Maximising Influence in Non-blocking Cascades of Interacting Concepts

James Archbold[1] and Nathan Griffiths[2]

[1] University of Warwick, UK, archbold@dcs.warwick.ac.uk
[2] University of Warwick, UK, nathan@dcs.warwick.ac.uk

Abstract. In large populations of autonomous individuals, the propagation of ideas, strategies or infections is determined by the composite effect of interactions between individuals. The propagation of concepts in a population is a form of influence spread and can be modelled as a cascade from a set of initial individuals through the population. Understanding influence spread and information cascades has many applications, from informing epidemic control and viral marketing strategies to understanding the emergence of conventions in multi-agent systems. Existing work on influence spread has mainly considered single concepts, or small numbers of blocking (exclusive) concepts. In this paper we focus on non-blocking cascades, and propose a new model for characterising concept interaction in an independent cascade. Furthermore, we propose two heuristics, Concept Aware Single Discount and Expected Infected, for identifying the individuals that will maximise the spread of a particular concept, and show that in the non-blocking multi-concept setting our heuristics out-perform existing methods.

1 Introduction

When autonomous individuals interact, as part of a large population, the propagation of ideas, strategies or infections throughout the population is determined by the composite effect of interactions between individuals. Populations can be viewed as complex systems, with net effects that are hard to predict or influence despite being due to individual behaviour. The propagation of concepts, strategies or infections is a form of influence spread and can be modelled as a cascade from a set of initial individuals through the population.

Understanding how to limit or increase the spread of cascades through a population provides valuable insight into how to influence populations towards a particular state. Such insight has many applications, from informing epidemic control and viral marketing strategies to understanding the emergence of conventions in multi-agent systems. For example, characterising the spread of disease aids in identifying groups of individuals who are at risk, enabling containment efforts to be focused intelligently to avoid wider spread. Understanding how ideas and their adoption propagates can inform viral marketing strategies, or find the network value of individuals in a population. In these cases the key is being able to identify the set of individuals who can help to spread an idea or product, or who can restrict future spreading (e.g. through their vaccination).

Several models have been developed to simulate how influence spreads in a network, and much attention has been focused on the *influence maximisation problem*: finding a set of k nodes (individuals) whose activation will maximise the spread of a particular concept. This problem has been shown to be NP-hard, which has led to the development of heuristics to approximate optimal solutions. Many models assume that cascades are *blocking*, in that a node that has been infected/activated by an idea or concept cannot be activated by any others. However, in many domains individuals can hold multiple opinions, adopt multiple strategies, or have multiple interacting infections. The concepts held by an individual will affect those that they are likely to adopt later, and those that they are likely to propagate to others. This informs the idea of cascades or concepts interacting, however most existing work on influence spread has considered single concepts, or small numbers of blocking concepts.

There has been relatively little consideration of cascades with multiple concepts, and such work has made simplifying assumptions. In the domain of epidemic spread Sanz *et al.* developed a model that allows two concepts to interact [20]. The concepts active on a node affect its ability to activate other nodes, and so the spread of a concept is affected by the other concepts within the network. Concept interaction could also be applied in other cascade models, requiring re-evaluation of existing influence maximisation heuristics. There is also the opportunity to develop heuristics that leverage concept interaction to improve concept spread.

In this paper we focus on non-blocking cascades, and propose a new model for characterising concept interaction in an independent cascade. Specifically, we propose a modification to the independent cascade that incorporates interactions for an arbitrary number of concepts. Furthermore, we propose two heuristics, Concept Aware Single Discount and Expected Infected, for identifying individuals that will maximise the spread of a given concept, and we show that in the non-blocking multi-concept setting our heuristics out-perform existing methods.

2 Related Work

In many application areas it would be valuable to leverage influential nodes to maximise the spread of a concept throughout the population. This is referred to as the *influence maximisation problem* where we aim to pick a (minimal) set of nodes that would maximise the spread of information through the population. Several influence propagation models have been proposed in social network analysis literature [8, 14]. The target set of nodes is activated at the start of influence propagation, and in subsequent cycles, neighbours of active nodes are activated according some model of propagation. Such models can be classified into two types: those that use node-specific thresholds and those based on interacting particle systems [14].

In the *linear threshold model* [14], a node is influenced by each of its neighbours to varying degrees, as defined by the edge weights. Each node v has a threshold θ_v, such that when the sum of the weights of v's active neighbours exceeds θ_v, v becomes active. Methods have been proposed for maximising influence spread within this model [7], but for now our focus is the independent cascade model.

