The Costs of Overambitious Seeding of Social Products

Shankar Iyer and Lada A. Adamic

Facebook, Inc., 1 Hacker Way, Menlo Park, CA 94025,
shankar94@fb.com and ladamic@fb.com

Abstract. Product-adoption scenarios are often theoretically modeled as "influence-maximization" (IM) problems, where people influence one another to adopt and the goal is to find a limited set of people to "seed" so as to maximize long-term adoption. In many IM models, if there is no budgetary limit on seeding, the optimal approach involves seeding everybody immediately. Here, we argue that this approach can lead to suboptimal outcomes for "social products" that allow people to communicate with one another. We simulate a simplified model of social-product usage where people begin using the product at low rates and then ramp their usage up or down depending upon whether they are satisfied with their experiences. We show that overambitious seeding can result in people adopting in suboptimal contexts, where their friends are not active often enough to produce satisfying experiences. We demonstrate that gradual seeding strategies can do substantially better in these regimes.

Keywords: product adoption, influence maximization, social networks

1 Introduction

How should product developers optimally introduce a new product to the people who they hope will adopt it? These scenarios are commonly theoretically modeled as "influence maximization" (IM) problems [9]. In IM, it is assumed that people can influence one another's decisions to purchase or adopt a product. Meanwhile, there is a budgetary constraint on the number of people to whom the developers can directly give or market the product. Then, the problem is to identify the set of people to whom the developers *should* introduce the product so as to maximize the expected long-term adoption. In most IM models, if there is no budgetary constraint, then the optimal approach is to seed everybody immediately [12]. In this paper, we argue that this approach can lead to suboptimal long-term usage in the case of products that enable people to communicate with their friends (i.e., "social products").

Prior research has identified several reasons why overambitious seeding can be detrimental, including direct costs of product rejection [1], downstream negative word-of-mouth [10], and aversion to products that are too popular [2]. These are all important considerations, but we argue that there can be a distinct negative mechanism at play, specifically for social products.

To identify this mechanism, we study a simplified model of social-product usage that takes into account the experiences of people who use the product, not just their binary adoption decisions. Our model encodes some plausible aspects of user experience in social products:

1. **Need for social support**: a person's satisfaction with a product experience depends upon how many of their friends are using it.
2. **Rate-of-usage adjustments**: When people gain access to the product, they begin using it at a low rate p_0 and then gradually ramp their rate of usage up or down depending upon whether they are satisfied with their experiences.
3. **Possibility of permanent churn**: if people have enough unsatisfying product experiences, they churn permanently and are unwilling to try the product again.

We simulate our model on synthetic and real-world networks and ask the following question: is long-term activity maximized by giving everyone access immediately, or is it preferable to seed only a subpopulation and then expand access out gradually?

Summary of Results We show that there is a key tradeoff at play in our model:

1. There is a cost to giving someone access early on, because that person's friends are (by assumption) using the product at low rates, so the risk of unsatisfying experiences is relatively high.
2. There is also a cost to *not* giving someone access early on: others miss out on that person's social support, and their experiences may therefore be worse.

Suppose we can identify a subpopulation of people who can sustain long-term activity of the product on their own, without needing the social support of their friends in other parts of the social network. Then, it may be advantageous to introduce the product to just that set of people at time $t = 0$ and then gradually expand access. In this way, we can avoid some of the costs of item 1 above, because we can expand access in conditions that are conducive to good experiences. In the case of real-world social networks, we show that the tradeoff generally favors this approach.

These findings cast the well-studied IM problem in a new light, by highlighting how thinking about people's experiences *after* the initial adoption can fundamentally change our perspective on the optimal deployment approach.

2 The Model, the Networks, and the Simulations

Model Definition We study a model of social-product usage on an undirected social network. Each node v represents a person who may use or not use the product in a given time step t. Each edge e represents a friendship tie between two people.

Our model proceeds in a sequence of time steps, beginning with $t = 0$. At any given time t, a person v can either have access to the product or not. If the person has access, then the person uses the product in that time step with probability $p_v(t)$ and abstains otherwise. At the time t_v when a person v initially gets access, $p_v(t_v)$ is initialized to a value p_0.

