Limiting Concept Spread in Environments with Interacting Concepts

James Archbold
Department of Computer Science
University of Warwick
J.Archbold@warwick.ac.uk

Nathan Griffiths
Department of Computer Science
University of Warwick
Nathan.Griffiths@warwick.ac.uk

ABSTRACT

The propagation of concepts in a population of agents is a form of influence spread, which can be modelled as a cascade from an initial set of individuals. In real-world environments there may be many concepts spreading and interacting. Previous work does not consider utilising concept interactions to limit the spread of a concept. In this paper we present a method for limiting concept spread, in environments where concepts interact and do not block others from spreading. We define a model that allows for the interactions between any number of concepts to be represented and, using this model, develop a solution to the influence limitation problem, which aims to minimise the spread of a target concept through the use of a secondary inhibiting concept. We present a heuristic, called *maximum probable gain*, and compare its performance to established heuristics for manipulating influence spread in both simulated small-world networks and real-world networks.

CCS Concepts

•**Computing methodologies** → **Multi-agent systems;**

Keywords

Influence spread; Influence limitation; Opinion propagation; Social epidemics; Social simulation

1. INTRODUCTION

In many environments it is possible for strategies, concepts or infections to 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 individuals [12].

Insight gained by understanding how to control cascades in abstract populations has many applications, such as informing epidemic control, viral marketing, and understanding convention emergence in multi-agent systems. For example, characterising the spread of disease aids in identifying at risk groups, enabling containment efforts to be focused to avoid wider spread. Understanding how ideas propagate can inform viral marketing or identify influential individuals. The key enabler is being able to identify the set of individuals who can help spread an idea or product, or who can restrict future spreading (e.g. through their vaccination).

Several models have been developed to characterise influence spread [6, 12], along with techniques to maximise spread [3]. These models represent a population as a network, with individuals being represented as nodes and edges representing the influence that can travel from one individual to another, with some that try to model complex social relationships [15]. Existing models of influence spread typically assume that only a single concept exists, or that concepts are *blocking*, preventing an individual from activating multiple concepts [9, 10]. However, individuals in real world environments can have many concepts active simultaneously. Furthermore, these concepts can interact, affecting the strength with which they spread between individuals.

Not every concept that can spread in a population is desirable or beneficial, for example a disease or rumour, and we may wish to limit its spread. Previous investigations assume blocking concepts, selecting nodes that are either immunised against the undesirable concept or chosen as seeds for a secondary blocking concept. While the use of blocking concepts is reasonable in an epidemiology context, as immunisation is often effective, it is less generally applicable, for example in limiting the spread of rumours or opinions. We assume that individuals can activate, or adopt, multiple interacting concepts, that can affect how other concepts spread.

Minimising the spread of a target concept, through the selection of a seed set for a secondary *inhibiting* concept, is known as the *influence limitation problem*. Previous investigations have focused on finding nodes present on a high number of shortest paths [26], or nodes connecting communities [5]. In environments where concepts are blocking, selecting these nodes prevents the undesirable concept from utilising the most influential network paths. However, when concepts merely inhibit each other, the blocking of a path cannot be guaranteed and such methods are less effective. If a node is on many shortest paths, but is not near to the start of the target concept's cascade, it is unlikely to encounter the target concept, and so cannot help to limit concept spread. Where concepts interact, the likelihood of a node to activate the target concept, and the expected gain from that activation, must be considered when attempting to limit spread.

In this paper, we focus on an adaptation of the influence limitation problem, where concepts can interact. We present

Appears in: *Proc. of the 16th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2017), S. Das, E. Durfee, K. Larson, M. Winikoff (eds.), May 8–12, 2017, São Paulo, Brazil.*

a model of concept interactions and propose the maximum probable gain heuristic, measuring its effectiveness against established methods. Furthermore, we investigate the effect of response time, by attempting to limit the spread of a concept at different stages of a cascade. The evaluation focuses on synthetic small-world networks, as many real-world networks exhibit small-world properties [4, 20], and on selected real-world networks from the Stanford SNAP project.

2. RELATED WORK

Many influence spread problems are sub-modular, and a greedy approach is often effective in approximating the optimal solution [5, 12]. Hill climbing can be used to select the node that provides the largest incremental increase to the performance of the current seed set [7, 8, 12]. However, for real-world problem sizes, this approach is often intractable, since it has a time complexity of $\mathcal{O}(n^3)$ or higher [23]. Despite this, hill climbing is often used as a baseline when new influence spread models are defined. Methods to efficiently evaluate the performance increase from selecting a node have been proposed, notably, the degree discount heuristic [3].

Some greedy approaches can be made tractable, such as the Greedy Viral Stopper (GVS) algorithm, for environments in which 'correct' and 'incorrect' information propagates, with the correct information superseding incorrect [21]. While typically intractable, if we divide the network into communities [1], the GVS algorithm can be performed on individual communities efficiently and the solutions combined.

Methods for influence limitation typically assume that concepts block. Fan *et al.* propose using nodes that connect one community within a network to another, known as *bridge ends* [5]. Similarly, Li *et al.* select nodes identified as 'protector' nodes, whose immunization against an undesired concept protects nodes identified as bridge ends connecting to other communities [17]. Other methods have also been proposed that remove edges between communities [19]. While protecting bridge ends can limit the spread of a concept, it becomes less effective when path blocking cannot be guaranteed. Measuring the betweeness of nodes is also an effective, but computationally expensive, approach [26].

Kotnis and Kuri propose a solution where individuals can be trained, at a cost based on their degree and the quality of training, to be better at deciding if information is a rumour [14]. For a given budget, having more low quality trained individuals produces better results than having fewer individuals with higher quality training. This model assumes a single cascade, but discusses the possibility of other messages affecting the spread of a rumour.