In the *independent cascade model* (ICM) [10], when a node v becomes active it gets one chance to activate each of its inactive neighbours w, with some probability p_{vw}. Kempe *et al.* showed that a hill-climbing approach can be guaranteed to find a set of target nodes that has a performance slightly better than 63% of the optimal set [14]. A key issue with the greedy approach is the need to estimate target set quality. Numerous heuristics have been proposed to improve the speed of estimating the influence spread of a node [1, 6], but it remains problematic in large networks. Building on the greedy approach, Chen *et al.* proposed a degree discount heuristic that accounts for the existing activations in the network and attempts to reduce the impact of 'double counting' [6]. The degree discount heuristic has been shown to have similar effectiveness to the greedy approaches, while remaining computationally tractable.

The problem of influence maximisation has been studied in many contexts. Early studies into influence spread and maximisation focused on the network worth of users [8, 18]. Influence cascades have also been studied in relation to epidemic spread [15, 16, 20]. The two most commonly applied models when characterising epidemics are the Susceptible Infected Susceptible (SIS) and Susceptible Infected Recovered (SIR) models [5, 9]. These both take a probabilistic approach to the independent cascade model, allowing nodes to become deactivated.

Many of these studies have used single cascade models. In many real-world environments, there may be many concepts vying for the attention of an individual. As such, the effect of multiple influence cascades within a single network has been the focus of more recent work on influence spread [11, 12], with consideration of competing cascades that model competing products [2], epidemics [13] or general influence spread [3]. Existing work in this area, has typically assumed that the cascades are blocking, meaning that nodes activated/infected by one cascade cannot be activated/infected by another. Additionally, most existing work assumes only two concepts, while in reality there could be many interacting concepts. It is also often assumed that once activated a node remains active, although there are exceptions to this [17].

Sanz *et al.* developed a multi-layer network model in which concepts may only spread on a given layer but nodes can be activated by more than one concept at a time. Other work on the spread of epidemics also limits their travel to a single layer [19]. Existing research typically either assumes blocking concepts on a network layer, or non-blocking concepts that are each limited to a single layer [12]. There has been little consideration of non-blocking concepts in a single layered network. Much of the work in epidemics focuses on the SIS model and the survival thresholds of viruses, with little exploration of multiple concepts interacting within other models [4].

3 Concept Interaction

To model concept interaction, we extend the work of Sanz *et al.* which modelled two interacting diseases [20]. When attempting to infect a susceptible *receiver*, the infectiousness will change if the receiver is infected with the other disease. Conversely, the infectiousness of a disease is affected by the state of the *infector* spreading it. If the infector has both diseases, their infectiousness will change. This model, which was originally intended for use with SIS and SIR cascade models, is the basis for our approach in the independent cascade.

We must allow for both positive and negative relationships when concepts can interact. If a concept c affects c' in a positive way, we call it *boosting*, while if c is *inhibiting* c' then the effect is negative. How concepts spread in a given cascade model will change the exact effect of boosting and inhibiting. In general, boosting a concept makes it more likely to activate on a node and inhibiting makes it less likely. These relationships can be asymmetric: a concept could boost another concept that inhibits it and vice versa.

The relationship between two concepts, c and c' is defined by two *concept interaction factor* (CIF) functions, which describe the effect of the interaction on the infector and receiver respectively. Each concept active on an infector will be able to affect the spread of any other concept active on the infector. Concepts do not act independently in the real world, their combination and interaction will affect which concepts a node spreads, and how infectious that spreading is. We refer to these interactions as the *internal* effect of a given concept on the infector. The function $CIF_{int}(c, c')$ represents how concept c' affects the spread of c when an infector with both concepts active attempts to spread c. For the receiver, we consider the concepts it has active and the *external* concept that attempts activation. The concepts already active on a node will affect how willing it is to adopt new concepts. Concepts a node has already activated may make it more or less amenable to new incoming concepts, affecting the chance of activation success. The function $CIF_{ext}(c, c')$ represents how concept c' affects the chance of a successful attempt by an infector to activate concept c on a receiver with c' active. These functions are both bounded in the range of [-1,1]. If c' inhibits c, these functions return a value below 0, while above 0 indicates a boosting relationship. If c' does not affect c the functions return 0.

Since real-world environments may have more than 2 concepts we must evaluate the effect the infector's *internal* and receiver's *external* environment will have on the concept currently spreading. Two *concept interaction environment* functions characterise these effects, $CIE_{int}(C_n, c)$ and $CIE_{ext}(C_n, c)$ describe the internal and external environment respectively for a spreading concept c and set of concepts active on node n, C_n. These functions will take into account whether each concept in C_n boosts or inhibits c and return a value that represents how the combined effects of all the concepts in C_n affect either the infector's ability to spread c or the receiver's receptiveness to c.

