $i \in V$ by $v_i : V \to \mathbb{R}_{\geq 0}$, so that $v_i(A) = v_i(A \cap N_i)$. We attempt to utilize the network structure by using the influence the relationships between clients (the aforementioned positive network externalities) exert over these clients' valuations of the product. One way to accomplish this is by offering the product to each client $i \in V$ in a certain order for a certain price $p_i \in \mathbb{R}$, both of which we want to compute.

Clients are presumed to be individually rational and to have quasi-linear utility functions. As only one offer is proposed to each client, speculations on falling prices on part of the clients would be pointless - they either buy the product under the circumstances prevailing when the offer is submitted or don't purchase it all. Therefore a client i purchases the product *only* and *always* when the valuation is larger than or equal to the price: $p_i \leq v_i(A)$.

Uniform additive model. We use the *uniform additive model* (UAM) for the client valuation function in our model; one of several valuation functions feasible for this purpose. Alternatives - not further discussed here - comprise the symmetric, concave graph, simple additive and (deterministic) submodular model among others. [8] and [4] contain more information on this topic. [8] also gives a description of the influence model in general and the UAM in specific, the latter of which we summarize here. We proceed on the assumption that the seller as well as the buyers know the distribution from which our randomized values are selected. That is, the seller knows how the values of $v_i(A)$ are distributed for all $A \subseteq V$ and $i \in V$. Buyers' values are premised to be distributed independently from each other.

Definition 1 ([8]). *In the uniform additive model, there [are] weights $w_i j$ for all $i, j, \in V$. The value $v_i(S)$, for all $i \in V$ and $S \subseteq V \setminus \{i\}$ is drawn from the uniform distribution* $[0, \sum_{j \in S \cup \{i\}} w_{i,j}]$.

Myopic prices. Given an influence set (IS), that is a set of clients already owning the product, the *myopic price* (MP) of a client $i \notin IS$ is the price that maximizes the expected payment of this client to the seller. In the UAM a client's myopic price is calculated as the sum of the weights of all incoming edges connected to members of the influence set - who we have no need to calculate a MP for, as they have already acquired the product for free. Given that m is the myopic price of a non-IS node, I the set of incoming edges connected to influence set members and w_i the weight attributed to edge i, it follows that:

$$m = \frac{\sum_{i \in I} w_i}{2} \tag{1}$$

As apparent from the preceding equation, the MP is an convenient indicator for the frequency and intensity of connections between the investigated client and it's neighbours in the influence set, whereas adjacent clients not part of it don't carry any weight.

Influence-and-exploit strategies. A marketing strategy in it's entirety defines when and at what prices the product is offered to the clients. As a subset of this class of algorithms, *influence-and-exploit* strategies are characterized by their two-phased approach aimed at the utilization of positive externalities in the network - 1) *influence* step: Give away items for free or discounted. 2) *Exploit* step(s): Offer to other inactive clients, who are, on grounds of the influence exerted on them by the clients addressed in the influence step, more willing to buy. Thus the totaled revenue is anticipated to increase.

***(k, l)-PP* strategies.** [4] discusses influence-and-exploit strategies with clients divided in l groups and k rounds (the influence step included, so $k = 2$ would be the influence step followed by one round in which the offering to inactive clients takes place). Moreover, to affect a client's bias towards the product, a neighbouring client has to have acquired the product in a previous round:

Definition 2 ([4]). *In* $(k, l) - PP$ strategies*, clients are split into k groups and l rounds such that*

(a) each client belongs to exactly one group and round,
(b) all clients in the same group are offered the same price independent of their round and
(c) only clients that have purchased the product in a previous *round can influence the valuation functions of the clients in the current round, but this influence is independent of their group.*

For instance, a group could model all clients that live in the same country or state, influential clients, or an age group.

2.2 Data

Four anonymized real-world graph datasets were used for testing purposes: Two during development and two others after the implementation was completed - to avoid overfitting the model to the data. A characteristic shared by all of them is that only node ID and, in two cases, edge weights signifying the intensity of contact between two parties are available for analysis due to the anonymization conducted on behalf of preserving the network member's privacy. While this action is understandable and necessary, access to personal data points like age, gender, location etc. would have opened up interesting possibilities in designing new heuristics.

In consideration of our requirements in terms of processing time and memory usage we deemed it sufficiently efficient to load the datasets directly from file; no graph databases, triple stores or other, more common SQL database software are used. Potential alternatives were SparkleDB, a triple store, as well as Neo4j[3] and AllegroGraph, both graph databases.

Two of the graphs (Epinions and Slashdot) were fetched from the Stanford network analysis platform (SNAP)[4], the other two (Advogato and DBLP) were retrieved from KONECT, a project similar to SNAP that was launched and is maintained by the Unversity of Koblenz-Landau. A short summary of the examined datasets follows, for additional information see their respective sources.

Dataset	**Vertices**	**Edges**	**Weights**	**Directed**	**Phase**
Epinions	75879	508837	None; no multiple edges.	No	Development
Slashdot	82168	948464	None; no multiple edges	No	Development
Advogato	6551	51332	Positive weights; no multiple edges.	Yes	Evaluation
DBLP	1248427	17631144	None; multiple edges.	No	Evaluation

- **Epinions** [13] A platform for consumer reviewing products. Members of epinions.com may "trust" each other - someone who is trusted by many other participants, besides his higher reputation in the community, is assessed as more objective by the site's algorithms and his/her reviews therefore treated as more valuable and important.
- **Slashdot** [14] slashdot.org is a popular news website strongly concerned with topics in technology and, to a somewhat lesser extent, in science. It allows users to mark peers as *friends* or *foes*, thus establishing connections between them that may be exploited for our purposes. Nonetheless, the data available doesn't contain weighted edges.
- **Advogato** [11] Being one of the first social network platforms and a self-referred *"[...] research testbed for group trust metrics and other social networking technologies"*, advogato.org is an obvious aspirant to consider when covering topics in this field of research. Advogato enables trust relationships in three different nuances, correlating with three different edge weights, between users.

[3] Or the upcoming C++ version of this framework, Neo4c.

[4] Quoting from snap.stanford.edu: "[SNAP] is a general purpose network analysis and graph mining library. It is written in C++ and easily scales to massive networks with hundreds of millions of nodes, and billions of edges."

- **DBLP** [12] The DBLP Computer Science Bibliography, hosted by the Universität Trier at dblp.uni-trier.de, is a database tracking scientific publications on computer science and related fields. In this respect it represents an exception among these datasets as it is not a social network per se. In the used graph each vertex represents an author, each edge a collaboration between two authors. Multiple edges between two vertices - representing multiple collaboration between the same two authors - are permitted. In preparing the dataset, these were condensed to a single edge, whose weight was adjusted accordingly. Given that E is the set of edges between two vertices, w_j the individual edge weight[5] and w the according total weight, the following applies:
$$w = log(1 + \sum\nolimits_{j \in E} w_j) \tag{2}$$

3 Investigated Marketing Strategies

As already discussed earlier, four different heuristics - all based on the approach of [4] - were designed and analyzed (see Section *Evaluation*). Additionally, we consider six different price selection mechanisms (PSMs) to be combined arbitrarily with each GMH. The integration of the library of the aforementioned Stanford Network Analysis Platform was considered but not deemed necessary for our purposes.