We associate a threshold s_v with each person. If v uses the product in time step t, then s_v is the number of friends of v who also need to use the product at time t for v to be satisfied. Then, v adjusts his or her probability of subsequent usage up or down as follows:

$$\begin{aligned} p_v(t+1) &= p_v(t) + \delta \text{ if } > s_v \text{ friends use in time step } t \\ &= p_v(t) \text{ if } s_v \text{ friends use in time step t} \\ &= p_v(t) - \delta \text{ if } < s_v \text{ friends use in time step t} \end{aligned}$$

This encodes the "need for social support" and "rate-of-usage adjustment" assumptions. We allow $p_v(t)$ to grow to 1 or drop to 0. The latter case encodes the "possibility of permanent churn" assumption, because the person in question will no longer have any opportunities to have positive experiences. On the other hand, $p_v(t) = 1$ is not necessarily a permanent state.

In our simulations, for simplicity, we consider situations where the threshold s_v can take on two values, 1 or 2. We refer to people who people who need only $s_v = 1$ active friend to be satisfied as "enthusiasts." Meanwhile, we refer to people who require $s_v = 2$ active friends as "skeptics."

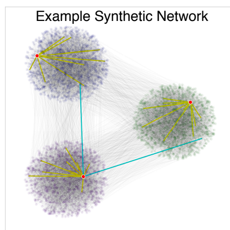
We typically simulate our model on networks that share the modular structure of many empirical social networks, where certain groups of nodes (a.k.a. clusters) are more densely connected to one another than they are to the other nodes in the network [4,13]. We allow the average propensity for adopting social products to vary between these clusters by assigning to each cluster c a probability $f_{sk,c}$ that a person within c is a skeptic. If we assign $f_{sk,c} = 0.4$, that means that each person in c is randomly designated as a skeptic with probability 40% and is otherwise an enthusiast.

Within these scenarios, we ask whether it is optimal to adopt the "single-shot universal seeding policy" of seeding all clusters at once, or if it is better to initially seed a few clusters and then gradually expand access. In the latter case, we need some protocol for implementing the gradual expansion of access. We expand access to a new person when they have had at least two friends using the product in each of five consecutive time steps, so that a skeptic would have been consistently satisfied. This is one example of what we have called a "gradual seeding strategy," and we refer to it below as the **gradual access expansion rule**[1].

Our model intentionally excludes (a) a budgetary constraint on seeding, (b) rejection of the product upon gaining access, and (c) negative word-of-mouth or nonconformism effects. Thus, the negative costs of overambitious seeding in our model arise through different mechanisms than those studied previously [1, 2, 7, 10].

[1] Other variants of this rule can certainly be considered and may even lead to better long-term outcomes, but this choice suffices to demonstrate our main results.

Synthetic Networks As mentioned above, we are interested in studying the performance of various cluster-level seeding strategies on networks composed of several modular clusters. To build synthetic clusters, we generate k_{ic}-regular random graphs, where k_{ic} denotes the "in-cluster" degree. The random graphs are built using the NetworkX Python package [5, 11, 16]. Then, with three k_{ic}-regular random graphs in hand, we connect them by randomly activating a set of intercluster links. We activate sufficiently many links such that the average "out-of-cluster" degree $\langle k_{oc} \rangle$ matches a prescribed value[2].



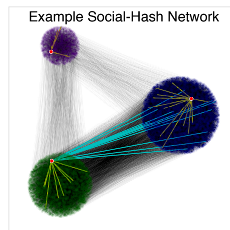


Fig. 1. On the left, an example synthetic three-cluster network, where each cluster is a 10-regular random graph of 1000 nodes and $\langle k_{oc} \rangle = 1$. On the right, an example social-hash network. One randomly chosen node is highlighted (in red) in each cluster: that node's within-cluster edges are shown in yellow and out-of-cluster edges are shown in cyan. These networks were generated for illustrative purposes at a later date than the networks used in the simulations.

Facebook Friendship Networks Real social networks can exhibit properties such as clustering and assortativity that are absent from our synthetic networks. To argue that overambitious seeding can be problematic on real-world networks, we also run our simulations on portions of the Facebook friendship graph. The Social Hash (SH) clustering was originally developed to enable faster data retrieval by physically collocating data for people who communicate frequently. Thus, many (but not all) of a person's frequently contacted friends belong to the same SH cluster [8, 15].

This property is well matched to the type of cluster-level approaches that we want to test, so we simulate our model on SH clusters containing US Facebook users who visited in the 28 days leading up to and including 2018-04-29. We use the clusters in a deidentified fashion, where we never access any information about the identity of individual people within the clusters or even aggregate demographic properties of the cluster (e.g., location within the US, age group, etc.). We also identify three-cluster networks such that each cluster has average

[2] We set $\langle k_{oc} \rangle < k_{in}$, so the average person has many more friendships within the same cluster than with people in the other clusters.

