Indirect Influence Manipulation with Partial Observability

James Archbold[1] and Nathan Griffiths[2]

[1] University of Warwick, UK, james.archbold@warwick.ac.uk
[2] University of Warwick, UK, Nathan.Griffiths@warwick.ac.uk

Abstract. The propagation of concepts through a population of agents can be modelled as a cascade of influence spread from an initial set of individuals. In real-world environments there may be many concepts spreading and interacting, and we may not be able to directly control the target concept we wish to manipulate, requiring indirect manipulation through a secondary controllable concept. Previous work on influence spread typically assumes that we have full knowledge of a network, which may not be the case. In this paper, we investigate indirect influence manipulation when we can only observe a sample of the full network. We propose a heuristic, known as Target Degree, for selecting seed nodes for a secondary controllable concept that uses the limited information available in a partially observable environment to indirectly manipulate the target concept. Target degree is shown to be effective in synthetic small-world networks and in real-world networks when the controllable concept is introduced after the target concept.

Keywords: influence spread · social networks · information diffusion

1 Introduction

In many environments, strategies, concepts or infections may spread within a population. The nature of propagation is determined by the interactions between individuals. Populations of autonomous entities are complex systems, meaning that the net effects of propagation are hard to predict or influence, despite being due to individual behaviour. Such propagation is a form of influence spread, which can be modelled as a cascade from a set of initial individual agents [8]. The propagation of concepts between agents can affect individual behaviour, which in turn can affect the behaviour of the overall system. Influence spread techniques have applications in epidemiology, marketing and behavioural science, and can involve an agent-based simulation of a real-world problem. Understanding concept propagation aids in the identification of influential individuals, who can help or hinder a concept's spread.

Several models have been developed to characterise influence spread [4], along with techniques to maximise spread [3]. A population can be represented as a network, through which concepts spread. Nodes represent individuals and edges represent the influence that exists between pairs of individuals. Concept spread is maximised through the strategic selection of a set of nodes, known as the seed set, to begin a cascade through the network. Selecting an effective seed set has been the focus of much work. Multi-concept models typically assume that concepts *block* each other, preventing a node from activating multiple concepts simultaneously [6]. However, in the real-world an individual may have many concepts active, which may interact and affect how concepts spread.

For example, consider how political beliefs are developed. The news stories and opinion articles that a person reads, and decides are reputable, will affect the political affiliation they are likely to adopt. Furthermore, this will affect the types of articles and opinions that they share with those within their social circle. Similarly, in epidemiology, a disease may cause symptoms in an individual that encourages the spread of other diseases.

If concepts interact, we can consider how to indirectly affect the spread of a concept that cannot be directly controlled. A common example of this is inoculation and education to limit the spread of a disease, but we could also promote particular news stories to improve the spread of a particular political opinion, or lower the sales of a product through competing products. In all these cases, the target concept cannot be controlled directly and so we use another controllable concept to boost or inhibit the spread of the target concept.

Recent work has utilised concept interaction to indirectly affect the spread of a target concept that cannot be directly controlled. Liontis and Pitorua proposed the MoBoo heuristic, which evaluates the possible gain from selecting a node [10]. Archbold and Griffiths proposed the Maximum Probable Gain (MPG) heuristic to select seeds for a secondary controllable concept to indirectly affect the spread of the target concept [2].

Previous work assumes that we have knowledge of the entire network, but mapping real-world social networks is expensive and is often infeasible. Typically, when working with real-world networks, only a small sample can be observed, which may not be representative and limits the information known about the network. Thus, we wish to effectively manipulate influence within a network when we can only observe a sample of that network to inform our decisions.

In this paper, we study the problem of indirect influence manipulation when we have only partial network information. We present a heuristic to select a seed set for a controllable concept from a small sample of a full network, known as Target Degree, and compare it to random selection, degree-based selection, single discount, degree discount, MPG and MoBoo.