The notion of concept interaction is independent of the cascade model considered. For illustration, in this paper we focus on the independent cascade [14], as it has been the focus for much influence maximisation research, and extend it to account for multiple interacting concepts. In the standard independent cascade, an infector has chance p of making a neighbour active. With multiple concepts this probability is affected by the CIE_{int} function of the infector and the CIE_{ext} function of the receiver. When node n attempts to activate concept c on node m, the probability of success in the interactive independent cascade becomes:

$$p_c^s = p_c * (1 + CIE_{int}(C_n, c) + CIE_{ext}(C_m, c))$$

Where p_c is the baseline probability for that concept. CIE_{int} and CIE_{ext} are bounded to prevent unbalanced boosting compared to inhibiting. Boosting and inhibiting should have similar impact, rather than one offering more significant change. Therefore, we define the *concept interaction environment* functions as:

$$CIE_{int}(C_n, c) = max(-1/2, min(1/2, \sum_{c' \in C_n} CIF_{int}(c, c')))$$

$$CIE_{ext}(C_n, c) = max(-1/2, min(1/2, \sum_{c' \in C_n} CIF_{ext}(c, c')))$$

This means that p_c^s can range between $[0, p_c * 2]$. Since we must consider each node's environment and the resulting effect on the current concepts spread, this value will be calculated for each interaction.

Cascades proceed in rounds, with an initial set of active nodes for each concept. Nodes can be in more than one initial set. Each node in the initial set for a concept will attempt to active that concept on each neighbour that is inactive for that concept. Each successfully activated neighbour will attempt to activate its neighbours in the next round. Nodes make one attempt on each neighbour for each concept they have active, and when no concept activates new nodes the cascade ends. For simplicity in this paper, we adopt the assumption that nodes will never deactivate a concept.

4 Heuristics for Node Selection

Several heuristics have been proposed for influence maximisation, as discussed in Section 2. In this section we introduce the main existing heuristics and propose two new methods: Concept Aware Single Discount and Expected Infected, which aim to take advantage of concept interaction.

Degree based selection. Degree based selection is the simplest heuristic, and has the advantage of only using attributes of the network, meaning that it is cheap to compute. With the degree heuristic we simply select the k nodes with the highest degree, an approach that has previously been shown to be effective [14].

Single Discount. When a node is added to the selection set, each of its neighbours has a chance to be activated in the first subsequent round of a cascade. However, if it is known that a node will become activated, adding it to the selection set provides no additional value, since that node will be activated regardless of whether it is added to the selection set. This is the motivation behind the single discount heuristic. When a node n is placed into the selection set, the degree of all neighbouring nodes is lowered by 1 to represent their reduced network value (i.e. the number of potential activations they can create has reduced since n is already known to be active). Selection using the single discount heuristic proceeds in rounds, selecting the highest degree node and discounting its neighbours until the desired selection size is reached [6].

Concept Aware Single Discount Heuristic. Introducing concept interaction into the environment requires reconsideration of how concepts spread through a network. Each node can now affect the reach of a concept's spread based on the other concepts they have active. Node degree is typically a good indicator of influence, however in a concept interactive environment this is not always the case. A node with many inhibiting concepts will be less desirable than a node with many boosting concepts if their degrees are equal. Similarly, if a node has many neighbours with active inhibiting concepts, its influence is likely to be low.

We propose a new heuristic, Concept Aware Single Discount (CASD), that weights the degree of a node based on its own concept environment and that of its neighbours, with the aim of providing a more accurate value of node desirability. Specifically, for CASD we define node utility as:

$$U_c(v) = CIE_{int}(C_v, c) + \sum_{n \in N(v)} 1 + CIE_{ext}(C_n, c)$$

where $N(v)$ is v's set of neighbours. Since we are attempting to select nodes that would help to maximise the spread of the targeted concept, the internal environment of a node is a good indicator of node value along with its weighted degree. The external environment of a neighbour of v affects the likelihood of v activating it. The aim of the heuristic is to target nodes with many boosting neighbours and avoid those surrounded by inhibiting nodes. Therefore, $CIE_{ext}(C_n, c)$ is used to increase or decrease the contribution a neighbour makes to the degree of a node. This allows for the concept environment of a node and its neighbours to be considered when evaluating it's worth to the selection set.