SH Cluster Properties							
ID	N_c	$\langle k_{ic} \rangle$	P5 k_{ic}	P50 k_{ic}	P95 k_{ic}	C	r
A	3772	8.6	1	6	26	0.27	0.25
B	1572	23.6	1	7.5	93	0.44	-0.21
C	527	4.3	1	2	30	0.25	0.97
D	7362	13.8	1	8	48	0.34	0.24
E	2174	7.5	1	5	24	0.27	0.32
F	3413	10.8	1	7	33	0.28	0.25
G	1488	7.2	1	5	20	0.28	0.28
H	806	7.7	1	4	27.75	0.32	0.4
I	835	17.9	1	7	70.3	0.22	0.22
J	11942	145.2	6	83	517	0.22	0.18
K	1473	75.0	2.6	54	220	0.34	0.03
L	22774	64.1	9	51	165	0.29	0.31
M	1849	27.6	1	20	82.6	0.29	0.26
N	17255	35.7	4	29	89	0.20	0.14
O	2372	13.0	1	6	49	0.27	0.47
P	9839	26.4	3	20	73	0.23	0.39
Q	12917	54.5	5	44	139	0.21	0.23
R	1692	18.2	1	10	65	0.31	0.30
S	1479	10.3	1	6	37	0.24	0.34
T	3036	12.0	1	8	35	0.21	0.27

Table 1. Statistics of individual SH clusters. C refers to the average clustering coefficient over all nodes, and r is the degree assortativity.

$\langle k_{oc} \rangle >= 1$[3]. Tables 1 and 2 show that there is considerable structural diversity amongst the SH clusters and three-cluster networks respectively[4].

SH Network Properties								
		Percentage of nodes where k_{oc}						
ID	N	$= 0$	$= 1$	$= 2$	$= 3$	> 3	C	r
FGH	5783	49.1%	17.6%	9.3%	6.3%	17.7%	0.27	0.23
IJK	14259	64.8%	13.3%	6.6%	3.6%	11.7%	0.23	0.22
LMN	41797	46.9%	22.2%	11.0%	5.7%	14.2%	0.24	0.36
OPQ	25222	41.3%	22.3%	13.0%	7.0%	16.4%	0.20	0.35
RST	6268	53.8%	16.4%	8.9%	5.2%	15.7%	0.22	0.36

Table 2. Statistics of networks composed of three SH clusters from Table 1.

[3]On average, each person in each cluster has at least one out-of-cluster friend.

[4]The size of the three-cluster network can differ slightly from the size of the three clusters individually because, in both cases, we exclude people with zero degree, who would inevitably churn under our model. In a small percentage of cases, a person who has no within-cluster friends may still have friends in another cluster when three clusters are considered together.

3 Simulation Results and the Costs of Overambitious Seeding

3.1 Synthetic Networks

Simulations on Individual Clusters Figures 2 and 3 report the results of simulations on individual synthetic clusters, where the entire cluster is given access at $t = 0$. Figure 2 shows examples of full time series on individual clusters. These time series illustrate that 10^4 time steps are generally sufficient to reach the steady-state regime in terms of activity. Meanwhile, they also provide a window into what is going on dynamically. There are cases, for example when $p_0 = 0.16$, where the average activity initially drops but eventually rebounds. This indicates two different processes are going on simultaneously early on: some people are finding their product experiences unsatisfying and decreasing their usage while others are finding their experiences satisfying and increasing their usage. Initially, the former process dominates the trend in average activity. Eventually however, the latter process takes over and dominates the curve, with the activity rate of the non-churned population steadily rising. In our simple model of social-product usage, the non-churned population is actually active in every time step in the steady state, so the percent that is *not* active at late times represents the population that has churned.

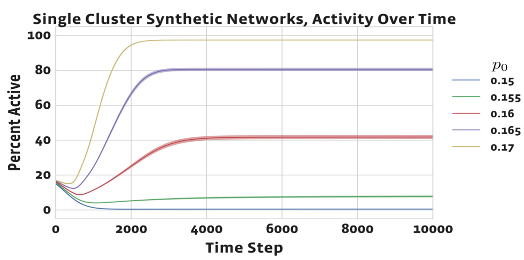


Fig. 2. Average time series of the active percentage in simulations on single synthetic clusters. Everyone in the cluster is given access at time $t = 0$. In this and all subsequent plots, we include 95% confidence intervals, but they are often smaller than the plot line. [Parameters: $k_{ic} = 10$, $N_c = 10^4$, $f_{sk,c} = 0.75$, $\delta = 0.005$, 100 simulations per time series.]