A related problem, selecting a group of nodes and improving their ability to spread a target concept, is discussed by Liontis and Pitoura [18]. Only the selected nodes have this improved spreading ability, which cannot be passed on to neighbouring nodes. This approach is based on the PMIA algorithm for influence spread, proposed by Wang *et al.*, which considers nodes likely to activate a concept, and the expected activations gained from that node's influence [25]. Liontis and Pitoura focus on *boosting* concepts, but these techniques may also prove effective for inhibiting concepts.

An individual's opinions can affect the concepts they activate or spread. These opinions can be represented through network and node attributes, as in the adaptation of the linear threshold model proposed by Kaur and He [11]. Each edge has two separate influence strengths, representing a positive and negative opinion respectively. Similarly, a node has both a positive and negative threshold. A node will activate the opinion that first exceeds its corresponding threshold. Nodes with high positive influence are selected to block the negative opinion from being further spread. Stitch *et al.* present another method of representing opinions, by assigning an attitude score to individual nodes [24]. Nodes with a high attitude score are more likely to spread negative word-of-mouth, even if they have been mostly exposed to positive opinions, and vice versa for nodes with low attitude scores. Both of these models again utilise the blocking assumption.

Budak *et al.* propose the *highest infectees* heuristic that, for environments with 'good' and 'bad' cascades, gives results comparable to greedy hill climbing [2]. This heuristic assumes knowledge of the 'bad' cascade's seed set, and simulates a large number of cascades using that seed set. Nodes are ranked by the number of simulations in which they became infected with the 'bad' cascade, and are selected as seeds for the 'good' cascade in descending order.

Others have focused on modifying edges in the network, rather than selecting nodes to block the spread of a concept. Li *et al.* proposed reducing the edge weight to limit the rate of transmission [16]. The weight of an edge is expressed as a function of the degree of its end points, and the transmission rate between two nodes is proportional to the edge weight. The use of inflammation immunization, which reduces edge weight by a chosen factor, is shown to be effective in this model and may translate well to real-world scenarios. Notably, the reduction of edge weights lowers the transmission weight while not compromising network efficiency as a whole. Conversely, the bond percolation approach suggested by Kimura *et al.* curbs the spread of an infection but, by removing links, damages the network structure and in turn, the ability of the network to transmit other concepts [13].

3. CONCEPT INTERACTION MODEL

To model complex concept interaction, we propose a model for concept interaction based on that presented by Sanz *et al.* [22], extended to be applicable to any number of concepts. We focus on the Independent Cascade Model (ICM), due to its widespread use in previous work. In the ICM, newly activated nodes make one attempt to spread the concept to each of their neighbours, with a probability, p, of success [6]. Although we define propagation and influence strength in terms of the ICM, the model for concept interaction is more generally applicable.

We assume that an environment consists of nodes in a network. Each node, v, has a set of *incoming neighbours*, N_v^i, and *outgoing neighbours*, N_v^o, where N_v^i can influence v, and v can influence N_v^o. These sets are not necessarily disjoint and may be equivalent in some environments, allowing directed and undirected graphs to be represented. A node can adopt, or activate, multiple concepts. The set of *active concepts* for node v at time j is C_v^j.

We must also define how the concepts active on a node affect interactions with other concepts. The concepts already active on a node will affect its ability to *spread* and *adopt* future concepts. For each node, v, we represent the *adopting context* and the *spreading context*. The function $Context_{adopt}^{v,j}(c)$ describes how the concepts active on v at time j affect the chance of v adopting concept c, while $Context_{spread}^{v,j}(c)$ describes how the concepts active on v modify v's chance to spread concept c.

When nodes interact, the spread of a concept can be affected by other concepts active on both the *infector* and the *receiver*. The *concept relationship* function $CR_{spread}(c, c')$ describes the effect of c' on the chance of c being successfully activated on a receiver node, when an infector spreading c also has c' active. Similarly, the concept relationship function, $CR_{adopt}(c, c')$, describes how c' affects the chance of c being adopted by a receiver with c' active.

These functions describe the relationship, positive or negative, between any two concepts. A concept relationship function with a positive value defines a boosting relationship, a negative value defines an inhibiting relationship, while 0 implies that the concepts do not affect each other. Concept relationship functions are used when evaluating the adopting or spreading context of a node, and are node independent.

Some concepts may prevent others from spreading to a node, which is known as blocking. Each concept c has an *excluded set*, X_c, of concepts it blocks. No concept in X_c can activate on a node with c already active.

The ability of a concept to propagate from one node to another is dependent on the strength of the influence one node exerts on the other. The *influence strength*, $I^c_{m,n}(j)$, that node m can exert on node n at time j for concept c, captures this. If $I^c_{m,n}(j) = 0$, then m cannot influence n with respect to concept c.

For this work, we define $I^c_{m,n}(j) = p$ for any pair of nodes such that $m \in N^i_n$ and p is the chance of infection. In each timestep, each node that activated a concept in the previous timestep can attempt to spread that concept to each of its neighbours. To consider concept interaction, we define the concept relationship functions, as follows:

$$CR_{adopt}(c, c') = CR_{spread}(c, c') = r$$

where $r \in [0, 1]$ is a feature of the environment that represents the extent to which concept c' affects concept c. With this definition, c is affected by c' in the same way when c' is active on either the infector or the receiver. For a given concept, c, we must define how it interacts with the context of a node, v, at time j. The adopting and spreading contexts for concept c are defined as:

$$Context^{v,j}_{adopt}(c) = 1 + \sum_{c' \in C^t_v \setminus c} CR_{adopt}(c, c')$$

$$Context^{v,j}_{spread}(c) = 1 + \sum_{c' \in C^t_v \setminus c} CR_{spread}(c, c')$$

Here, we sum the contribution of the corresponding concept relationship function for each concept already active on the node. We add 1 to this sum, allowing the context functions to scale the strength of the influence. Finally, we define how these context functions affect the influence exerted on a receiver by an infector. For a node v, the *contextual influence*, $CI^c_{w,v}(j)$, exerted by incoming neighbour w in relation to concept c at timestep j is defined as:

$$CI^c_{w,v}(j) = I^c_{w,v}(j) \times Context^{v,j}_{adopt}(c) \times Context^{w,j}_{spread}(c)$$

The influence strength is weighted by the adopting context of the receiver and the spreading context of the infector.