Selection proceeds in rounds, with the highest valued node selected each round. When a node n is added to the seed set, neighbour v has its utility updated accordingly:

$$U_c(v) = U_c(v) - (1 + CIE_{ext}(C_n, c))$$

In the same way as Single Discount, we remove the value contributed by the neighbour as it can no longer be activated. Once all neighbours have been updated, the next selection is made, until the required number of nodes is selected.

Degree Discount. Degree discount has been shown to be effective in approaching the optimal solution with reasonable computational overhead [6]. It relies on calculating the expected nodes gained from adding a given node to the selection set. When a node is added, the expected gain of adding its neighbours is lowered. Additionally, those neighbours now have a chance to be activated in the first round of a cascade. The heuristic therefore weights the degree of a node based on these factors, updating the value for any neighbours when a node is added to the selection set. Nodes are initially ranked by degree, and when a node is added to the seed set 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 [6]).

Expected Infected Heuristic. It is important to consider the environment of a node and its neighbours when selecting nodes. The expected payoff from a node will change if it is surrounded by inhibiting concepts compared to boosting ones. Degree Discount is successful because it considers the expected number of activations for a node to decide its value. However, since it is intended for a single cascade model, it requires updating to consider concept interaction. We propose a new heuristic, Expected Infected, with the aim of accounting for these effects.

For each node, v, we consider the set of neighbours with chosen concept c active, $AN_c(v)$. Each of these neighbours will have a chance to activate v, which if successful removes any additional value v would have. The probability of v having c activated by one of these neighbours, $p_a(c, v)$, is:

$$p_a(c, v) = \sum_{n \in AN_c(v)} OP_c(n, v)$$

The sum of the individual chances of each neighbour to activate concept c on v, known as $OP_c(n, v)$, can be defined as:

$$OP_c(n, v) = p_c * (1 + CIE_{int}(C_n, c) + CIE_{ext}(C_v, c))$$

We can now determine the number of activations from $N(v)$ that can be expected as a result of activating v, as follows:

$$EA_c(v) = 1 + \sum_{n \in N(v) \setminus AN_c(v)} OP_c(v, n)$$

In addition to v itself, for each non-active neighbour, we have a $OP_c(v, n)$ chance to activate concept c. Summing the probabilities for each neighbour gives the expected number of neighbours v will activate. However, the chance that v will be activated by a neighbour must also be considered, and so the expected utility for adding v to the seed set is given by:

$$U_c(v) = (1 - p_a(c, v)) * EA_c(v)$$

where $1 - p_a(c, v)$ is the chance of v not being activated. If activated anyway, v will give no additional value. Accounting for this requires scaling $EA_c(v)$ by the

Table 1: Experimental Parameters

Parameter	Values
Graph Type	Small-world, Scale-free
Graph Size (nodes)	1000, 5000
Boost Proportion	0, 0.1, 0.2, 0.3, 0.4
Inhibit Proportion	0, 0.1, 0.2, 0.3, 0.4
Initial set size	1%, 2.5% or 5% of graph size
Intervention set size	1%, 2.5%, 5%, 7.5% or 10% of graph size
Rounds before intervention	5, 10, 25

probability that v does not get activated. Initially nodes have a value of $EA_c(v)$, since there will be no active neighbours.

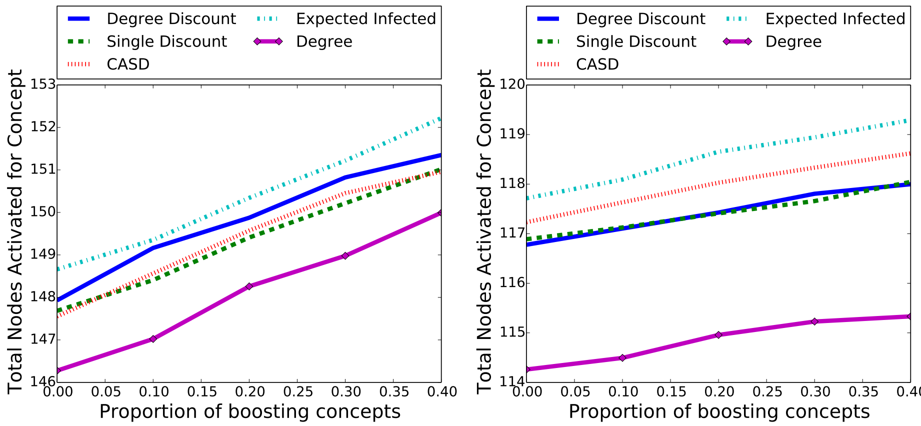
In each selection round, we add the node with the highest value for this heuristic and update its neighbours accordingly, continuing until the selection set is the desired size.