The subsequent section elucidates structure and specifics of all GMHs and PSMs. Every algorithm sticks to a sequence of four phases:

1. Picking the influence set The influence set comprises all clients granted the product for free and is picked randomly. Given a constant value q between 0 and 1, every node has the probability q for being selected as a member the influence set.
2. Calculating the myopic prices All clients are iterated in a random order and their respective myopic prices calculated. See (1) for more information.
3. Mapping clients to groups After the clients' myopic prices are determined, groups are formed according to the criterion specified by the mapping heuristic. Regardless of the criterion chosen, the underlying calculation scheme remains fixed (see (4) for details). Depending on the applied PSM a group price is selected (Section 3.1 for more information).
4. Iterative offering As already discussed in 2.1, *(k, l)-PP* strategies encompass k rounds and l rounds. The probability ω_i to submit an offer to a random client in round i is
$$\omega_i = \frac{(1-q)}{k-1} \tag{3}$$
and therefore constant over all rounds. According to our model the client compares his valuation of the product with the proposed product price if

[5] Which, in this case, alwayse equals 1.

an offer is presented. If the former equals or exceeds the latter, the offer is accepted; if not, declined. As an implication of the fact that the number of clients in posession of the product rises with the number of rounds passed, the client valuation function and therefore the probability to buy is expected to increase on average.

While the first two phases are identical with all heuristics, the two latter differ. All heuristics shown here use the same technique, presented in [4], for calculating group thresholds: Use the highest myopic price - or another criterion - as a reference point, divide it by a certain constant factor c, apply the result as new reference point, repeat. Thereby a decreasing exponential distribution for group limits is achieved. Given j is the current index, G_j the set of all clients in group j, p_i the myopic price of client i, c the specific constant divisor and highest myopic price over all clients, this can be expressed as (see [4]):

$$G_j = \{i \mid \frac{\hat{p_1}}{c^{j-1}} \geq p_i \geq \frac{\hat{p_1}}{c^j}\} \tag{4}$$

3.1 Group Mapping Heuristics

CDHS Gathered from [4], incorporated without any changes and therefore named for its authors, CDHS reflects the cited paper's proposal for posted-price strategies with multiple groups and rounds: Clients are sorted and grouped by their myopic prices according to (4).

Chaos A as a random approach, Chaos[6] first determines group sizes between one and the possible maximum, then picks and distributes clients purely by chance inconsiderate of any qualities the may exhibit. It's purpose is to serve as a benchmark, a opportunity for comparison. It is hardly surprising that it does not score very well compared to other heuristics. As a consequence of the different group limit calculation mechanism and contrary to the three other heuristics group sizes decrease in correlation with higher group numbers (see Figure 2). Besides, group limits may very well do not match up as they are adopted from arbitrarily picked clients (i.e., it is not certain that the group's upper limit is higher than it's lower limit).

Degrees Replaces a client's myopic price with it's degree as the pivotal criterion. Therefore vertices are sorted by and mapped according to their degree (group limits are calculated via the vertices' degrees), whereas only incoming edges are factored in. If in ascertaining the group price the minimum is sought, the myopic price of the client with the smallest degree - which is not necessarily the one with the lowest myopic price - is established as group price.

Neighbourhood The crucial property that hallmarks this heuristic is that in

[6] Named that way because *Random* is already assigned to a PSM.

addition to the MP it also considers the weight of edges connected to neighbours who are *not* part of the influence set. As already stated, clients have to have acquired the product already in an earlier round to influence events in the current round though. Otherwise they are not of any significance to the valuation function of the examined client, regardless of whether they are adjacent to it or not. A more detailed explanation follows.

To allow for this complication, we introduce ψ_r to reflect the probability of one given client receiving an offer *earlier* than another client and is calculated as a function of k: Assume $r = k-1$, as the influence step doesn't bear any relevance to this effect. In r rounds, there are r^2 possible combinations of rounds for two clients if they are permitted to enter the same round. Since there are r cases of two clients doing so and we attempt to identify the probability of a client getting an offer in a *prior* round, this probability amounts to:

$$\psi_r = \frac{\frac{r^2-r}{2}}{r^2} = \frac{r-1}{2 \cdot r} = \frac{1-\frac{1}{r}}{2} \tag{5}$$

Obviously, ψ approaches $\frac{1}{2}$ rather rapidly as r approaches infinity:

$$\lim_{r \to \infty} \psi_r = \frac{1}{2} \tag{6}$$

Thus, in analogy to the calculation of a myopic price m as defined in (1) (like with the myopic price, the weight sum is divided by two) and assuming J represents the set of edges to nodes not in the influence set, the extended myopic price $\hat{m}$ can be expressed as

$$\hat{m} = m + \frac{\sum_{j \in J} w_j}{2} \cdot \psi_r \tag{7}$$

Consequently, with $k = 2$ this algorithm is identical to CDHS as the ψ_r in the second term yields zero and therefore $\hat{m} = m$ applies.

3.2 Price Selection Mechanisms

As aforementioned heuristics follow a *(k, l)-PP* strategy, each group has to have exactly one price that is to be offered to all members of this groups. A PSM defines how these prices are selected among the participating clients. While [4] doesn't consider this detail, the model used during the research for this paper provides the opportunity to choose from the pool of individual client offers the following settings:

- Minimum Adopts the lowest myopic price in the group as group offer.
- First Quartile Picks the 0.25 quantile in the set of all myopic prices in the group.
- Median Same as above, but with the median.
- Third Quartile Same as above, but with the 0.75 quantile.

- Maximum Analogical to the first option, the highest myopic price is available too.
- Random Chooses one of the clients in the group by chance and uses it's myopic price as group price.

Presuming the clients are sorted descendingly by whatever characteristic specified (e.g. the myopic price), the mechanism works straightforward: The group price is determined by calculating the index fitting the picked PSM. Consider the obvious example of a group consisting of 100 clients - the group price will be derived (1) with the *minimum* option: from the lowest-ranking client, (2) with the *first quartile* option: from the 25th-lowest-ranking client, etc. If there is no exact index match - e. g., if there are ten members in the group and the first quartile is desired - the next lowest index is chosen. The *random* option will arbitrarily pick one of the vertices and adopt it's individual price as group price.

4 Evaluation

All simulations reviewed in this section were conducted with both the DBLP and Advogato datasets and averaged over ten runs for the former and one hundred runs for the latter. Only the total amount of income was evaluated as a measure quality, other data points - e.g. the success rate in offering to clients or the average amount spent - were neglected.[7] Several research questions were to be answered:

- What is the highest total amount of myopic prices when varying the probability to enter the influence set?
- How does the myopic prices distribution in groups look like?
- How are myopic prices distributed among groups?
- Which combinations of group mapping heuristic and price setting mechanism are the most lucrative?
- How do the number of groups and rounds relate to a marketing strategy's success?