2 Related Work

Several influence propagation models have been proposed in the literature [8]. Two of the most widely discussed are the *Independent Cascade Model* (ICM) and the *Linear Threshold Model* (LTM). The ICM treats influence as particle diffusion, with each node given a single chance, on activation, to activate a concept on each inactive neighbour, with some probability, p [4]. The LTM emulates social peer pressure, with every active neighbour of a node being considered [8].

Each node v has a threshold θ_v, and if the sum of the weights of v's active neighbours exceeds θ_v, then v becomes active.

The problem of influence maximisation has been widely studied, resulting in many different approaches. A basic hill-climbing approach that selects the node that provides the largest incremental increase to the performance of the current seed set can be effective [5], but is typically intractable in practice. Chen *et al.* proposed the Degree Discount heuristic for the ICM, which accounts for existing activations by ranking nodes by degree and decrementing that degree when a node's neighbours are selected to be seeds [3]. Degree Discount has been shown to be both effective and tractable.

In real-world environments, there may be many concepts spreading through a population. Thus, recent work has considered multiple influence cascades, but typically assumes that cascades are blocking, preventing a node from activating multiple concepts [7]. If concepts are not blocking, then their interactions must be considered. Sanz *et al.* developed a multi-layer network model in which each concept spreads on a separate layer, but nodes can have multiple concepts active, and concepts can interact [12]. Through interaction a concept can *boost* or *inhibit* the spread of another concept. Considering these interactions has resulted in new influence maximisation strategies [1].

Liontis and Pitoura developed the MoBoo algorithm for indirectly boosting the spread of a concept [10]. Nodes are selected to have an increased probability to spread the target concept. MoBoo constructs a series of trees from nodes with the target concept active, and selects nodes based on how many trees they appear in and the number of child nodes they have in each tree. Concept interaction introduces the ability to contain rumours, by indirectly limiting the spread of a concept [9]. Existing approaches to rumour containment also often assume that concepts block, and attempt to partition the network and make traversal difficult [11]. However, this approach becomes less effective when concept, and therefore path, blocking can not be guaranteed. The MPG heuristic uses local exploration, identifying those nodes that are likely to activate the target concept [2]. These nodes also have their neighbourhood explored, to determine their expected gain. MPG selects nodes with both high activation probability and high expected gain, for the controllable concept's seed set.

There has been relatively little work that considers partial observability in the context of influence spread. Partially observable Markov decision process planners [14] and greedy, oracle-based, algorithms [13] have been used to account for uncertainty when directly maximising influence. However, these methods only consider single concepts, rather than multiple interacting concepts.

3 Concept Interaction

In this paper, we consider the indirect influence maximisation problem and the indirect influence limitation problem, which require selecting a seed set of size k for a controllable concept with the aim of affecting the spread of a target concept. In the indirect influence maximisation problem, we aim to increase the spread of the target concept and in the indirect influence limitation problem, we aim to minimise the spread of the target concept. Both problems assume that concepts interact and affect each other's spread, which we model in a similar way to Sanz *et al.'s* approach for two interacting concepts [12].

We model a set of agents as a network, where nodes represent individual agents and edges represent a connection that allows for influence to be exerted. When two agents interact, influence is sent by the *infector* to the *receiver* and the receiver will potentially activate a particular concept that is active on the infector. We denote the strength of the influence exerted by node v on node u with respect to concept c as $I_{v,u}(c)$. Any value $I_{v,u}(c) > 0$ means that v has some influence over u, represented as an edge in a network.

The relationship between two concepts is defined by the effect that one has on the other's ability to spread, represented as a numerical value. The variable $CR(c, c')$ describes the effect that c' has on the ability of c to spread when present on the infector or receiver. In this paper, we assume that c' affects concept spread in the same way, regardless of whether it is active on the infector or receiver.