5 Experimental Approach

To evaluate the effectiveness of our proposed heuristics in the context of multiple interacting concepts, we perform simulations using the interactive independent cascade model proposed in Section 3. Each simulation is performed using 10 concepts, with an activation probability for any concept of 0.05. We use the heuristics introduced in Section 4, along with random selection to provide a baseline for comparison. The network topologies listed in Table 1 were used, since they exhibit characteristics found in real-world social networks.

For each simulation, we determine the number of concepts boosted and inhibited by a given concept by selecting from a Gaussian distribution, with a mean of $boost\_proportion * 10$ and $inhibit\_proportion * 10$ respectively, and a standard deviation of 2.5. This, with the proportions defined in Table 1, prevents concepts being too similar and allows for more realistic environments. The final number of concepts boosted or inhibited by a single concept is restricted to be in the range $[0, 5)$.

The initial set of nodes for each concept is selected uniformly at random, and is the size same for all concepts. The cascade proceeds for a fixed number of iterations (a burn-in period) until an intervention occurs, at which point the targeted concept will activate an additional set of nodes selected using a chosen heuristic. The burn-in period before intervention is necessary since the concept aware heuristics require nodes to have concepts activated prior to selection. When selecting intervention nodes, the initial value of a node is discounted considering active neighbours as dictated by the chosen heuristic. This helps to compensate for heuristics that assume no nodes are active at the start. Each heuristic is used for interventions in 100 runs for each combination of parameters in Table 1.



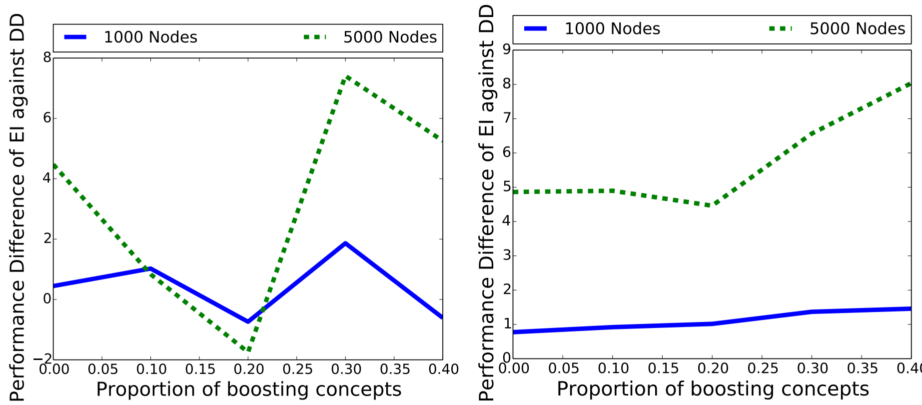
(a) Scale-free network of 1000 nodes (b) Small-world network of 1000 nodes

Fig. 1: Total activations against proportion of boosting concepts

6 Results

We initially compare the performance of each of the heuristics introduced above for a range of parameters. Random selection performed significantly worse than other heuristics in all cases, and while Degree was less effective than the other heuristics it was by a much smaller margin. Therefore, for simplicity of presentation, we do not consider Random selection further. Fig 1a shows the performance of the heuristics as the proportion of boosting concepts increases. We can see that Expected Infected performs best and out-performs our other proposed heuristic, CASD, with results for other topologies and populations mirroring this result. It can be seen that CASD's performance varies based on the environment, at times out-performing degree discount but not consistently. Therefore, the remainder of our analysis focuses on comparing Expected Infected to the best performing of the existing heuristics, namely Degree Discount.

Expected Infected generally outperforms Degree Discount, although in scale-free environments it suffers. Overall, the difference in performance is larger in small-world networks than in scale-free, as shown by Fig 1. This is likely due to the difference in connectivity these two network environments present. In a small-world network, most nodes can be reached with just a few hops from any node, but most nodes do not neighbour each other. Scale-free networks, in comparison to small-world, tend to be more connected and nodes are more likely to be direct neighbours. Boosting concepts seem more advantageous in small-world networks, perhaps since they can allow concepts to cascade for extra hops, and in an environment where most nodes can be reached in a few hops this can greatly help with activation numbers. In scale-free networks however, the high density means each node is more likely to have a high degree, making it easier to spread a concept to at least one neighbour.