These questions are assessed in detail below. The software used to run these simulations was written by the author of this paper and is available online[8]; flowcharts describing the workflow of the applied algorithm can be found in appendix A. Raw data backing up the presented charts and logfiles for every run are accessible at the same source. The runtime for one simulation ranges roughly between fifteen and twenty seconds with DBLP and a half and one second with Advogato, depending mainly on the used heuristic. No further analysis regarding runtime performance and memory (and potential disk space) consumption or the complexity of the applied algorithms was undertaken.

[7] Additional information can be retrieved via the produced logfiles, if necessary.
[8] https://code.google.com/p/mheson-sn/

4.1 Distribution of Myopic Prices

As increasing the average myopic price raises the expected revenues and therefore is a crucial goal to pursue during the first stage of a influence-and-exploit marketing strategy, it it is important to clarify whether the constant q determining the probability of clients joining the influence set is optimal. The obvious way to do so is by measuring and comparing the sum of the myopic prices of all clients with varying values for q.

As to the choice of the optimal value for q, although the probabilistic submodular[9] instead of the uniform additive model is examined, [4] postulate the following theorem:

> **Theorem 2.** *Consider the SM-model with $UMP[\gamma, 1]$ with $\gamma \geq 0$. For $q = 1 - \frac{k-1}{2(k-1)(1-\gamma)+k\gamma}$ it follows that $R(k\text{-}\overline{PD}(q)) \geq \frac{k-1}{2(k-1)(2-\gamma)+2\gamma}\hat{R}$ and thus $\hat{R}_{k\text{-}PD} \geq \frac{k-1}{2(k-1)(2-\gamma)+2\gamma}\hat{R}$.*

Here, $k\text{-}\overline{PD}(q)$ represents an IE strategy over k rounds with the probability q for a client being selected as a member of the influence set. $\hat{R}_{k-PD}$ is the revenue of the optimal IE strategy with price discrimination over k rounds. γ is the probability of an offer in the amount of the myopic price being accepted by a client - more precisely, γ is the lower threshold of said probability while Γ is the upper threshold. In the case of the UAM, which we use in our model, $\gamma = \Gamma = 0.5$ applies. Therefore, under the assumption of the UAM as the valuation function, the result of the cited theorem can be reformulated as:

$$\hat{R}_{k\text{-}PD} \geq \frac{k-1}{3 \cdot k - 2}\hat{R} \tag{8}$$

Hence, based on the derivations in Theorem 2 the conclusion can be drawn that there is always a strategy for $k = 2$, namely with $q = 0.5$, achieving an approximation of the optimal revenue by a factor of 0.25 or better, independent of γ: $\hat{R}_{k\text{-}PD} \geq \frac{1}{4}\hat{R}$. [4] stress that this strategy makes only use of the network structure when the myopic prices are computed.

Later on, they introduce the idea of a group segmentation into their framework. They combine the mentioned strategy for a $\frac{1}{4}$-approximation of $\hat{R}$, $2\text{-}\overline{PD}(0.5)$, with another one of their theorems (namely, *Theorem 8*). By assigning the value of $\Theta(\log n)$ to l, they deduce a constant factor approximation for $\hat{R}$:

> **Corollary 3.** *Let $c > 1$ be a constant. Assuming valuation functions from the SM-model with $UMP[\gamma, \Gamma]$ with $\gamma > 0$ then*
>
> $$R\left((2, \log_c(n))\text{-}\overline{PP}^*_G(c, 1/2)\right) \geq \frac{\gamma}{4\Gamma \cdot c} \cdot \hat{R}.$$

We again adapt their findings for our purposes and set $\gamma = \Gamma = 0.5$ as we use the UAM, so that

$$R\left((2, \log_c(n))\text{-}\overline{PP}^*_G(c, 1/2)\right) \geq \frac{1}{4 \cdot c} \cdot \hat{R} \tag{9}$$

[9] A generalization of the UAM. See section 2 of [4] for more information.

can be assumed.

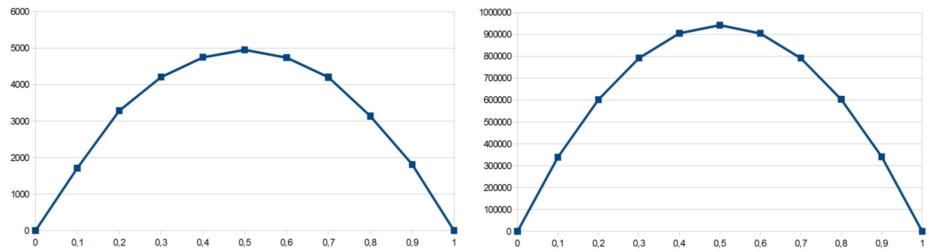

Fig. 1. Total myopic price sum under varying influence set probabilities. Left: Advogato, number of runs $i = 100$; right: DBLP, $i = 10$.

Therefore we conducted simulations with our model to verify if the predicted settings actually yield the optimum. The results, presented in Figure 1, confirm that the best hypothetical approximation for q indeed equals the empirical optimum for the two sampled networks: The myopic price sum peaks at $q = 0.5$.

As already addressed, it may also be of interest how the myopic price sum is distributed among the participating groups. To answer this question in depth, at first group sizes with different heuristics are presented in Figure 2. With *Chaos* being the obvious and expected exception, all heuristics display roughly the same pattern. As evident in Figure 3, the ratio of the number of group members to group myopic price sums show a great degree of similarity, although they differ notably in the bottom groups' populations. Notice that *Chaos* has remarkably even quotients in all groups.

Finally, in Figure 4, the myopic price group sums indicate that while the groups more extreme in their pricing both at the top and the bottom may not be disregarded entirely, the groups in between have a significantly higher share of the pool and thus are significantly more relevant: The "middle-class" contributes, even though paying considerably less per unit, far more than the few clients at the top. Therefore they presumably pose a more profitable target for optimizations.

4.2 Combinations of Group Mapping Heuristic and Price Setting Mechanisms

After an analysis of both the optimal value of q and the distribution of myopic price capital, it would stand to reason that the next question to answer is which

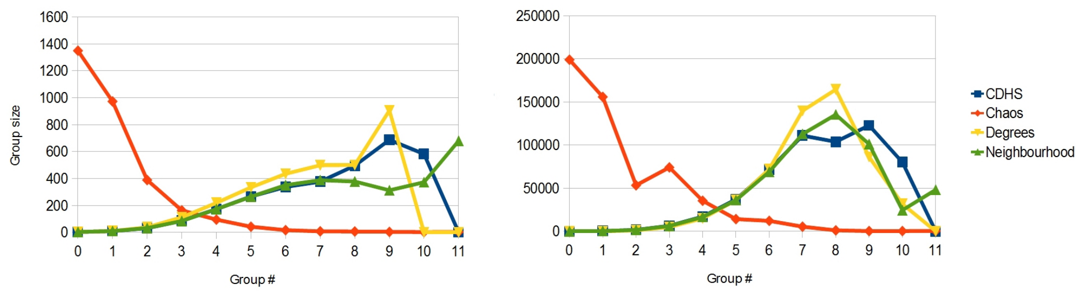

Fig. 2. Comparison of group sizes with different GMHs. Left: Advogato, $i = 100$; right: DBLP, $i = 10$.