4. INFLUENCE SPREAD LIMITATION

Previous work has focused on blocking concept spread, typically allowing a node to have a maximum of one concept active [5, 11, 17, 24]. For the influence limitation problem, we aim to select the seed set for a secondary inhibiting concept to minimise the spread of a primary *target* concept. We do not assume blocking concepts, and so require a method of selection that uses inhibiting concepts effectively.

At the start of a cascade, a set of nodes will begin with a concept active, which can then spread to other nodes. This initial set of nodes is known as the *seed set* for that concept. Selecting seed nodes for a concept in order to limit the spread of another concept requires consideration of a node's position in the network in relation to nodes that have the target concept active. Nodes that are closer to target concept nodes will be more likely to encounter the target concept. Typically, in the ICM, each successive round of a cascade infects fewer nodes than the previous round due to the low probability of infection. This means that early rounds are when a concept maximises its spread, and so limiting early spread is likely to be effective.

To limit spread we should focus on nodes that have a high expected value to the target concept, such as nodes with a high degree, or those present on the shortest path to more central nodes. Infecting these nodes with the inhibiting concept will lower the expected gain of the target concept, and the influence limitation problem becomes a case of maximising that loss of expected gain.

Aiming to maximise the spread of the inhibiting concept may also prove effective. By maximising the spread of the inhibiting concept, we aim to maximise the number of negative interactions with the target concept, inhibiting its ability to spread. In the remainder of this section we describe an established influence maximisation heuristic, namely degree discount, and our proposed influence limitation heuristic, called Maximum Probable Gain.

4.1 Degree Discount

Degree discount has been shown to be an efficient method of influence maximisation, approaching the performance of the greedy algorithm [3]. It relies on calculating the expected nodes gained from adding a node to the seed set. When a node is selected as a seed, the expected gain of selecting its neighbours is lowered. Additionally, those neighbours now have a chance to be activated in the first round of a cascade. Nodes are initially ranked by degree, and when a node is added to the seed set 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. This calculation is based on the expected benefit of such nodes (details of its derivation can be found in [3]).

4.2 Maximum Probable Gain

Typically, heuristics to limit influence spread select nodes based on the assumption that they block the target concept, and as such focus on nodes that are likely to be encountered or that link groups of nodes together. Intuitively, under the blocking assumption, the aim is to remove paths that the target concept could travel, limiting its spread. However, without the blocking assumption, a focus on these nodes may not be the best approach.

Without the blocking assumption, the local influence of a node is more important. We wish to select locally influential nodes, that can spread the inhibiting concept and minimise the spread of the target concept. This is similar to betweeness, and selecting nodes with high betweeness means that they are likely to be encountered by the target concept and to reach a higher number of nodes. However, betweeness is expensive and requires the calculation of a large number of shortest paths. As such, we need an alternative method.

To maximise the chance of interactions between the two concepts, we should focus on nodes more likely to be reached by the target concept. By inhibiting the ability of a node to spread, we also lower the chance of its neighbours activating the target concept. Therefore, the higher the expected gain in target concept activations of a node, the greater the impact to the spread of the target concept if that node activates the inhibiting concept. As such, the aim of the Maximum Probable Gain (MPG) heuristic is to select nodes that are likely to activate the target concept, and provide a high number of expected activations for it.

To calculate a node's viability as a seed node for the inhibiting concept, we set a threshold for exploration, θ. From any node, we consider only nodes that can be reached with a probability higher than θ.

We define the set S_t as the set of nodes with the target concept, t, active. Each node $n \in S_t$ has a set of reachable nodes, R_n, which contains all nodes with a probability of being reached more than, or equal to, θ. The probability of a concept to reach node m from n is the *propagation probability*, $p(n,m)$, which can be calculated recursively using the most probable path from n to m, $MPP(n,m)$. This is the path with the highest chance of traversal, which can be calculated for a neighbourhood of nodes relatively simply.

Since all paths from a node will be affected by that node adopting the inhibiting concept, we consider only the most probable path for simplicity. The most probable path acts as an indicator of the influence between two nodes and has been similarly used by other heuristics with positive results [25].

$MPP(n,m)$ is an ordered set, with length l, of nodes, $\{n_0, n_1 ..., n_{l-1}\}$, where $n_0 = n$ and $n_{l-1} = m$. Thus, for a given node, $n_i \in MPP(n,m)$ we define $p(n,n_i)$ as:

$$p(n, n_i) = p(n, n_{i-1}) \times p(n_{i-1}, n_i) \tag{1}$$

where $p(n,n) = 1$.

We can calculate $MPP(n,m)$, where $m \in R_n$, as we construct the R_n set of reachable nodes, through a snowball sampling method. Starting at n, we sample each neighbour, v, and calculate their $p(n,v)$. Since we are looking at one hop neighbours, this is also their most probable path, $MPP(n,v)$. We add each v for which $p(n,v) > \theta$ to R_n, and explore the neighbours of these nodes. We calculate the propagation probability and most probable path for each of these neighbours, adding each neighbour, u with $p(n,u) > \theta$ to R_n, then repeat our exploration until no nodes are added to R_n. If a node w is encountered multiple times, we choose the neighbour that results in the highest propagation probability, and define $MPP(n,w)$ as:

$$\begin{aligned} MPP(n,w) = \{ & MPP(n,v) \\ & + w \mid \forall m \in N_w^i \wedge m \in R_n : p(n,v) \geq p(n,m)\} \end{aligned} \tag{2}$$

where N_w^i is the set of incoming neighbours of node w.

The process continues until we can add no more nodes that have a propagation probability above θ. Since probability decreases with each hop, all paths to each node in R_n that could possibly have a propagation probability above θ are explored. If we re-consider a node when we encounter it again, we will eventually find the most probable path for all nodes in R_n. Furthermore, if a node, w has the target concept, t, already active then $p(n,w) = 0$ and we do not add it to R_n and so do not explore w's neighbour nodes.