(a) Scale-free network

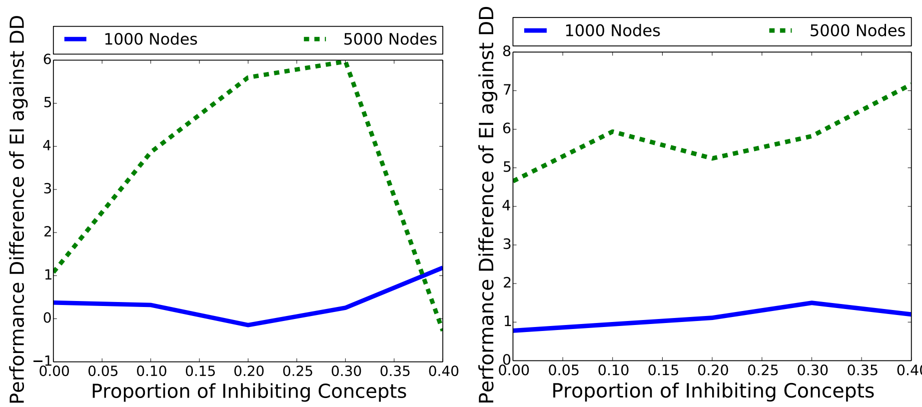
(b) Small-world network

Fig. 2: Difference in activations of Expected Infected (EI) and Degree Discount (DD) against boosting proportions for 1000 and 5000 node graphs

As the proportion of boosting concepts rises, all heuristics improve their total activations, demonstrating the impact of concepts interacting. The advantage of Expected Infected is stable for smaller populations, but is more varied in larger populations. It seems that other network aspects counteract the benefit of targeting boosting concepts. For instance, in larger graphs encountering other concepts may be rarer, making smaller populations more sensitive to concept interaction.

Targeting boosting concepts seems more effective in small-world environments, as their proportion within the network increases. As Fig 2 shows, there is a small overall decrease in performance difference between Expected Infected and Degree Discount for the smaller scale-free environments. In larger populations there is, overall, a small increase. The results for small-world environments demonstrate a steady performance, increasing slightly at higher proportions for both population sizes. The drop off in performance at high boosting proportions for scale-free networks shows that as boosting concepts become more numerous in this environment, explicitly targeting them becomes less advantageous. Naturally, the more boosting concepts exist the easier it is to encounter them by chance. Furthermore, due to their construction through preferential attachment, scale-free networks often have a core group of nodes with high degree. Both Expected Infected and Degree Discount will target nodes of high degree, and such nodes will be more capable of utilising nearby boosting concepts without explicitly targeting them as the proportion of boosting concepts increases. Small-world networks are less likely to have these central nodes, and so the advantage gained from boosting concepts is more valuable.

It can be seen that the performance of Expected Infected compared to Degree Discount suffers a drop of six activations from 0 to a 0.2 proportion of boosting concepts in larger scale-free networks. Then, from a proportion of 0.2



(a) Scale-free network

(b) Small-world network

Fig. 3: Difference in activations of Expected Infected (EI) and Degree Discount (DD) against inhibiting proportions for 1000 and 5000 node graphs

to 0.3 there is a dramatic rise in the performance difference, with Expected Infected having, on average, 9 more activations. The small drop after this at a proportion of 0.4, considering the previous increase, suggests the existence of an optimum proportion of boosting concepts where explicitly targeting them gives a significant benefit. The environment that creates a particular optimum is not clear. It appears to be a more significant factor within scale-free networks, possibly due to their construction being based around hub nodes. Furthermore, the method of choosing the proportion of boosting and inhibiting concepts for a given single concept may also have an effect. The fairly large standard deviation for the Gaussian distribution, considering there are only 10 concepts within any network, means then number of boosting concepts can be quite varied and at lower mean proportions will often be 0. Future work will explore the effect of changing the standard deviation of the Gaussian distribution to investigate this hypothesis.