We assume that $CR(c, c') \in [0, \infty)$ is a feature of the environment. If $CR(c, c') < 1$ then c' is inhibiting and decreases the chance of c to spread, and if $CR(c, c') > 1$ then c' is boosting and increases the chance of c to spread. If $CR(c, c') = 1$ then c' has no effect on the ability of c to spread. $CR(c, c')$ is used to define the contextual influence that v can exert on u with respect to concept c, $CI_{v,u}(c)$, which accounts for concept relationships as follows:

$$CI_{v,u}(c) = \begin{cases} I_{v,u}(c) & \text{if } c' \text{ is not active on either } v \text{ or } u \\ I_{v,u}(c) * CR(c, c') & \text{if } c' \text{ is active on either } v \text{ or } u \\ I_{v,u}(c) * CR(c, c')^2 & \text{if } c' \text{ is active on both } v \text{ and } u. \end{cases} \tag{1}$$

Note that the effect of c' is compounded when it is active on both the infector and receiver, in the same way that individuals can be more easily influenced by people who they perceive as similar to themselves.

To model concept propagation, we adapt the ICM to allow for multiple simultaneous cascades. Cascades proceed in rounds, with the nodes in each concept's seed set activating that concept in round 0. In each subsequent round, nodes that activated a concept in the previous round have a chance to activate that concept on each of their neighbours. This continues until there are no new activations for any concept. Activations happen simultaneously, and so the contextual influence can only be affected by concepts active before the current round. In this model, we use $I_{v,u}(c)$ as the chance of node v successfully activating c on node u.

4 Target Degree

We wish to indirectly manipulate the spread of a target concept in an environment where we are only able to observe a small sample of a network. Previous heuristics assume full knowledge of the network, and may require in-depth exploration of a node's neighbourhood in order to select a seed set for the controllable concept. However, such analysis is impractical when only a sample can be observed. Sampling a node does not guarantee that all edges of the node are observed, making degree-based selection unreliable. Thus, a node's observable, explorable, neighbourhood may not be representative of the influence it can exert in the full network, and so we require a new method of seed selection.

We propose the ***Target Degree*** (TD) heuristic, which ranks nodes by the number of neighbours with the target concept active. If a node is in the observable sample, we assume that we know whether it has the target concept active. If a node with the target concept active has unobserved edges, activating the controllable concept on that node will not only affect the spread of the target concept in the observable area, but also increase the chance of interacting with the target concept in the unobserved network. A node with the target concept active is likely to have neighbours with the target concept active or is likely to spread the target concept to its neighbours. Without full network knowledge, we focus on the immediate benefits and activate the controllable concept on nodes that either have the target concept active or have many neighbours with the target concept active.

As such, for TD, we create two ranked lists, L_t and $L_{\neg t}$, of nodes with the target concept active and nodes without the target concept active respectively. In each list, nodes are ranked based on the number of neighbours they have with the target concept active, in descending order. We then append $L_{\neg t}$ to end of L_t to create the combined list, L. For a seed set of size k, we select the first k elements of list L. TD can be efficiently calculated and does not require additional calculations after the selection of individual seed nodes, as is the case with several existing heuristics such as degree discount, MPG and MoBoo. We evaluate TD against the following heuristics.

Degree-based selection. Degree-based selection is a simple heuristic, that is cheap to compute and has been shown to be effective [8]. In degree-based selection, the k nodes with the highest degree are selected as the seed set.

Single Discount. This heuristic accounts for the fact that a node selected to be in the seed set cannot be activated by its neighbours, meaning that the neighbours of a selected seed node suffer a decrease in non-active neighbours that can potentially be activated. In single discount, the highest degree node is selected, and the degree of its neighbours is lowered by 1. This process is repeated until the full seed set is selected [3].

Degree Discount. Selecting a node as a seed lowers the expected gain of its neighbours, and increases the chance its neighbours may be activated in the first round. Degree discount therefore weights a node's degree based on the number of its neighbours previously selected to be seed nodes. Nodes are ranked by degree, and when a node is selected its neighbours have their degree set to $d_v - 2t_v - (d_v - t_v) * t_v * p$, where d_v is the original degree, t_v is the number of neighbours in the seed set and p is the probability of infection. The full derivation of this calculation can be found in [3].

MoBoo. In this heuristic, nodes are evaluated based on their expected gain if they were to be selected as a boosting node. Using the two most probable independent paths for the target concept to reach a node v, the activation prob-

Table 1. Experimental parameters.