We wish to find the nodes that are most likely to be reached by the target concept, and so we consider each node in the R_n sets for all $n \in S_t$. We wish to find the probability that a node, v will activate the target concept from any possible source, $n \in S_t$. This is known as v's activation probability, $ap(v)$, defined as:

$$ap(v) = \sum_{n \in S_t, v \in R_n} p(n,v) \tag{3}$$

That is, the sum of the propagation probability to v from each node in S_t that can reach v. While $ap(v)$ can exceed 1, the term activation probability is used, relating it to terminology used in other heuristics to calculate similar properties. After calculating $ap(v)$, we must consider the expected gain of a node if it activates the target concept.

To do this, for a node v that is a member of at least one R_n set, where $n \in S_t$, we construct R_v, using the same method as before. Again, we focus on the target concept and, if a node has the target concept already active we do not add it to R_v and do not explore its neighbours. We can find the expected number of nodes that will activate the target concept, $E(v)$, if v activates the target concept by totalling the propagation probability of nodes in R_v:

$$E(v) = \sum_{w \in R_v} p(v,w) \tag{4}$$

If a node has both the target and inhibiting concepts active, its expected gain will lower by a proportion determined by the strength of the inhibiting relationship. The higher the expected gain, the higher the loss of infections for the target concept, and as such we desire nodes with a high $E(v)$.

The expected gain of a node is weighted by the probability it will activate the target concept, to help identify influential nodes. This weighted expected gain for a node v, $WE(v)$, represents the value of v to the spread of the target concept:

$$WE(v) = E(v) \times ap(v) \tag{5}$$

We select the node with the highest $WE(v)$ value as a seed node for the inhibiting concept, as it represents a node likely to interact with the target concept and presents a high possible gain for the target concept.

If we select a node v to be a seed node for the inhibiting concept, its target concept activation probability will change, and similarly for its outgoing neighbours. The expected gain for v's incoming neighbours will also change. As such, we update two groups of nodes.

First, for any node w such that $v \in R_w$, we must update $E(w)$. This involves considering the effect of the inhibiting concept on not only the propagation probability of our chosen node, v, but any nodes m where $v \in MPP(w,m)$. As such, we must recalculate $E(w)$ using the new chance of infection, which takes the relationship between the two concepts into account. For a node m, such that $v \in MPP(w,m)$, we replace the original value of $p(w,m)$ with $p_t(w,m)$, which will include the concept context between v and its preceding and proceeding nodes in the path. The context functions can scale $p(w,m)$ to calculate $p_t(w,m)$, due to $p(w,m)$ being

Table 1: Experimental parameters.

Parameter	Values
Graph Size (nodes)	1000, 5000, 10000, 25000, 50000, 100000
Clustering Exponent	0.25, 0.75
Seed Set Size	10, 25, 50, 100, 250, 500
CR function values	-0.2, -0.4, -0.6, -0.8, -1
Burn-in Timesteps	0, 2, 5

calculated through a series of multiplications:

$$p_t(w,m) = p(w,m) \times Context^{w_{i-1},j}_{spread}(t) \times Context^{w_i,j}_{adopt}(t) \times Context^{w_i,j}_{spread}(t) \times Context^{w_{i+1},j}_{adopt}(t) \quad (6)$$

where t is the target concept, j is the current timestep, and $v = n_i \in MPP(n,m)$. The updated propagation probabilities are used to recalculate $E(w)$ as before.

Second, any node m such that $v \in MPP(n,m)$, where $n \in S_t$, will have its activation probability, $ap(m)$, affected by v being chosen to activate the inhibiting concept. We subtract the original $p(n,m)$ for all n where $v \in MPP(n,m)$ from $ap(m)$ and add $p_t(n,m)$, calculated as above. If $p_t(n,m)$ falls below θ, then we remove it from R_n and do not include $p_t(n,m)$ in m's updated $ap(m)$ calculation. If this causes $ap(m) = 0$, then we do not consider m any further.

With this update, we can recalculate $WE(v)$ for all nodes and select the new highest valued node, then update the required nodes. We repeat this until the seed set for the inhibiting concept has reached its desired size.

5. EXPERIMENTAL SETUP

We wish to evaluate the effectiveness of our proposed heuristic, *Maximum Probable Gain*, for providing a solution to the influence minimisation problem against heuristics used to maximise the spread of an inhibiting concept. For this work, we assume $\theta = 0.001$, to allow for the exploration of the local neighbourhood of a node in the MPG heuristic.

The primary target concept has a randomly selected seed set in all our experiments. We use the following heuristics to select a seed set for our secondary inhibiting concept:

- Random nodes – nodes are chosen randomly
- Highest Degree – we select nodes with highest degree
- Single Discount – nodes with the highest degree that do not connect to a previously chosen seed are selected
- Degree Discount – nodes with the highest 1-hop expected gain are selected
- Maximum Probable Gain (MPG) – nodes likely to activate target concept, with high expected gain, are selected

We evaluate our heuristics using a variety of 'burn-in' times, to explore the impact of response time on effectiveness. The primary target concept will spread for a given number of timesteps, referred to as the 'burn-in', before we introduce the secondary inhibiting concept and select seeds for it. If the burn-in is too long, the primary concept will have completed its cascade by the time we introduce our second concept, and so influence minimisation will be ineffective. Therefore, we focus on short burn-in times.

Furthermore, we test a range of inhibiting relationship strengths. This allows us to determine if different strategies may be more viable at different levels of inhibition.

Simulated small-world networks are used to evaluate the performance of our heuristic, with a range of sizes and two different clustering components, 0.25 and 0.75. These are generated using the Kleinberg small world generator provided in the JUNG graph framework[1]. We also run tests on selected real-world topologies from the Stanford SNAP project[2], namely *DBLP*, *CA-CondMat* and *soc-Epinions1*. The experimental parameters are given in Table 1.