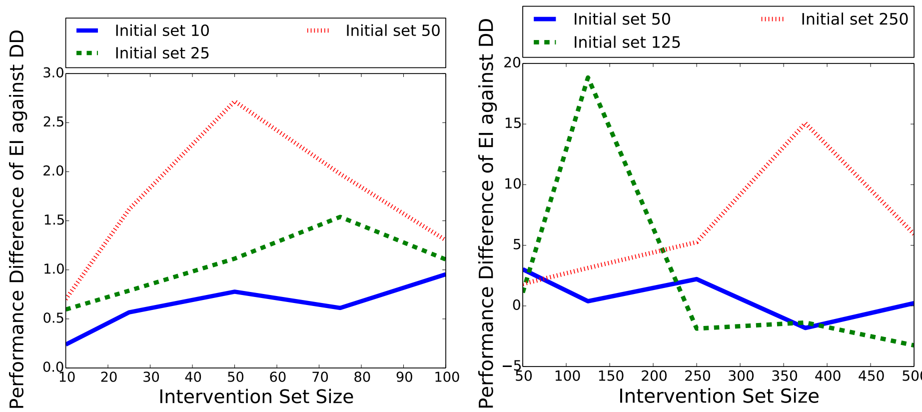
Observing performance against the proportion of inhibiting concepts, Fig 3 shows that performance tends to increase with more inhibiting concepts. This increase is more pronounced in small-world networks, especially in the larger networks. At a high proportion of inhibiting concepts, it becomes difficult to avoid them by chance. In a heavily inhibiting environment, high degree nodes have a higher chance to encounter inhibiting nodes and consequently have their influence diminished. Part of the effectiveness of Expected Infected appears to be in avoiding inhibiting concepts, rather than in taking advantage of boosting ones, as the highest proportion of inhibiting concepts generally exhibits the best performance difference. Scale-free networks demonstrate this behaviour at lower inhibiting proportions, though at higher levels Expected Infected begins to perform worse compared to degree discount. Once again, this could be due to the hub nodes finding the avoidance of inhibiting concepts impossible.

In many of the environments represented within Fig 2 and Fig 3 there is a decline performance when the proportion of boosting or inhibiting concepts reaches 0.2. At this level, it is likely that the boosting/inhibiting concepts are numerous enough that they can be encountered by chance. This will lower the advantage that can be gained by actively targeting or avoiding these concepts. This decline is more prominent in larger networks, likely due to the number of nodes causing the concepts to be sufficiently spread out that they will likely not significantly interact with each other. Together, these factors provide an environment that diminishes the advantages of Expected Infected, hence a drop in performance. Expected infected seems to perform best in two scenarios, namely, when interaction between concepts is low and it can target the advantageous areas that exist or when inhibiting concepts are common and it can actively avoid their detrimental effects.

The size of the initial and intervention sets also impacts performance as shown by Fig 4. For small-world networks we can see in Fig 4a that larger initial sets results in better performance and, mostly, increasing intervention size also improves the performance over degree discount. This increase in coverage makes concept interaction more likely and the consideration of other concepts more advantageous. Furthermore, as the initial and intervention set sizes increase, Degree Discount will find avoiding inhibiting concepts harder, demonstrating the advantage of avoiding them. In scale-free networks the relationship is less consistent, as the environment around the central nodes likely plays a bigger part. Fig 4b shows that at higher populations, performance often actually decreases as intervention set size increases. With the central nodes likely being targeted by all intervention set sizes and the importance of concept interaction consideration, the extra nodes in bigger intervention sets likely do not have many concepts nearby to take advantage of. This means that degree is more important for these extra nodes, and Expected Infected loses the advantage of boosting and inhibiting nodes. Furthermore, we can see most environments in Fig 4 have a peak, followed by a harsh decline in performance. In small-world environments the peak happens earlier for larger initial sets, likely a result of the structure of small-world networks. Since most nodes can be reached by any node in a small number of hops, the chance of choosing a node for the intervention set near other concepts increases for all heuristics as the initial set grows. This means that other heuristics gain from concept interaction, lessening Expected Infected's advantage.

A relationship also appears to exist between network density and performance of Expected Infected, since as density increases Expected Infected performs better. The denser a network, the more edges each node has, and avoiding inhibiting nodes by chance becomes less likely, potentially impacting performance. The effect of such network properties will be a key focus in our future work.

Overall the Expected Infected heuristic performs slightly, but consistently, better then Degree Discount in a non-blocking multi-concept environment aside from in some scale-free environments, and out-performs all other heuristics considered. This includes the second proposed heuristic, CASD, which itself per-



(a) 1000 node small-world network

(b) 5000 node scale-free network

Fig. 4: Difference in activations of Expected Infected (EI) against Degree Discount (DD) for different initial sets against intervention set size

formed inconsistently. CASD occasionally outperforms Degree Discount, particularly in small-world environments or those focused on inhibiting concepts. These are also the environments in which Expected Infected performs best, highlighting that these environments contain key properties for making use of concept interaction. The avoidance of inhibiting concepts seems to be a particular advantage of considering concept interactions, preventing the spread of a concept from being hindered.