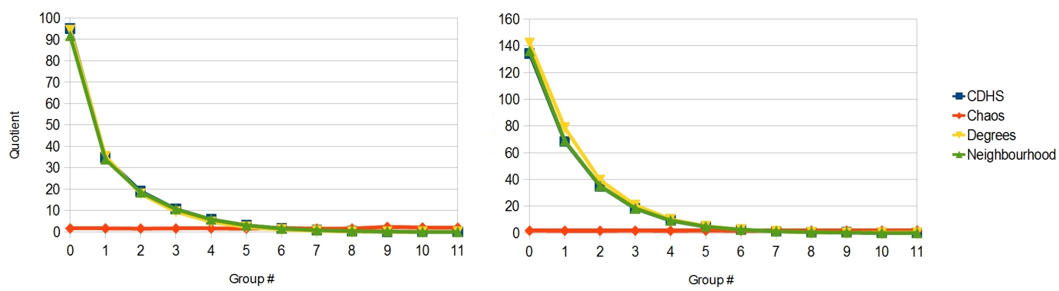

Fig. 3. Comparison of myopic price sum / group size quotients with different GMHs. Left: Advogato, $i = 100$, right: DBLP, $i = 10$.

group mapping heuristic yields the highest returns and how different price setting mechanisms infuence behaviour and revenues of these heuristics. To this end, every combination of group mapping heuristic and price setting mechanism - all in all 24 variations - was tested. Leaning on the terminology of [4], the particular specification for these tests can referred to as $(2,8)$-PP strategy, i.e. strategies with one influence and one exploit step and eight groups.

Regarding the importance of the latter, the simulation output displayed in Figures 5 and 6 demonstrates that albeit there is little but measurable difference between the first three quartiles and the randomized option, submitting offers to the amount of the group maximum has a strongly negative effect. As expected a further examination of the log files reveals that even though the clients buying paid more in average, the losses caused by the decreasing number of buyers due to the risen price outweigh these gains.

At first glance, the probably most striking feature in the chart are the poor and very volatile results of the randomized heuristic. Both the ranking as a dis-

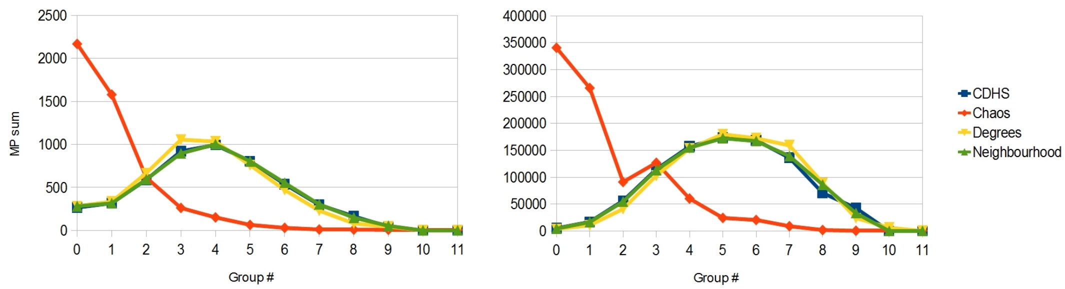

Fig. 4. Comparison of myopic price sums with different GMHs. Left: Advogato, $i = 100$, right: DBLP, $i = 10$.

tant fourth of four and the great degree of fluctuation are to be expected with this kind of strategy though. The other heuristics perform approximately two times as good than *Chaos*.

Degrees ranks third, closely behind the first two. It is notable that while it scores almost as good as them, it acts substantially more erratic. Otherwise, it's behaviour related to the price setting very closely resembles that of the first two: The averaged income peaks at the median and decrease markedly at the maximum, while the randomized setting generates neither the highest nor the lowest numbers.

As already discussed, on condition that $k = 2$ *Neighbourhood* is equivalent to *CDHS*, hence no sensible statements in the matter of the former's achievements can be made at this point. Naturally, in consequence of the equality one would anticipate the results of both to be virtually identical. Based on the measurements it seems indeed quite plausible to attribute the occuring marginal dissimilarities to statistical noise instead of any traits inherent to the algorithms in use.

Summed up: Out of all variants, *CDHS* (and *Neighbourhood*) in combination with the median as group price tend to do best. However, using the right heuristic affects the performance considerably more than the choice of the price setting mechanism. Ergo, since we want to investigate the most promising approaches in greater detail and in consideration of various values for k and l, the group price settings seem negligible. Therefore the two mentioned heuristics linked with the *minimum* price setting are subjected to a more thorough testing.

4.3 Selected Alternatives in the Context of (k, l)-PP strategies

Here the two selected constellations *CDHS/MIN* and *Neighbourhood/MIN* are studied more exhaustively in the context of $(k,l) - \overline{PP}^*_G(2, 0.5)$ strategies, i.e. strategies with the optimal probability for entries into the influence set $q = 0.5$, the dividend for calculating group limits $c = 2$ and variable values for the number of groups and rounds.

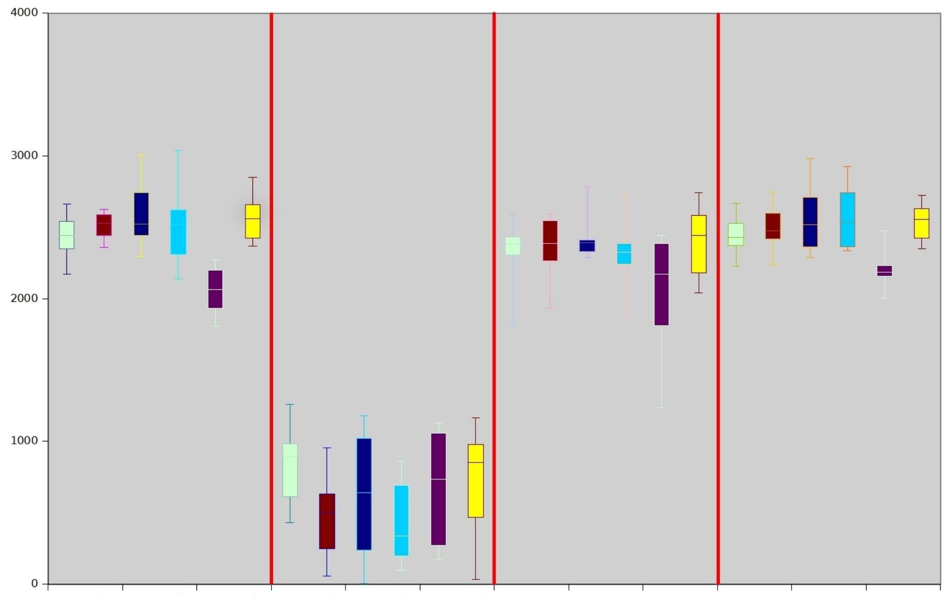

Fig. 5. Performance as measured by total revenue of all possible combinations of GMH and PSM with the Advogato graph ($i = 100$). Red lines separate GMHs - from left to right: CDHS, Chaos, Degrees, Neighbourhood. Boxes show different PSMs, from left to right: Minimum, first quartile, median, third quartile, maximum, random.