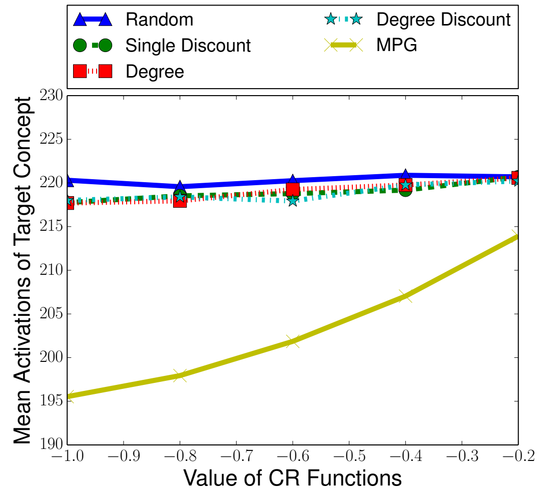

Figure 1: Mean activations of the target concept given the heuristic used to select the inhibiting concept, for small-world networks of 25000 nodes, with a clustering coefficient of 0.75, a seed set of 100 nodes and no burn in time.

6. RESULTS

MPG performs best with no burn-in, as shown in Table 2 where we see a significant difference between the performance of MPG and other heuristics. The other, degree based, heuristics are consistently outperformed by MPG, regardless of seed set size or network clustering coefficient. Further, we note that for the larger networks displayed in Table 2, the performance difference of MPG compared to others begins to extend beyond the range of MPG's standard deviation. In the results displayed in Table 2, there is a statistically significant difference between the performance of MPG and other heuristics ($p < 0.05$), with our certainty increasing as the seed set size increases. We observe no significant difference between the performance of the degree based heuristics, with their expected ranges overlapping significantly. Seed set size impacts performance more than network size, with MPG performing similarly for the same seed set sizes across different networks.

As the burn-in time increases, there is a significant impact in the performance of the MPG heuristic. Figure 1 shows its

[1]http://jung.sourceforge.net/
[2]http://snap.stanford.edu/data

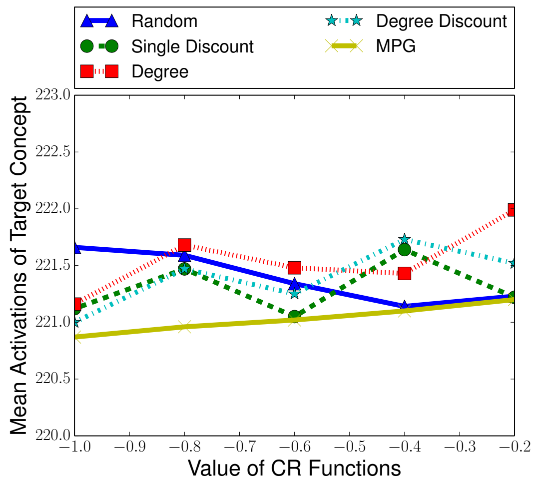

Figure 2: Mean activations of the target concept given the heuristic used to select the inhibiting concept, for small-world networks of 25000 nodes, with a clustering coefficient of 0.75, a seed set of 100 nodes and a burn in of 5 timesteps.

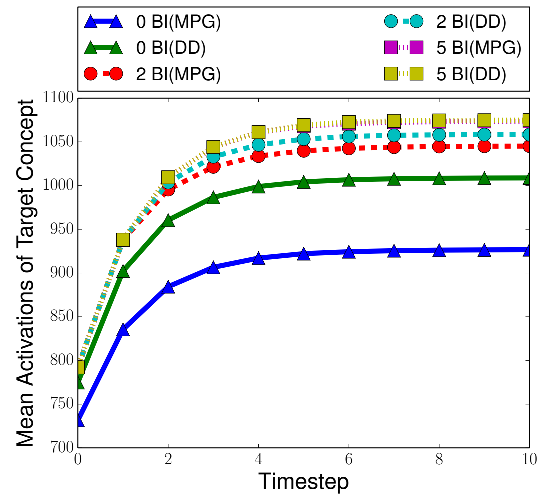

Figure 3: Mean activations of the target concept each timestep in a 25000 node small-world network (0.75 clustering exponent), for varying heuristics and burn-in (BI) times for introducing the inhibiting concept, and a CR relationship of -1 and seed set of 500 nodes.

performance with no burn-in, and Figure 2 presents the performance for a 5 timestep burn-in, highlighting this effect. Higher burn-in times decreases the performance difference between the heuristics, with a burn-in of 5 resulting in nearly identical performance. The expected ranges of each heuristic also begin to converge, and while with larger seed sets we still see MPG outperforming the other heuristics with a low burn-in, at higher burn-in times there is no statistically significant difference between the performance of any heuristic ($p > 0.5$ in all cases). In Figure 2, the results vary by a maximum of 1.2 nodes, which is much smaller than the range of 30 shown in Figure 1. The reason for this performance impact appears to be that, after the first 3 timesteps, activations begin to plateau, as seen in Figure 3. Regardless of the selection strategy, introducing an inhibiting concept as the primary concept's cascade stops will naturally be less effective. In particular, the approach of MPG, which attempts to find nodes with high expected gain and inhibit their ability to spread the target concept, is less effective when a cascade is nearing its end, as there will be fewer activations, reducing the importance of expected gain. Overall, lower burn-in times result in better MPG performance, with no difference between performance at the highest burn-in times.

Thus far we have focused on the larger networks used in our evaluation, however in smaller networks, and smaller seed sets, we still see a statistically significant difference in the performance of MPG compared to the other degree based heuristics. Across all network sizes, it can be seen that MPG performs best with a low burn-in, becoming less effective as the burn-in increases. The performance difference scales with network size, as seen by comparing Figure 4 and Figure 1, although seed set size is also a factor, as discussed above. Figure 5 demonstrates that a high burn-in results in no significant difference between heuristics for smaller networks. A high burn-in results in erratic performance within a narrow range for all heuristics, and this range only minimally increases with network size.