7 Conclusion

The study of how ideas, strategies or concepts propagate through a network has many applications. For example, simulations of disease and their infection characteristics can help identify areas at risk of an epidemic that should be the focus of containment, and detecting the influential individuals in a social network allows for the improvement and refining of marketing strategies. This paper introduced an extension of the concept interaction model by Sanz *et al.* [20] to allow for n concepts within the independent cascade. We also proposed two new heuristics, Expected Infected which uses concept relationships to find the expected value of activating a node and Concept Aware Single Discount which adapts the Single Discount heuristic for an environment with concept interactions. Expected Infected was found to out-perform Degree Discount consistently in an interacting concept environment. Specifically, the avoidance of inhibiting factors helps Expected Infected to avoid a concept's spread being hindered. Concept Aware Single Discount was found to be inconsistent, and while it could out-perform Degree Discount in some environments, it is always out-performed by Expected Infected.

Further work to quantify the effect of network properties on concept interactions is needed, to give a better understanding of when best to utilise concept interaction based heuristics. This will include investigating the effect on performance from changing the density of scale-free graphs and the Gaussian function for selecting boosting and inhibiting concept proportions. Observing how our results scale with the increase of concepts within the network would also be of interest, to see if the consideration of inhibiting concepts remains important. The current model is also simplistic in its approach to concept interaction, and extending the model to allow for nodes to deactivate concepts or for concepts to deactivate other concepts could provide more realism to these simulations. If concepts can be deactivated it is likely avoiding inhibiting concepts will become even more important, however the extent to which this is the case remains to be seen.

References

1. A. Anagnostopoulos, R. Kumar, and M. Mahdian. Influence and correlation in social networks. In *Proc. of the 14th ACM SIGKDD Int. Conf. on Knowledge discovery and data mining*, pages 7–15, 2008.
2. K. R. Apt and E. Markakis. Diffusion in social networks with competing products. In *Algorithmic Game Theory*, pages 212–223. 2011.
3. A. Borodin, Y. Filmus, and J. Oren. Threshold models for competitive influence in social networks. In *Internet and network economics*, pages 539–550. 2010.
4. C. D. Brummitt, K.-M. Lee, and K.-I. Goh. Multiplexity-facilitated cascades in networks. *Physical Review E*, 85(4):45–102, 2012.
5. D. Chakrabarti, Y. Wang, C. Wang, J. Leskovec, and C. Faloutsos. Epidemic thresholds in real networks. *ACM TISSEC*, 10(4):1, 2008.
6. 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.
7. W. Chen, Y. Yuan, and L. Zhang. Scalable influence maximization in social networks under the linear threshold model. In *IEEE 10th Int. Conf. on Data Mining*, pages 88–97, 2010.
8. P. Domingos and M. Richardson. Mining the network value of customers. In *Proc. of the 7th ACM SIGKDD Int. Conf. on Knowledge discovery and data mining*, pages 57–66, 2001.
9. S. C. Ferreira, C. Castellano, and R. Pastor-Satorras. Epidemic thresholds of the susceptible-infected-susceptible model on networks: A comparison of numerical and theoretical results. *Physical Review E*, 86(4):41–125, 2012.
10. J. Goldenberg, B. Libai, and E. Muller. Using complex systems analysis to advance marketing theory development. *Academy of Marketing Science Review*, 2001(9):1–20, 2001.
11. 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.
12. X. He, G. Song, W. Chen, and Q. Jiang. Influence blocking maximization in social networks under the competitive linear threshold model. In *SIAM Int. Conf. on Data Mining*, pages 463–474, 2012.

13. B. Karrer and M. Newman. Competing epidemics on complex networks. *Physical Review E*, 84(3):36–106, 2011.
14. 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.
15. J. Leskovec, A. Krause, C. Guestrin, C. Faloutsos, J. VanBriesen, and N. Glance. Cost-effective outbreak detection in networks. In *Proc. of the 13th ACM SIGKDD Int. Conf. on Knowledge discovery and data mining*, pages 420–429, 2007.
16. R. Pastor-Satorras and A. Vespignani. Epidemic spreading in scale-free networks. *Physical review letters*, 86(14):3200, 2001.
17. N. Pathak, A. Banerjee, and J. Srivastava. A generalized linear threshold model for multiple cascades. In *IEEE 10th Int. Conf. on Data Mining*, pages 965–970, 2010.
18. M. Richardson and P. Domingos. Mining knowledge-sharing sites for viral marketing. In *Proc. of the 8th ACM SIGKDD Int. Conf. on Knowledge discovery and data mining*, pages 61–70, 2002.
19. F. D. Sahneh and C. Scoglio. May the best meme win!: New exploration of competitive epidemic spreading over arbitrary multi-layer networks. *arXiv:1308.4880*, 2013.
20. J. Sanz, C.-Y. Xia, S. Meloni, and Y. Moreno. Dynamics of interacting diseases. *arXiv:1402.4523*, 2014.