Given two sets K and L containing the numbers of rounds and groups applied and be n the total client count including the influence set, these sets can be defined as

$$k \in K(2, 3, 5, 10, 20, 50) \tag{10}$$

$$l \in L(1, 2, 5, 10, log_2(\frac{n}{2})) \tag{11}$$

Each of the thirty combinations for each heuristic was run a hundred times on the Advogato and ten times on the DBLP set.[10] The numbers produced are shown in Figure 7 for *CDHS* respectively Figure 8 for *Neighbourhood*.

Two notable patterns emerge when analyzing the presented charts. On the one hand, the number of groups is far more influential than the number of rounds. The revenue earned rises in correlation with the number of distinct groups up to certain limit - which seems to be about 10 for both datasets researched - and then remains stagnant.

[10] This choice reflects that DBLP is almost two-hundredfold the size of the Advogato graph.

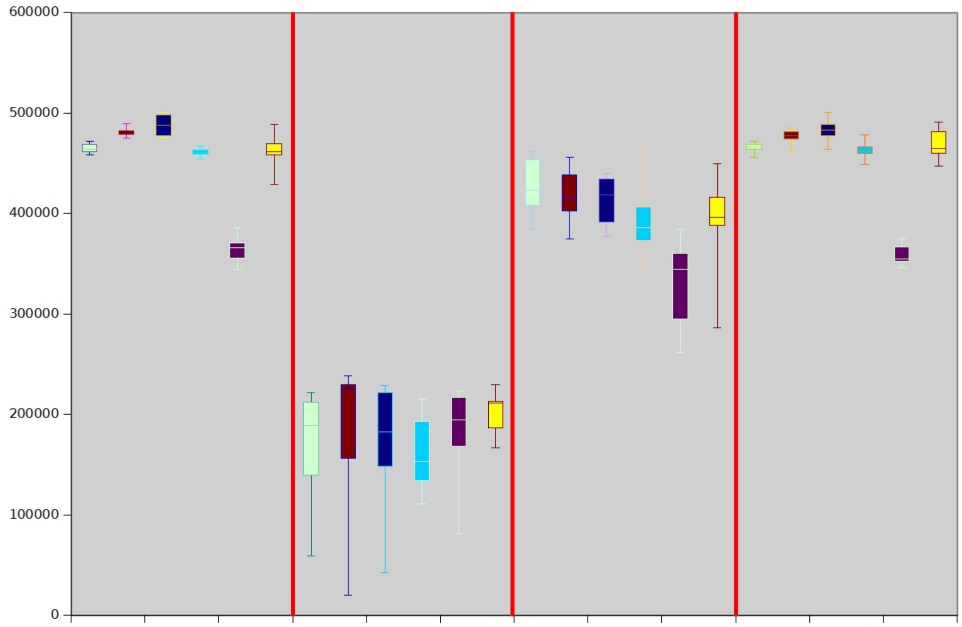

Fig. 6. Performance as measured by total revenue of all possible combinations of GMH and PSM with the DBLP graph ($i = 10$). Red lines separate GMHs - from left to right: CDHS, Chaos, Degrees, Neighbourhood. Boxes show different PSMs, from left to right: Minimum, first quartile, median, third quartile, maximum, random.

On the other hand, while running the simulation with *CDHS* generates numbers steadily rising with the count of both groups and rounds, *Neighbourhood* unexpectedly isn't quite as easy interpretable. This arises from three interconnected features - first, the decline of revenue with $k \geq 3$. Second, the sharp rise in revenue from $(2, l \geq 10)$ to $(3, l \geq 10)$. Third, whereas $k = 2$ behaves exactly as expected - i. e. identical to *CDHS* - $(k \geq 3, l)$ follows two different trends.

As already explained, *Neighbourhood* is equivalent to *CDHS* under the condition that $k = 2$. With $k > 2$, the additional mechanism accounting for the influence of neighbours not in the influence set kicks in and increases the group prices (since $\hat{m}$ doesn't equal m any longer). From this increase in prices we expected additional revenue - or, at worst, a stagnation - although in some bad cases an agent might refuse the offer due to the risen price. Why the total revenue diminishes instead of rising persists as an open question and is discussed more exhaustively in section 6.1.

Finally, we address the two divergent trends that surface for $k \geq 3$. Apparently, these are ascribable to different values for l. With $(k, l \leq 5)$, revenue declines progressively for greater values of k; in contrast $(3, l \geq 10)$ produces a peak, whereas the values for k following still deliver reduced revenue although on

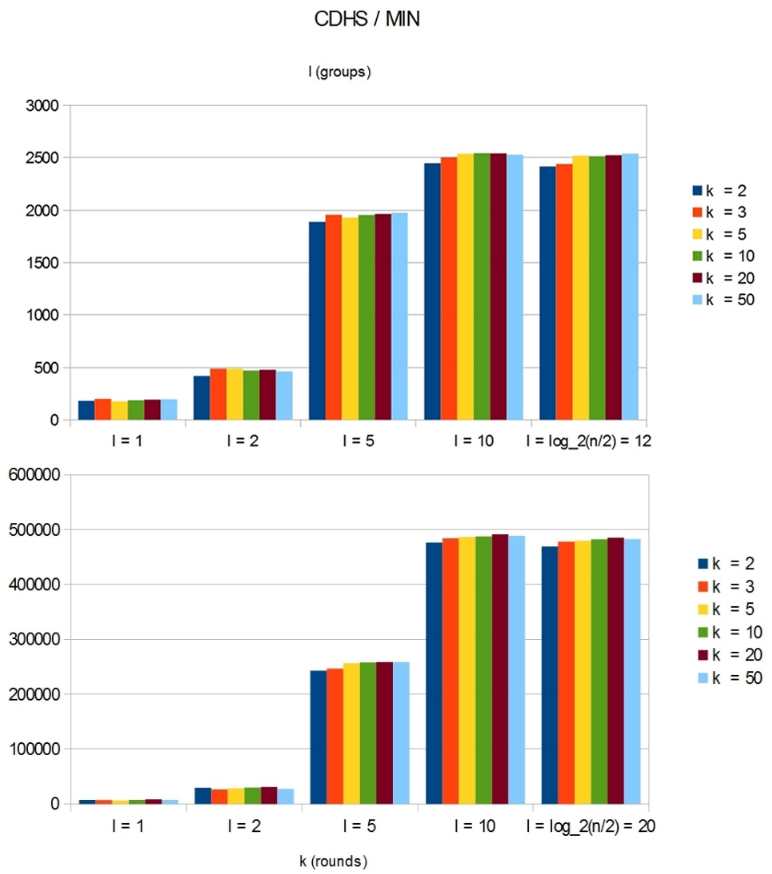

Fig. 7. Development of revenues with changing specifications of k and l, using CDHS / MIN. Upper chart derived from Advogato ($i = 100$), lower chart from DBLP ($i = 10$) graph.

a higher level than before (as $(k > 3, l \geq 10)$ may yield, other than $(k > 3, l \leq 5)$, better returns than $(2, l)$).