While our discussion in this paper mainly focuses on small-world networks, we also evaluated the performance of MPG for a small set of scale-free networks for comparison. In scale-free networks, all non-random heuristics performed at a similar level. It is only in scale-free networks with a non-zero burn-in time that a statistically significant result is observed, with MPG being outperformed by degree based heuristics. This further highlights the importance of burn-in time. We also note that in scale-free networks, the target concept can spread much further than in similar small-world networks. For scale-free networks with 25000 nodes, and a seed set of 100 nodes, the target concept achieves an average of 389 nodes with the least effective heuristic, random selection. Alternatively, in small-world networks of the same network and seed set size the target concept gains 218 activations when using random selection.

Part of the reason for this reduction in performance could be linked to the degree distribution of the different networks. The MPG heuristic focuses on finding nodes with a high expected cost, that are likely to be activated by the target concept, typically nodes of high degree or the neighbours of those nodes. Due to the existence of hub nodes within scale-free networks, this can result in many seed nodes focused on blocking the target concept from the same node, which is still likely to activate the target concept because of its extremely high degree. High degree nodes are less vulnerable to inhibiting effects, as they will be exposed to the target concept enough times that the activation probability remains high. If a majority of our resources are focused on hub nodes, which in a 25000 node network may include up to 6 nodes with more than 500 edges each, smaller clusters of highly connected nodes can activate and spread the target concept unimpeded. Furthermore, there is a high chance of a hub node activating the target concept regardless and spreading it to its neighbours. Comparatively, in a small-world network of 25000 nodes, we see on average a majority (more than 18000) of nodes have a degree of 5 or 6. This more balanced degree distribution avoids the situation where

Table 2: Average infections for the target concept for small-world networks, with no burn-in and a CR strength of -1, with standard deviation in brackets, and the best performing heuristic in bold.

Network Size	Clustering Exponent	Seed Set Size	MPG	Degree Discount	Single Discount	Degree
10K	0.25	10 Seeds	**19.88 (5.71)**	21.63 (6.39)	21.73 (6.44)	21.74 (6.46)
10K	0.25	25 Seeds	**48.23 (8.94)**	54.33 (10.1)	54.6 (10.43)	54.45 (10.54)
10K	0.25	50 Seeds	**97.01 (13.17)**	108.16 (14.12)	107.71 (14.16)	107.88 (13.95)
10K	0.75	10 Seeds	**19.57 (5.48)**	22.06 (6.4)	21.78 (6.4)	21.78 (6.43)
10K	0.75	25 Seeds	**48.95 (9.12)**	54.03 (10.47)	54.72 (10.26)	54.37 (10.5)
10K	0.75	50 Seeds	**97.81 (13.28)**	107.77 (14.39)	108.21 (14.83)	107.98 (14.17)
25K	0.25	100 Seeds	**196.17 (16.91)**	218.26 (18.59)	218.71 (19.71)	218.13 (19.68)
25K	0.25	250 Seeds	**477.96 (29.03)**	527.64 (30.84)	527.4 (31.53)	528.88 (30.69)
25K	0.25	500 Seeds	**924.11 (39.33)**	1005.3 (40.94)	1007.34 (40.4)	1007.11 (39.46)
25K	0.75	100 Seeds	**195.53 (17.04)**	218 (19.33)	217.75 (17.96)	217.7 (18.56)
25K	0.75	250 Seeds	**476.78 (28.14)**	527.74 (30.88)	526.58 (30.11)	528.58 (30.12)
25K	0.75	500 Seeds	**926.95 (35.66)**	1008.87 (40.34)	1007.51 (39.1)	1006.26 (38.15)
50K	0.25	100 Seeds	**199.03 (18.17)**	220.93 (19.62)	220.66 (19.77)	220.93 (19.79)
50K	0.25	250 Seeds	**487.13 (28.19)**	541.93 (32.733)	541.96 (33.94)	541.58 (33.92)
50K	0.25	500 Seeds	**954.39 (38.48)**	1053.61 (43.77)	1053.82 (43.87)	1052.27 (44.89)
50K	0.75	100 Seeds	**197.37 (17.55)**	219.29 (19.54)	220.23 (19.5)	220.35 (19.8)
50K	0.75	250 Seeds	**485.69 (30.17)**	540.7 (32.74)	541.18 (31.91)	541.1 (32.48)
50K	0.75	500 Seeds	**954.79 (38.74)**	1055.72 (43.79)	1054.37 (43.68)	1054.72 (41.29)
100K	0.25	100 Seeds	**197.88 (19.52)**	221.83 (20.93)	221.8 (20.84)	221.45 (19.87)
100K	0.25	250 Seeds	**491.21 (29.61)**	548.68 (33.08)	548.47 (32.84)	548.44 (33.39)
100K	0.25	500 Seeds	**974.5 (42.95)**	1079.57 (47.63)	1080.02 (47.32)	1079.74 (45.82)
100K	0.75	100 Seeds	**196.83 (17.52)**	221.07 (19.41)	221.72 (19.33)	221.68 (19.6)
100K	0.75	250 Seeds	**490.74 (28.78)**	549.35 (33.95)	549.02 (32.9)	549.17 (33.46)
100K	0.75	500 Seeds	**971.04 (40.17)**	1079.89 (44.8)	1079.78 (46.07)	1079.64 (46)

one node is significantly increasing the expected activations of many, allowing for more influence paths to be affected.

We also evaluated the performance of MPG using real-world network topologies. While these networks have small-world properties, they are not pure small-world networks, as is the case for the synthetic networks considered above. We would therefore expect to see a reduction in the performance of MPG. As in the synthetic networks, we see that a non-zero burn-in time removes the difference in performance and that MPG begins to be less effective, becoming comparable to, or sometimes worse than, other heuristics. Additionally, when compared to synthetic networks, the inhibiting relationship strength has a larger impact on the performance of MPG.