In Figure 3, we plot the asymptotic level of activity for various values of p_0 and $f_{sk,c}$. There is an approximate but intuitive pattern that emerges[5]. When everyone is an "enthusiast" ($f_{sk,c} = 0$) and is satisfied with one active neighbor, then $p_0 = \frac{1}{k_{ic}}$ approximately demarcates the boundary between a regime where people are typically satisfied by their early experiences ($p_0 > \frac{1}{k_{ic}}$) and where they are not. Meanwhile, when everyone is a "skeptic" ($f_{sk,c} = 1$), this dividing point gets shifted to approximately $p_0 = \frac{2}{k_{ic}}$. Intermediate skeptic fractions interpolate between these two extremes, with a crossover occurring somewhere between $p_0 = \frac{1}{k_{ic}}$ and $p_0 = \frac{2}{k_{ic}}$.

[5] Figure 3 shows the case of $k_{ic} = 10$, but this approximate pattern holds for $k_{ic} = 50$ as well.

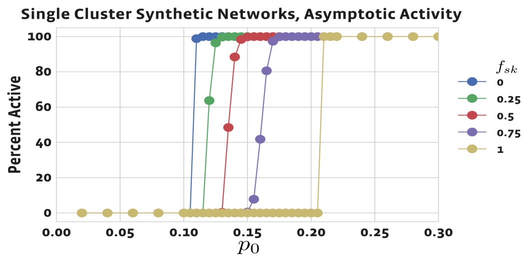


Fig. 3. Average asymptotic activity vs. initial activity probability p_0 in simulations on single synthetic clusters. Everyone in the cluster is given access at time $t = 0$. Here and in all subsequent plots, "average asymptotic activity" refers to an average over the last 100 of 10^4 time steps. [Parameters: $N_c = 10^4$, $k_{ic} = 10$, $\delta = 0.005$, 100 simulations per data point.]

Comparison of Seeding Strategies on Three-Cluster Networks We now turn our attention to simulations on networks composed of three synthetic clusters, as shown on the left-hand side of Figure 1. We label the three clusters C_0, C_1, and C_2, such that $f_{sk,C_0} \leq f_{sk,C_1} \leq f_{sk,C_2}$. When seeding n clusters, we choose the n with the lowest values of the skeptic fraction. We then ask if the optimal strategy is to seed $n = 3$ clusters at $t = 0$ or if $n < 3$ leads to a better long-term outcome.

Figure 4 shows how the asymptotic activity varies with the seeding strategy as p_0 is varied. Consider the case in Figure 4b where $p_0 \leq 0.02$ and $k_{ic} = 50$. In this regime, if the cluster C_0 had no connections to C_1 and C_2, we would expect from the argumentation around Figure 3 that activity would die out. We see from Figure 4b that adding the social support of the other two clusters does not change this situation. As p_0 grows however, it eventually reaches levels where:

1. The value of p_0 is high enough that C_0 can sustain activity in isolation.
2. The value of p_0 is high enough that the *combination* of activity in C_0, C_1, and C_2 at early times provides enough social support to C_0 to sustain activity in that cluster.

Note that the value of p_0 where case 2 occurs cannot be higher than the value of p_0 where case 1 occurs, since the support from C_1 and C_2 strictly adds to within-cluster social support. This implies that, as p_0 grows, there will initially be a regime where the single-shot universal seeding strategy will be optimal, as we see in the $p_0 = 0.022$ curve in Figure 4b.

As p_0 continues to grow however, eventually C_0 and C_1 collectively will be able to sustain activity in isolation from C_2. An example of this is the $p_0 = 0.105$ curve in Figure 4a. Here, the single-shot universal seeding strategy results in *lower* asymptotic activity than seeding two clusters. At still higher p_0, the long-term activity is monotonically *decreasing* in the number of clusters that are initially seeded, because C_0 can sustain activity entirely on its own, and access and usage can then successfully diffuse into C_1 and C_2.

However, Figure 4 also suggests that the regime of p_0 where seeding strategy matters at all is small. We will now see that, when we look at more realistic networks, the gradual seeding strategies win over wide ranges of p_0.