Subsumed *CDHS* delivers a better performance than *Neighbourhood* in most cases due to the latter's still unexplained drop in income in situations where $(k \leq 5, l)$ applies. However, there are configurations - to be more specific: $(3, l \geq 10)$ and, to a somewhat lesser extent, also $(5, l \geq 10)$ - with which *Neighbourhood* is more effective. Moreover, while *CDHS* is usually the better choice, the without ambiguity highest revenues of all combinations are produced by the former of the two mentioned *Neighbourhood* specifications:

$$(3, 10) - \overline{PP}^{*}_{G}(2, 0.5) \tag{12}$$

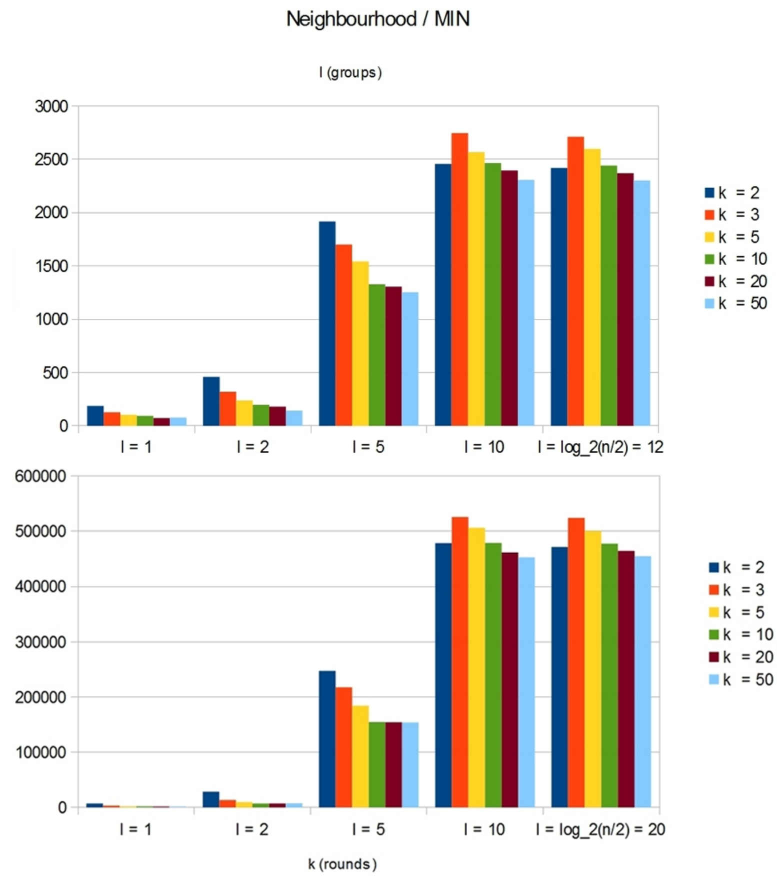

Fig. 8. Development of revenues with changing specifications of k and l, using Neighbourhood / MIN. Upper chart derived from Advogato ($i = 100$), lower chart from DBLP ($i = 10$) graph.

5 Related Work

We rely heavily on the antecedent work of Hartline et al. [8] and, even more so, Cigler et al. [4]. The former adopts the idea of social contagion - the propagation of influence of ideas or other memes in the broadest sense - explored by Kleinberg at al. in [10] with the goal of maximizing revenue instead of influence over the application of influence-and-exploit strategies. Their model offers the product to one client at once at individual prices. [4] suggests a limitation of full price-discrimination through the introduction of several groups and exactly one posted price for each of them. Furthermore, the number of grounds is confined.

The original model put forward by [8] has been improved by Fotakis et al. [7], who enhanced approximation algorithms for the uniform additive model and additionally gave discount factors. Babaei et al. [2] have undertaken an empiric evaluation of marketing strategies omitting the influence step, discounting the product for the most influential clients instead. Posted-price strategies without group assignments in a concave graph model are studied by Mirrokni et al. [15]. They provide IE-strategies that use submodular function maximization for influence set selection.

Settings with price discrimination as well as posted prices are the subject of Ehsani et al. [6], which presumes full information about the clients' valuation functions and production costs per unit. Ultimately, another variation is to be found in Akhlaghpour et al. [1]: The product to sell is offered to all of the randomly arriving clients during a limited number of days. If the clients' valuation of the product exceeds its price on a given day, the offer is accepted.

Finally, there is also research on the the identification of influential nodes, which may improve results via better structured influence sets (as already discussed, the members of the influence set are picked randomly in our model), in Domingos et al. [5] and Kempe et al. [9]. Sääskilahti et al. [17] and Cabral et al. [3] are more economically than algorithmically inclined and concerned with the effects of network topology in selling networks goods respectively multi-round pricing games. Oliver et al. [16] hint at the possible negative effects of full price-discrimination on customer satisfaction.

6 Conclusion

In this paper, various specifications of $(k, l) - \overline{PP}^*_G(2, q)$ influence-and-exploit strategies with different heuristics for mapping clients to groups and the selection of posted group prices are evaluated and compared. The valuation of choices - whether or not to accept an offer - is based on the uniform additive model and takes positive network externalities into account. The influence set is picked randomly with the same probability for all clients independently from each other. The distribution of myopic prices among groups is investigated.

Conducted simulations implementing the mentioned assumptions show that under particular circumstances the incorporation of the influence wielded by adjacent clients is benefical, as the revenue generated that way exceeds the results produced when applying the original approach. We also present the finding that the lowest myopic price of all clients in a group is not necessarily the best choice in search of a posted price for this group. Finally, we confirm that in $(k = 2, l)$ settings the theoretical best approximation of $q = 0.5$ for the probability of any node to enter the influence set matches the factual optimum for the real-world social network datasets analyzed.

6.1 Open Questions

Two open questions regarding the researched topics remain not resolved. The first, already discussed in 4.3, is: Why does the income generated by neighbourhood declines with $(k \geq 2, l \leq 5)$ respectively $(k \geq 3, l \geq 10)$, but rises between $(2, l \geq 10)$ and $(3, l \geq 10)$? Whereas the results with $(2, l)$ meet the expectations

- they are virtually identical to those of CDHS with the same configuration - we anticipated numbers to increase continuously or, at worst, to plateu with $(k > 2, l)$.

The second is related to the distribution of myopic price sums among groups - the groups at both the bottom and the top, e.g. the groups where the average myopic price is the highest and the lowest, contribute very little to the overall total of myopic prices. Does this unequal distribution affect performance negatively? If so, adapted mechanisms for the calculation of group limits (see (4)) could pose a topic of interest for further research.