Parameter	Values
Proportion of network sample (nodes) (SN)	0.1, 0.2, 0.3
Sampling Methods (SM)	Snowball, MHDA
Seed set size (SS)	10, 25, 50, 100, 250, 500
CR function values for the controllable concept	0, 0.2, 0.4, 0.6, 0.8, 1.2, 1.4, 1.6, 1.8, 2
Burn-in time for the controllable concept (BI)	0, 2, 5

ability $ap(v)$ is calculated as the probability that the target concept reaches v from one or both of the paths. The gain for a node, v, is then calculated as:

$$g(v) = \sum_{u \in Out(v)} \left(\frac{p'_{v,u}}{p_{v,u}} - 1 \right) \sum_{w\ descendant\ of\ u} ap(w)$$

where $Out(v)$ is the set of nodes that are the children of v in either path, $p_{v,u}$ is the probability of the concept spreading from node v to u and $p'_{v,u} = p_{v,u} + b$, with b being the improvement gained by a node being a boosting node. Each round, the node with the highest gain is chosen until we have the desired number of nodes. Full details of MoBoo can be found in [10].

MPG. This heuristic also calculates the activation probability and expected gain of each node. However, MPG limits its exploration to paths with an influence value higher than a set threshold, θ. The influence value, I_P, of a path $P = \{v_1 \rightarrow \ldots \rightarrow v_n\}$ with respect to target concept t, is calculated as the product of all $CI_{v_i,v_{i+1}}(t)$ values in the path.

The most influential path to node u from node v with respect to t is defined as $MIP(v,u) = \text{argmax}_{P \in AP_{v,u}}(I_P)$, where $AP_{v,u}$ is the set of all paths that start with v and end in u. The influence value of $MIP(v,u)$ is denoted as $I_{MIP(v,u)}$, and an influence value less than θ is treated as 0. Thus, the influence received from v by u, $IR(v,u)$ is set to $I_{MIP(v,u)}$ or to 0 if $I_{MIP(v,u)} < \theta$.

The activation probability of node u, $ap(u)$, is defined as the sum of all $IR(v,u)$ values where v is actively spreading the target concept. In the case of the ICM, this means that v was activated in the previous cascade round. The expected gain of u, $E(u)$, is the sum of all $IR(u,w)$ values where w is a node without the target concept active. The weighted expected gain, $WE(u)$, for node u is defined as $WE(u) = E(u) \times ap(u)$. The node with the highest $WE(u)$ value is selected, and $WE(v)$ is recalculated for unselected nodes, until the seed set reaches its desired size. Full details can be found in [2].

5 Experimental Approach

To simulate partial observability, we select controllable concept seeds from an observable subset of nodes in a network, sampled through either snowball sampling or through a Metropolis Hasting random walk with Delayed Acceptance

Table 2. Characteristics of the real-world networks used for evaluation.

Network	Nodes	Edges	Avg. Degree	Avg. Clustering Coefficient	Num. of Triangles	Diameter
CA-CondMat (CM)	23133	93497	4.04	0.6334	173361	14
cit-HepPh (HP)	34546	421578	12.2	0.2848	1276868	12
DBLP (DB)	317080	1049866	3.31	0.6324	2224385	21

(MHDA). Snowball sampling maintains the local structure of an area in the network, while MHDA produces a sample with characteristics, such as degree distribution, more in line with those of the full network, but at the expense of maintaining local structure. We consider various sampling proportions to represent varying degrees of observability, as listed in Table 1.

For each combination of parameters in Table 1, we perform 50 simulations for each of the heuristics described in Section 4 along with random selection to act as a baseline. We perform two tailed t-tests between heuristics to test for statistical significance. The controllable concept is introduced after a fixed number of time steps, known as the burn-in time. This is kept low, as high burn-in times result in indirect influence manipulation being ineffective [2]. The target concept is introduced at time step 0, and its seed set is randomly selected from the full network. Simulations are performed using the ICM, and each concept has a probability of spreading to a neighbour of $I_{v,u}(c) = 0.1$.