Table 3 shows the performance of different heuristics, when the inhibiting relationship is strongest. We see that MPG's performance within this networks varies. Clustering coefficient is important, as the network that MPG performs least effectively in is also the network with the lowest clustering exponent, namely *soc-Epinions1*. A high clustering coefficient is a characteristic of the pure small-world networks used, highlighting this property as important to the performance of MPG. Looking at the *CA-CondMat* network, we see the impact of diameter and average degree. The higher average degree and lower diameter of *CA-CondMat* when compared to *DBLP* appears to significantly affect MPG's performance. MPG performs significantly better than other heuristics in *DBLP*, while performing inconsistently in *CA-CondMat*. A node with a higher degree has more paths available to spread a concept, and will have a higher number of expected activations. If a network has low diameter, the average path to any other node is shorter. These characteristics are similar to those observed in the pure scale-free networks, and result in spending resources on a node that is likely to activate the target concept regardless.

This may also help to explain why MPG performs comparatively worse in *CA-CondMat* with a higher number of seed nodes, as the target concept will also be wider spread, and MPG will still spend the majority of its resources on any nodes with hub-like properties. MPG performs well in the *DBLP* network and pure small-world networks due to their similar properties. Overall, we observe that these characteristics may be a better indicator of MPG's performance than the use of the labels 'small-world' and 'scale-free', which can both be applicable to real-world networks.

Furthermore, we see that the impact of the inhibiting relationship is greater in real-world networks than the synthetic networks. Particularly within the *DBLP* network, where we see the greatest difference in the performance of MPG and other heuristics. At a CR strength of -0.6, we still see a significant difference, but at -0.4 this diminishes. With a CR strength of -0.2 the difference is minimal, sometimes only 30 nodes, with a standard deviation of 300. This shows that there is no major difference between the heuristics when the inhibiting relationship is weak, even in a network that is favourable towards the use of MPG.

Overall, we see that MPG performs best in small-world networks, with no burn-in time. In these environments, MPG consistently, significantly, outperforms degree based heuristics. With a high burn-in time, all heuristics performed at a similar level, due to the inhibiting concept be-

Table 3: Average infections for the target concept for real-world networks, with no burn-in and a CR strength of -1, with standard deviation in brackets, and the best performing heuristic in bold.

Network	Nodes	Edges	Seed Set Size	MPG	Degree Discount	Single Discount	Degree
DBLP	317080	1049866	100	**695.68 (341.04)**	1076.65 (335.48)	1102.83 (318.76)	1161.06 (389.88)
DBLP	317080	1049866	250	**1117.76 (252.22)**	1710.72 (283.72)	1773.45 (281.41)	1954.52 (381.47)
CA-CondMat	23133	186936	100	**367.31 (70.26)**	415 (63.26)	439.66 (74.06)	440.62 (82.92)
CA-CondMat	23133	186936	250	722.86 (75.84)	**676.51 (55.69)**	699.74 (69.06)	735.29 (65.65)
soc-Epinions1	75879	508837	100	469.36 (123.01)	324.14 (71.84)	**320.57 (72.06)**	324.51 (71.96)
soc-Epinions1	75879	508837	250	938.72 (162.6)	**575.31 (73.9)**	580.86 (71.29)	584.92 (71.17)

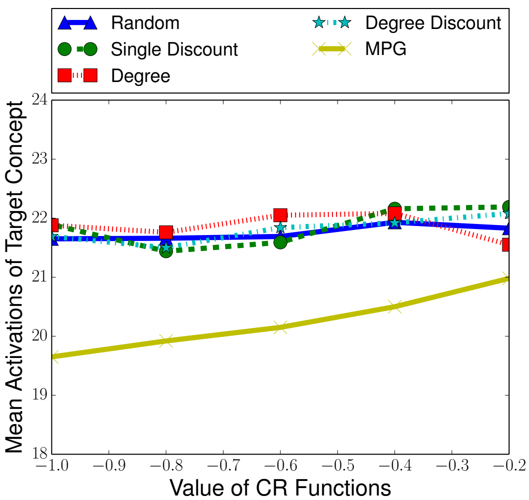

Figure 4: Mean activations of the target concept given the heuristic used to select the inhibiting concept, for small-world networks of 5000 nodes, with a clustering coefficient of 0.75, a seed set of 10 nodes and no burn-in.

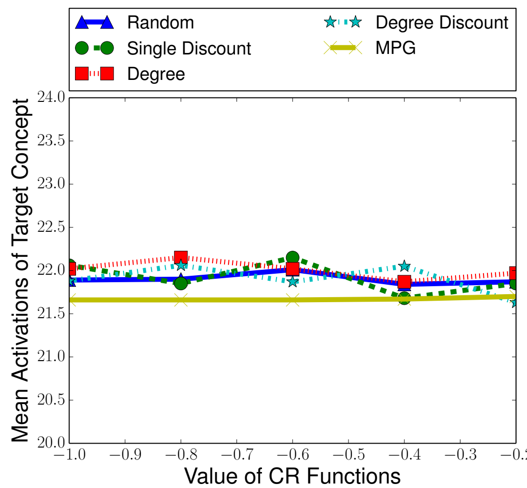

Figure 5: Mean activations of the target concept given the heuristic used to select the inhibiting concept, for small-world networks of 5000 nodes, with a clustering coefficient of 0.75, a seed set of 10 nodes and a burn-in of 5 timesteps.

ing introduced as the target concept's cascade is ending. In scale-free networks, we see no heuristic significantly performing better than others. MPG is outperformed in scale-free networks when the burn-in time is higher, showing that concept maximisation heuristics for the inhibiting concept should be used in those cases. Real-world network tests further show the importance of burn-in time, with lower burn-in times again resulting in better performance for the MPG heuristic. The inhibiting relationship strength is more important in real-world networks, with MPG performing better as it becomes stronger, and the MPG heuristics favours networks with strong small-world properties.

7. CONCLUSIONS

In this paper, we introduce the influence limitation problem, where we aim to limit the spread of a target concept through the use of a secondary inhibiting concept. We propose the maximum potential gain (MPG) heuristic as a solution to this problem, and evaluate its effectiveness against influence maximisation techniques. Our evaluation focused on small-world networks, but also explored a small set of scale-free networks and selected real-world networks. We have shown that MPG performs significantly better than other heuristics in small-world environments with no burn-in time. Burn-in time was shown to be a significant factor in the performance of MPG, with higher burn-in resulting in no statistically significant difference between the heuristics in small-world environments.