References

[1] Akhlaghpour, H., Ghodsi, M., Haghpanah, N., Mirrokni, V.S., Mahini, H., Nikzad, A.: Optimal iterative pricing over social networks (extended abstract). In: Saberi, A. (ed.) Internet and Network Economics, pp. 415–423. No. 6484 in Lecture Notes in Computer Science, Springer Berlin Heidelberg (Jan 2010), `http://link.springer.com/chapter/10.1007/978-3-642-17572-5_34`

[2] Babaei, M., Mirzasoleiman, B., Jalili, M., Safari, M.A.: Revenue maximization in social networks through discounting. Social Network Analysis and Mining 3(4), 1249–1262 (Dec 2013), `http://link.springer.com/article/10.1007/s13278-012-0085-5`

[3] Cabral, L.M.B., Salant, D.J., Woroch, G.A.: Monopoly pricing with network externalities. International Journal of Industrial Organization 17(2), 199–214 (Feb 1999), `http://www.sciencedirect.com/science/article/pii/S0167718797000283`

[4] Cigler, L., Dvorak, W., Henziger, M., Starnberger, M.: Posted price strategies to exploit positive network externalities. Working Paper (2013)

[5] Domingos, P., Richardson, M.: Mining the network value of customers. In: Proceedings of the Seventh ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. pp. 57–66. KDD '01, ACM, New York, NY, USA (2001), `http://doi.acm.org/10.1145/502512.502525`

[6] Ehsani, S., Ghodsi, M., Khajenezhad, A., Mahini, H., Nikzad, A.: Optimal online pricing with network externalities. Information Processing Letters 112(4), 118–123 (Feb 2012), `http://www.sciencedirect.com/science/article/pii/S0020019011003061`

[7] Fotakis, D., Siminelakis, P.: On the efficiency of influence-and-exploit strategies for revenue maximization under positive externalities. In: Goldberg, P.W. (ed.) Internet and Network Economics, pp. 270–283. No. 7695 in Lecture Notes in Computer Science, Springer Berlin Heidelberg (Jan 2012), `http://link.springer.com/chapter/10.1007/978-3-642-35311-6_20`

[8] Hartline, J., Mirrokni, V., Sundararajan, M.: Optimal marketing strategies over social networks. In: Proceedings of the 17th International Conference on World Wide Web. pp. 189–198. WWW '08, ACM, New York, NY, USA (2008), `http://doi.acm.org/10.1145/1367497.1367524`

[9] Kempe, D., Kleinberg, J., Tardos, v.: Maximizing the spread of influence through a social network. In: Proceedings of the Ninth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. pp. 137–146. KDD '03, ACM, New York, NY, USA (2003), `http://doi.acm.org/10.1145/956750.956769`

[10] Kleinberg, J.: Cascading behavior in networks: Algorithmic and economic issues. Algorithmic game theory 24, 613–632 (2007), `http://www.cs.cmu.edu/afs/.cs.cmu.edu/Web/People/sandholm/cs15-892F11/algorithmic-game-theory.pdf#page=634`

[11] Kunegis, J.: Advogato network dataset - KONECT (Jul 2013), `http://konect.uni-koblenz.de/networks/advogato`

[12] Kunegis, J.: DBLP network dataset - KONECT (Jul 2013), `http://konect.uni-koblenz.de/networks/dblp_coauthor`

[13] Leskovec, J.: SNAP: network datasets: Epinions social network (Jul 2013), `http://snap.stanford.edu/data/soc-Epinions1.html`

[14] Leskovec, J.: SNAP: network datasets: Slashdot social network (Jul 2013), `http://snap.stanford.edu/data/soc-Slashdot0902.html`

[15] Mirrokni, V.S., Roch, S., Sundararajan, M.: On fixed-price marketing for goods with positive network externalities. In: Goldberg, P.W. (ed.) Internet and Network Economics, pp. 532–538. No. 7695 in Lecture Notes in Computer Science, Springer Berlin Heidelberg (Jan 2012), `http://link.springer.com/chapter/10.1007/978-3-642-35311-6_43`

[16] Oliver, R.L., Shor, M.: Digital redemption of coupons: satisfying and dissatisfying effects of promotion codes. Journal of Product & Brand Management 12(2), 121–134 (2003), `http://www.emeraldinsight.com/10.1108/10610420310469805`

[17] Sääskilahti, P.: Monopoly pricing of social goods (2007), `http://mpra.ub.uni-muenchen.de/3526/`

A Simulation Workflow

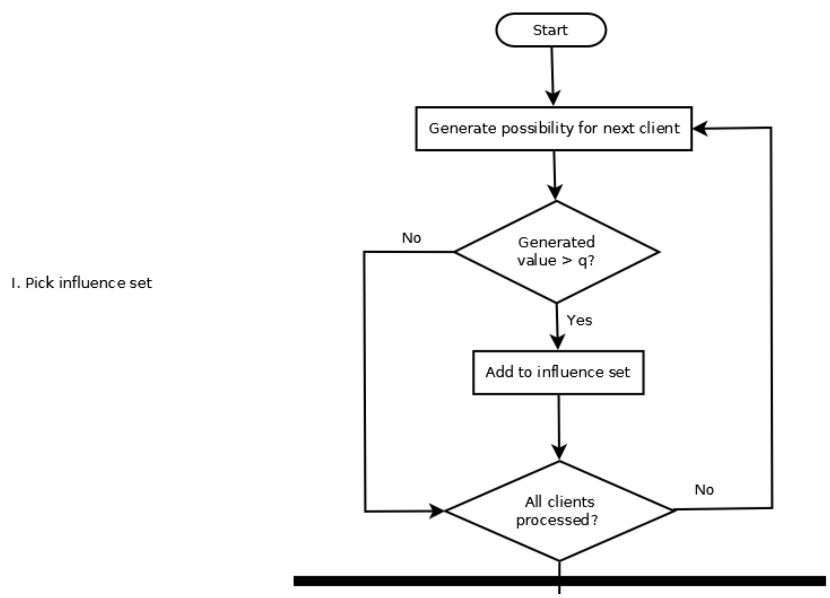

Fig. 9. Flowchart depicting the first phase (picking the influence set, see item 1 in Section 3) of internal workflow of the simulation used for our model.

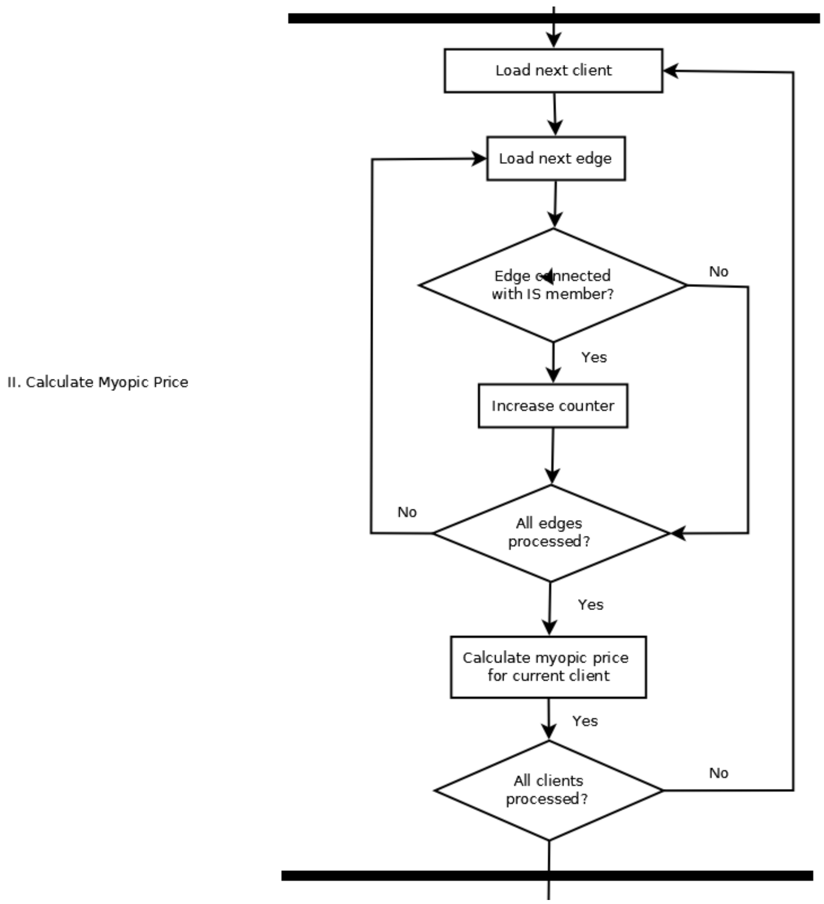

Fig. 10. Flowchart depicting the second phase (calculation of the myopic prices, see item 2 in Section 3) of internal workflow of the simulation used for our model.