In this paper, we consider a selection of representative networks. Synthetic small-world networks, with a size of 100000 nodes and a clustering exponent of 0.75, are generated using the Kleinberg small world generator in the JUNG graph framework[3]. Synthetic scale-free networks with 100000 nodes are constructed using the Barabási-Albert generator provided in JUNG, which begins with a set of unconnected nodes, 10 in this case, and introduces a new node each evolution step. The new node gains a number of edges, 4 in this case, connected to existing nodes using preferential attachment. A number of real-world networks[4] are used, based on datasets from the Stanford SNAP project[5], as listed in Table 2.

For the small-world, scale-free, and DB real-world networks we use seed set sizes of $100, 250, 500$. Since the observable samples of the CM and HP networks often contain less than 7500 nodes, we use the seed set sizes of $10, 25, 50$ to prevent the seed set from a majority of the sample.

6 Results

To begin, we discuss the synthetic networks. Results for the synthetic small-world and scale-free networks can be seen in Table 3, for when the burn-in time is 0.

[3] http://jung.sourceforge.net/
[4] These networks are samples of full social networks, but for the purposes of this paper we treat them as the complete network.
[5] http://snap.stanford.edu/data/index.html

Table 3. Average infections for the target concept in networks with $SN = 0.2$, $SS = 250$ and $BI = 0$, with standard deviation in brackets, and the best performing heuristic in bold.

Network Type	Sampling Method	CR Value	Target Degree	MPG	MoBoo	Degree Discount	Single Discount	Degree
SW	MHDA	0	**494.14 (32.82)**	533.86 (36.17)	554.88 (37.13)	553.72 (37.01)	553.66 (37.01)	554.44 (37.49)
SW	Snow	0	**489.96 (33.12)**	525.62 (35.53)	555.8 (37.3)	553.02 (37.83)	553.08 (37.86)	553.78 (37.5)
SW	MHDA	2	**651.82 (45.96)**	635.76 (43.58)	611.2 (44.14)	562.48 (37.98)	562.48 (37.98)	562.58 (38.92)
SW	Snow	2	**657.52 (42.55)**	603.72 (38.15)	624.98 (40.45)	569.88 (39.09)	569.14 (39.31)	569.3 (39.99)
SF	MHDA	0	7565.02 (1676.26)	6852.68 (1420.88)	6887.36 (1091.64)	6827.02 (1402.93)	**6805.14 (1392.88)**	6866.82 (1510.26)
SF	Snow	0	2613.22 (754.62)	1724.64 (343.79)	10480 (2114.67)	1420.98 (238.4)	1403.96 (245.64)	**1402.9 (298.49)**
SF	MHDA	2	33673.52 (756.82)	35075.02 (852.79)	35230.6 (891.54)	35495.22 (893.7)	35528.9 (883.2)	**35588.1 (872.15)**
SF	Snow	2	37017.62 (765.79)	39252.76 (634.52)	38806.04 (659.84)	40675.74 (534.25)	**40687.26 (582.2)**	40684.02 (511.53)

In general, the sampling proportion and seed set size only impact the magnitude of the results, but not the relative performance, and so we do not discuss them further, in regards to the synthetic networks.

As can be seen in Table 3, we see that in small-world networks with a burn-in of 0, TD statistically significantly outperforms ($p < 0.01$) all other heuristics for each sampling proportion, sampling method and seed set size for both inhibiting and boosting the target concept. Comparatively, in scale-free networks we see that degree-based heuristics perform best. In general, we see that TD is the worst performing heuristic in scale-free networks, particularly when attempting to inhibit the target concept. When attempting to boost the target concept, the difference in performance between all heuristics is relatively small and, in general, not statistically significant.