In the future, we wish to evaluate the performance of MPG when used for indirect influence maximisation. Since the heuristic aims to find nodes with a high expected potential, and utilise the inhibiting relationship to lower it, it may also be effective if we aim to increase that potential. In addition, we also wish to investigate the performance of MPG in more complex network environments, including multi-layer networks, representing the different social networks an individual may belong to.

REFERENCES

[1] V. D. Blondel, J.-L. Guillaume, R. Lambiotte, and E. Lefebvre. Fast unfolding of communities in large networks. *Journal of statistical mechanics: theory and experiment*, 2008(10):P10008, 2008.

[2] 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.

[3] W. Chen, Y. Wang, and S. Yang. Efficient influence maximization in social networks. In *Proc. of the 15th ACM SIGKDD Int. Conf. on Knowledge Discovery and Data Mining*, pages 199–208, 2009.

[4] I. de Sola Pool and M. Kochen. Contacts and influence. *Social networks*, 1(1):5–51, 1978.

[5] L. Fan, Z. Lu, W. Wu, B. Thuraisingham, H. Ma, and Y. Bi. Least cost rumor blocking in social networks. In *Proc. of the 33rd IEEE Int. Conf. on Distributed Computing Systems (ICDCS)*, pages 540–549, 2013.

[6] J. Goldenberg, B. Libai, and E. Muller. Using complex systems analysis to advance marketing theory development. *Academy of Marketing Science Review*, 9(3):1–18, 2001.

[7] A. Goyal, W. Lu, and L. V. Lakshmanan. Celf++: optimizing the greedy algorithm for influence maximization in social networks. In *Proc. of the 20th ACM Int. Conf. companion on world wide web*, pages 47–48, 2011.

[8] A. Goyal, W. Lu, and L. V. Lakshmanan. Simpath: An efficient algorithm for influence maximization under the linear threshold model. In *Proc. of the 11th IEEE Int. Conf. on Data Mining (ICDM)*, pages 211–220, 2011.

[9] S. Goyal and M. Kearns. Competitive contagion in networks. In *Proc. of the 44th annual ACM symposium on Theory of computing*, pages 759–774, 2012.

[10] X. He, G. Song, W. Chen, and Q. Jiang. Influence blocking maximization in social networks under the competitive linear threshold model. In *Proc. 12th SIAM Int. Conf. on Data Mining*, pages 463–474, 2012.

[11] H. Kaur and J. S. He. Blocking negative influential node set in social networks: from host perspective. *Transactions on Emerging Telecommunications Technologies*, 2015.

[12] D. Kempe, J. Kleinberg, and É. Tardos. Maximizing the spread of influence through a social network. In *Proc. of the 9th ACM SIGKDD Int. Conf. on Knowledge Discovery and Data Mining*, pages 137–146, 2003.

[13] M. Kimura, K. Saito, and H. Motoda. Minimizing the spread of contamination by blocking links in a network. In *Proc. of 23rd AAAI Conf. on Artificial Intelligence*, volume 8, pages 1175–1180, 2008.

[14] B. Kotnis and J. Kuri. Cost effective rumor containment in social networks. *arXiv preprint arXiv:1403.6315*, 2014.

[15] K. Lakkaraju. Modeling attitude diffusion and agenda setting: the mama model. *Social Network Analysis and Mining*, 6(1):1–13, 2016.

[16] H.-J. Li, C. Zhang, and X.-S. Zhang. A study of inflammation immunization strategy in weighted complex network. In *Proc. of the 11th IET Int. Symposium on Operations Research and its Applications in Engineering, Technology and Management*, pages 1–7, 2013.

[17] S. Li, Y. Zhu, D. Li, D. Kim, and H. Huang. Rumor restriction in online social networks. In *Proc. of the 32nd IEEE Int. Performance Computing and Communications Conference (IPCCC)*, pages 1–10, 2013.

[18] K. Liontis and E. Pitoura. Boosting nodes for improving the spread of influence. *arXiv preprint arXiv:1609.03478*, 2016.

[19] N. Masuda. Immunization of networks with community structure. *New Journal of Physics*, 11(12):123018, 2009.

[20] A. Mislove, M. Marcon, K. P. Gummadi, P. Druschel, and B. Bhattacharjee. Measurement and analysis of online social networks. In *Proc. of the 7th ACM SIGCOMM Conf. on Internet measurement*, pages 29–42, 2007.

[21] N. P. Nguyen, G. Yan, M. T. Thai, and S. Eidenbenz. Containment of misinformation spread in online social networks. In *Proc. of the 4th Annual ACM Web Science Conf.*, pages 213–222, 2012.

[22] J. Sanz, C.-Y. Xia, S. Meloni, and Y. Moreno. Dynamics of interacting diseases. *Physical Review*, 4(4):041005, 2014.

[23] S. Shirazipourazad, B. Bogard, H. Vachhani, A. Sen, and P. Horn. Influence propagation in adversarial setting: how to defeat competition with least amount of investment. In *Proc. of the 21st ACM Int. Conf. on Information and knowledge management*, pages 585–594, 2012.

[24] L. Stich, G. Golla, and A. Nanopoulos. Modelling the spread of negative word-of-mouth in online social networks. *Journal of Decision Systems*, 23(2):203–221, 2014.

[25] C. Wang, W. Chen, and Y. Wang. Scalable influence maximization for independent cascade model in large-scale social networks. *Data Mining and Knowledge Discovery*, 25(3):545–576, 2012.

[26] S. Wang, X. Zhao, Y. Chen, Z. Li, K. Zhang, and J. Xia. Negative influence minimizing by blocking nodes in social networks. In *Proc. of the 17th AAAI Conf. on Late-Breaking Developments in the Field of Artificial Intelligence*, pages 134–136, 2013.