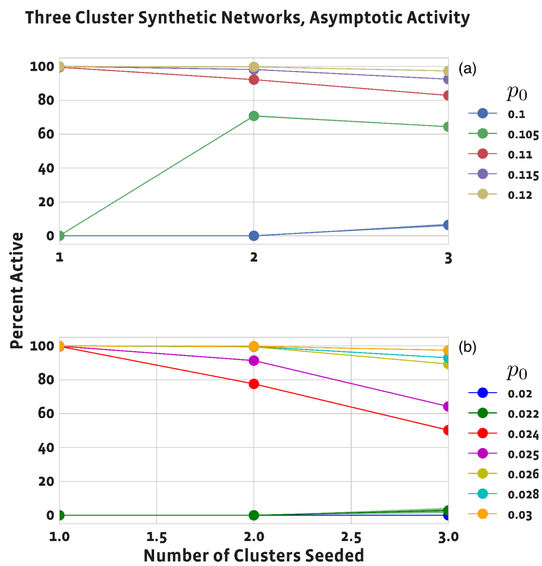


Fig. 4. Average asymptotic activity vs. number of initially seeded clusters in simulations on three-cluster synthetic networks. Clusters are prioritized for seeding by their skeptic fraction $f_{sk,c}$, as described in the text. [Parameters: $N_c = 10^4$, $\langle k_{oc} \rangle = 1$, $k_{ic} = 10$ in (a) and 50 in (b), $f_{sk,C_0} = 0$, $f_{sk,C_1} = 0.25$, $f_{sk,C_2} = 0.5$, $\delta = 0.005$, 100 simulations per data point.]

3.2 Simulations: Facebook Friendship Graphs

Simulations on Individual Clusters In Figure 5, we report the results of simulations on individual SH clusters from Table 1. Here, as in Figure 3, we simulate a situation where the entire cluster is given access at time $t = 0$. Compared to the results for k_{ic}-regular random graphs in Figure 3, the crossover from low asymptotic activity to high asymptotic activity is now smoothed out over a broader range of p_0.

This is because the SH clusters exhibit variance in the degree distribution, while the k_{ic}-regular random graphs have none. At the same level of initial activity p_0, people with a higher number of friends are more likely to be satisfied with their early experiences. Furthermore, most of the SH clusters exhibit assortativity, so people with higher degree are relatively likely to be friends with one another. This means that there can be local patches of the cluster that can sustain long-term activity, even at low values of p_0. Other patches of the cluster, with lower typical degree, will cross over at higher values of p_0, leading to the smeared out crossover observed in Figure 5.

Comparison of Seeding Strategies on Three-Cluster Networks We now turn our attention to simulations in networks composed of three SH clusters, as described in Table 2. Here, for simplicity, the three SH clusters have the same skeptic fraction, but there are still differences in how much activity the three clusters can sustain in isolation, owing to the between-cluster variance in k_{ic}. Thus, we can prioritize clusters for seeding in decreasing order by the median within-cluster value of k_{ic}.

Figure 6 shows the asymptotic activity in network RST under various seeding strategies. Compared to the case of synthetic networks, we observe a much wider range of p_0 where the asymptotic activity monotonically decays with the number

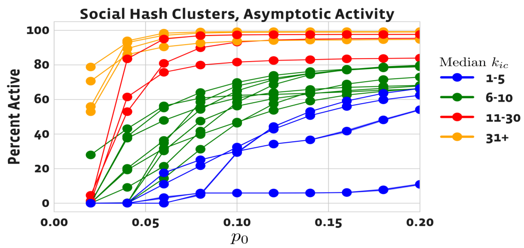


Fig. 5. Average asymptotic activity vs p_0 in simulations on SH clusters. Everyone in the cluster is given access at time $t = 0$. The curves are grouped by median k_{ic}. [Parameters: $f_{sk,c} = 0.75$, $\delta = 0.005$, 50 simulations per data point]

of initially seeded clusters. In Figure 7, we plot full time series for the case of $p_0 = 0.04$. The bottom panel shows time series under the single-shot universal seeding approach. Here, we can see that all three clusters incur the costs of early seeding. They all achieve higher asymptotic activity than we would expect from Figure 5, because people in each cluster experience social support from friends in the others. However, the top panel shows that it is possible to do much better. When we seed only cluster R, enough long-term activity develops in R to allow access to diffuse into the other clusters. When access does diffuse into S and T, it does so in contexts that are favorable for continued usage, leading to much higher asymptotic activity in these clusters.