As the burn-in time increases, for both small-world and scale-free networks, the heuristics begin to perform at the same level. Table 4 shows the two different patterns of performance we see as the burn-in time increases. In small-world networks, we see that only TD's performance significantly changes as the burn-in time increases to 2, but it continues to outperform the other heuristics and then, at the highest burn-in time, there is no statistically significant difference between the heuristics. In snowball sampled scale-free networks we see a similar pattern, in that all heuristics converge to similar performance as the burn-in increases, and the best performing heuristic does not change. For MHDA sampled scale-free networks, as seen in Table 4, TD's relative performance improves as the burn-in time increases. At a burn-in time of 2, TD is the best performing heuristic by a significant margin. At a burn-in time of 5, the target concept has performed the majority of its spreading, and so manipulating the concept at that point will

Table 4. Average infections for the target concept in networks with $SN = 0.2$, $SS = 250$, $CR = 0$ and $SM = MHDA$, with standard deviation in brackets, and the best performing heuristic in bold.

Network Type	Burn-in Time	Target Degree	MPG	MoBoo	Degree Discount	Single Discount	Degree
SW	2	**541.2 (34.24)**	551.92 (36.09)	551.96 (36.37)	557.0 (37.82)	557.0 (37.82)	557.38 (37.93)
SW	5	**555.94 (37.53)**	556.54 (37.63)	556.46 (37.46)	557.7 (38)	557.7 (38)	557.7 (38)
SF	2	**8390.06 (1052.69)**	11630.76 (1288.63)	11755.16 (1163.19)	11738.56 (1307.4)	11751.9 (1307.14)	11697.38 (1459.65)
SF	5	**16192.8 (1007.95)**	16786.46 (645.81)	17034.88 (760.98)	16556.06 (716.85)	16558.16 (718.6)	16554.48 (760.27)

yield minimal results, meaning that there is no significant difference between the heuristics.

Overall, when considering synthetic networks, we see that TD is the best choice for small-world networks, regardless of sampling method. We also see that TD is less affected by burn-in time than other heuristics, which allows it to perform well in MHDA sampled scale-free networks. Thus, if we can control the sampling method of a network, TD becomes a strong choice for manipulating the spread of a target concept.

Considering real-world networks, we see that the use of snowball sampling results in a similar performance to that seen in the scale-free synthetic networks. Thus, due to space limitations, we focus on the MHDA samples. Figures 1 and 2 show the difference in performance for a subset of heuristics at different burn-in times, for boosting and inhibiting the target concept respectively. We omit degree-based selection and single discount as they perform similarly to degree discount in all cases. MoBoo is omitted as it performs consistently poorly at inhibition and is comparable to MPG at boosting. Both figures are for the CM network, with the HP network exhibiting similar results.

Figure 1 shows that, as the burn-in increases, TD improves its performance and outperforms the other heuristics, as in the scale-free networks. In Figure 2, TD is the best performing heuristic when the CR value and burn-in time is low, and than similarly improves as the burn-in time increases.

This resistance to burn-in time is particularly advantageous for real-world applications, where it can be difficult to introduce a controllable concept to a network at the exact same time as the target concept.

In the DB network, Figure 3 shows that, when inhibiting the target concept from a MHDA sample, increasing the burn-in time increases TD's performance. However, increasing the sampling proportion, as in Figure 4, causes MPG to maintain superior performance at higher burn-in times. DB is the sparsest network, with the highest diameter, meaning that most nodes have few local connections and are not closely connected to the rest of the network. As such, at

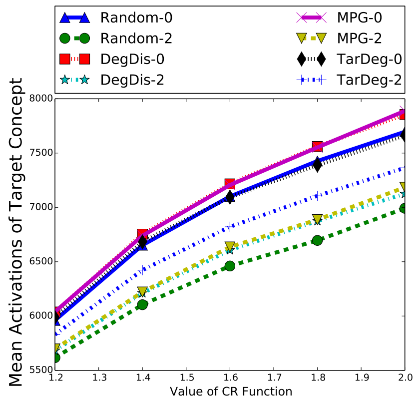

Fig. 1. Mean activations of the target concept given the heuristic and burn-in time used to select the boosting concept in the CM network, $SN = 0.3$, SM=MHDA, $SS = 50$, $CR = 2$.