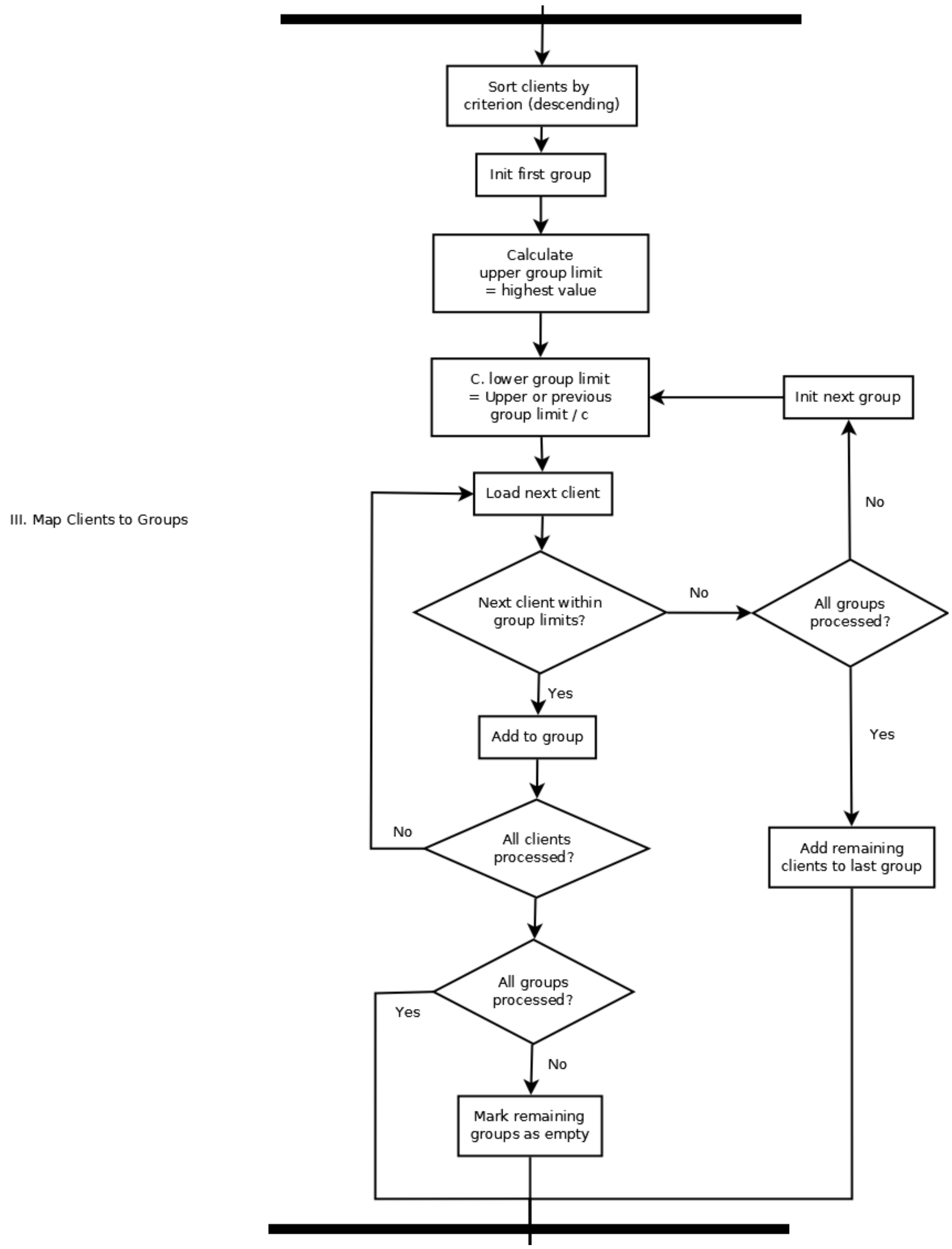

Fig. 11. Flowchart depicting the third phase (mapping clients to groups, see item 3 in Section 3) of internal workflow of the simulation used for our model.

IV. Iterative Offering

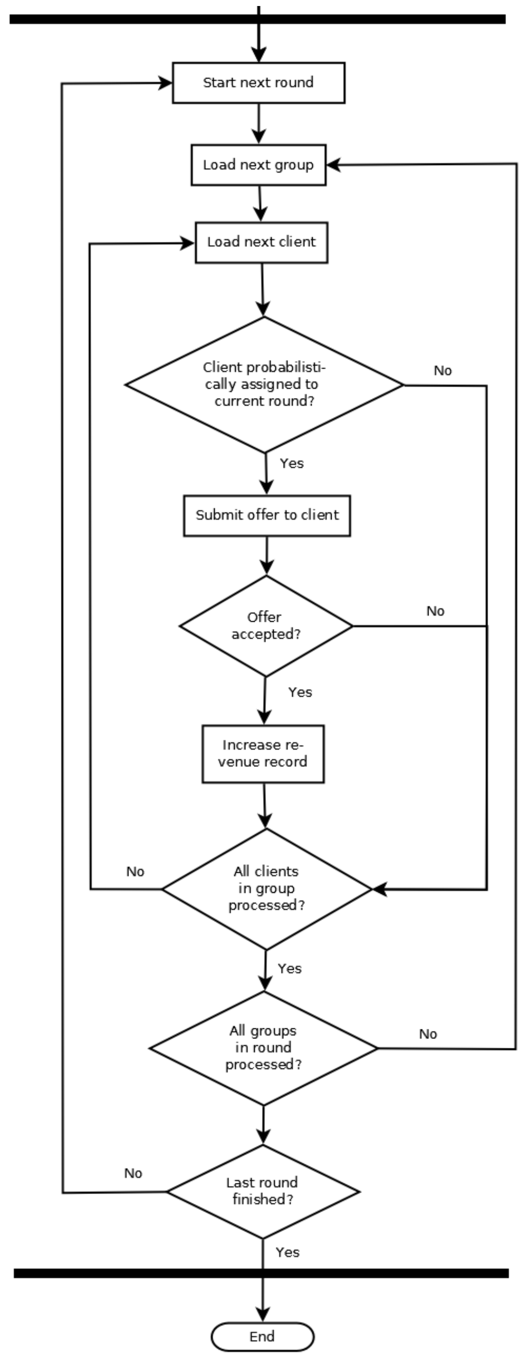

Fig. 12. Flowchart depicting the fourth phase (iterative offering, see item 4 in Section 3) of internal workflow of the simulation used for our model.