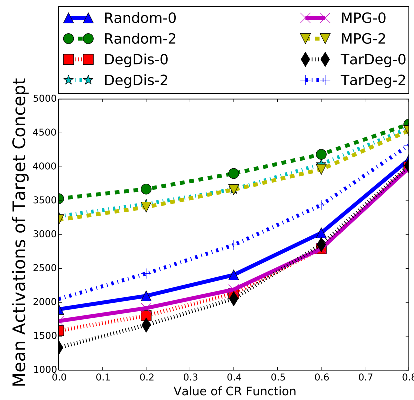

Fig. 2. Mean activations of the target concept given the heuristic and burn-in time used to select the inhibiting concept in the CM network, $SN = 0.3$, SM=MHDA, $SS = 50$, $CR = 0$.

higher burn-in times, it is less likely for unobserved nodes to affect the observed sample. Increasing the sampling proportion further removes the few unobserved connections that may exist, which in turn can improve the performance of exploration. This also occurs when boosting the target concept, although TD is never the best performing heuristic in this case.

Overall, in real-world networks, we see that increasing the burn-in time generally improves the comparative performance of TD to a point, after which every heuristic performs at a similar level. At higher burn-in times, there is a higher chance of observed nodes having activated the target concept from an unobserved node. By selecting these nodes, TD is more likely to influence the unobserved network and is more likely to interact with the target concept.

Furthermore, we see that MHDA sampling allows TD to perform better than snowball sampling. Snowball sampling is assumed to be capable of finding all edges of a node. As such, we perfectly sample a local area of the network. This means that, compared to the random walk approach of MHDA sampling, there are no unobserved edges. In an observed sample with no unobserved edges, in-depth path prediction is more reliable. When MHDA sampling is used, the number of unobserved edges increases and makes path prediction less effective. A node with the target concept active may have been activated by a neighbour in the unobserved part of the network, meaning that selecting these types of nodes increases the chance for the controllable concept to affect the target concept in the unobserved area of the network.

Finally, we see a distinction between the synthetic small-world networks and the other networks observed. It is only in the synthetic small-world networks that TD outperforms all other heuristics in every environment. The presence of

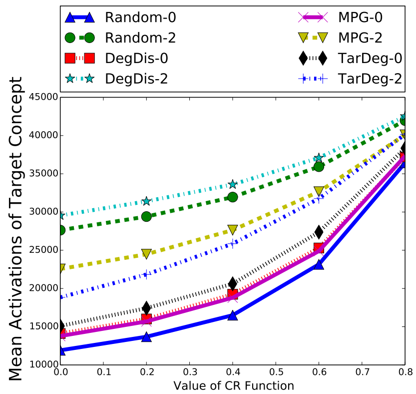

Fig. 3. Mean activations of the target concept given the heuristic and burn-in time used to select the inhibiting concept in the DB network, $SN = 0.1$, SM=MHDA, $SS = 500$, $CR = 0$.

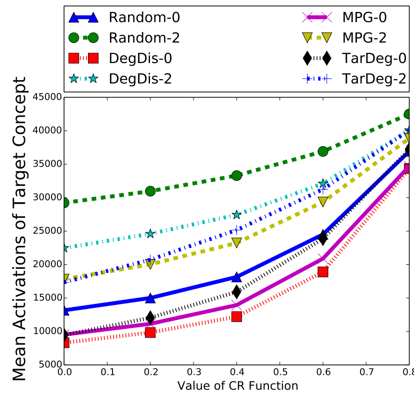

Fig. 4. Mean activations of the target concept given the heuristic and burn-in time used to select the inhibiting concept in the DB network, $SN = 0.3$, SM=MHDA, $SS = 500$, $CR = 0$.

scale-free properties in the other networks means that degree-based heuristics are favoured in most cases, particularly in snowball samples. This, combined with TDs improved performance in MHDA samples, implies that TD performs better when there is a higher number of unobserved edges.

7 Conclusions & Future Work

In this work, we discuss the problem of indirectly manipulating the spread of a concept through concept interaction, when we do not have full network knowledge. We proposed the Target Degree (TD) heuristic that utilises minimal information and does not rely on in-depth network exploration, and compared its performance to several other heuristics for indirect influence manipulation.

TD was the best heuristic in the synthetic small-world networks, and is effective in scale-free networks sampled using MHDA with a burn-in time greater than 0. Otherwise, degree-based heuristics proved superior, and we see a similar result in the real-world networks studied. In nearly all cases, TD was the best performing heuristic at a burn-in time of 2 in MHDA samples, with the exception of the DB network, implying that TD requires denser networks to be effective. In real-world applications it may be impossible to introduce the controllable concept at the same time as the target concept, especially since we assume no control over the target concept, making TD a suitable option. Overall, for both synthetic and real-world networks, if the sampling method that determines which nodes can be observed can be selected, then TD may provide the best result.

TD performed poorly in snowball sampled networks, implying that it performs better with a higher number of unobserved edges. Sampling real-world networks is unlikely to be perfect, meaning that unobserved edges are more likely, providing further evidence of TDs suitability to real-world applications.

In future work, we wish to explore this problem in other influence spread models, including the Linear Threshold model and the Susceptible-Infected-Susceptible model. Furthermore, we will consider dynamic networks, where the observed sampled may lose or gain nodes and the unobserved network is updated as the concepts cascade through the network.

References

1. Archbold, J., Griffiths, N.: Maximising influence in non-blocking cascades of interacting concepts. In: International Workshop on Multi-Agent Systems and Agent-Based Simulation. pp. 173–187. Springer (2015)
2. Archbold, J., Griffiths, N.: Limiting concept spread in environments with interacting concepts. In: Proceedings of the 16th Conference on Autonomous Agents and MultiAgent Systems. pp. 1332–1340 (2017)
3. Chen, W., Wang, Y., Yang, S.: Efficient influence maximization in social networks. In: Proceedings of the 15th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. pp. 199–208 (2009)
4. Goldenberg, J., Libai, B., Muller, E.: Using complex systems analysis to advance marketing theory development. Academy of Marketing Science Review **9**(3), 1–18 (2001)
5. Goyal, A., Lu, W., Lakshmanan, L.V.: Celf++: optimizing the greedy algorithm for influence maximization in social networks. In: Proceedings of the 20th ACM International Conference companion on World Wide Web. pp. 47–48 (2011)
6. Goyal, S., Kearns, M.: Competitive contagion in networks. In: Proceedings of the 44th annual ACM Symposium on Theory of Computing. pp. 759–774 (2012)
7. He, X., Song, G., Chen, W., Jiang, Q.: Influence blocking maximization in social networks under the competitive linear threshold model. In: Proceedings 12th SIAM International Conference on Data Mining. pp. 463–474 (2012)
8. Kempe, D., Kleinberg, J., Tardos, É.: Maximizing the spread of influence through a social network. In: Proceedings of the 9th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. pp. 137–146 (2003)
9. Li, S., Zhu, Y., Li, D., Kim, D., Huang, H.: Rumor restriction in online social networks. In: Proceedings of the 32nd IEEE International Performance Computing and Communications Conference. pp. 1–10 (2013)
10. Liontis, K., Pitoura, E.: Boosting nodes for improving the spread of influence. Preprint. Available from: arXiv:1609.03478 (2016)
11. Masuda, N.: Immunization of networks with community structure. New Journal of Physics **11**(12), 123018 (2009)
12. Sanz, J., Xia, C.Y., Meloni, S., Moreno, Y.: Dynamics of interacting diseases. Physical Review X **4**(4), 041005 (2014)
13. Wilder, B., Yadav, A., Immorlica, N., Rice, E., Tambe, M.: Uncharted but not uninfluenced: Influence maximization with an uncertain network. In: Proceedings of the 16th Conference on Autonomous Agents and MultiAgent Systems. pp. 1305–1313 (2017)
14. Yadav, A., Chan, H., Xin Jiang, A., Xu, H., Rice, E., Tambe, M.: Using social networks to aid homeless shelters: Dynamic influence maximization under uncertainty. In: Proceedings of the 2016 International Conference on Autonomous Agents & Multiagent Systems. pp. 740–748 (2016)