Why does this story hold, as Figure 6 suggests, over a much broader range of p_0 than in the synthetic case? Even at low p_0 values, long-term activity can be sustained in at least some patches of the single seeded cluster. Often, these patches are sufficient to induce the diffusion of access (and subsequently usage) to the other clusters. Thus, we should expect gradual seeding strategies to generally outperform single-shot universal seeding on realistic networks.

In Figure 8, we plot the difference in asymptotically active percentage under the two strategies, on RST and four other SH networks. This plot confirms that gradual seeding strategies lead to better results over wide ranges of p_0 for most networks, despite the structural differences reported in Table 2[6].

4 Conclusion

In this paper, we have identified a mechanism through which overambitious seeding can harm long-term usage of social products, which depends crucially on the contextual value of products that enable social communication: if people begin using these products before their friends are doing so sufficiently frequently, this can lead to people having unsatisfying experiences and eventually abandoning the product. Seeding policies that involve distributing the product as widely as possible as soon as possible, though favored by usual IM models, are likely to

[6]There is only one network (FGH) where single-shot universal seeding wins, and then only at high p_0

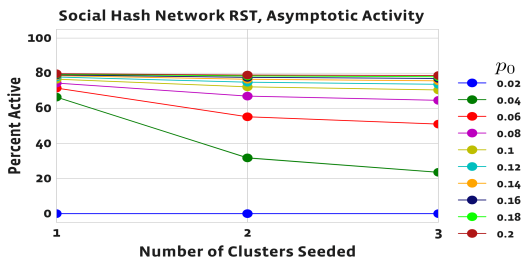


Fig. 6. Average asymptotic activity vs. number of clusters initially seeded in simulations on SH network RST. Clusters are prioritized for seeding by their median k_{ic}, as described in the text. [Parameters: $f_{sk,c} = 0.75$, $\delta = 0.005$, 50 simulations per data point.]

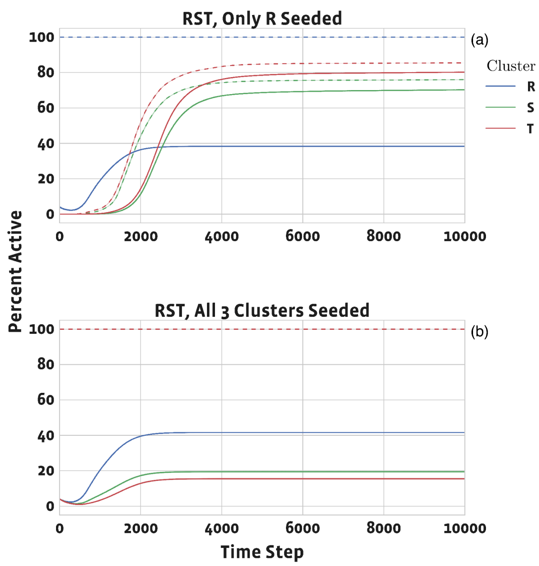


Fig. 7. Average time series of the percentage with access (dashed lines) and the active percentage (solid lines) in simulations on SH network RST. The cluster with the highest median k_{ic} is seeded in (a), and all three clusters are seeded in (b). [Parameters: $p_0 = 0.04$, $f_{sk,c} = 0.75$, $\delta = 0.005$, 50 simulations per time series.]

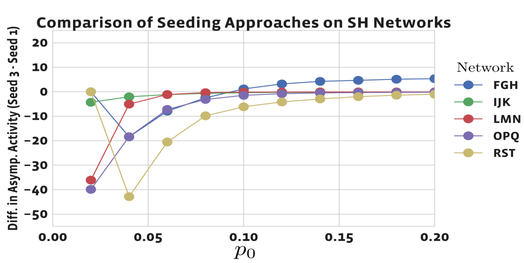


Fig. 8. Difference in asymptotic activity under two seeding strategies (seed one cluster at $t = 0$ vs. seed three clusters) vs. initial activity probability p_0 in simulations on three SH-cluster networks. [Parameters: $f_{sk,c} = 0.75$, $\delta = 0.005$, 50 simulations for one-cluster strategy and 50 simulations for three-cluster strategy per data point.]

lead to more of these events than are strictly necessary. Our simulations suggest that relatively simple gradual seeding approaches can do dramatically better[7].

Although we have demonstrated the costs of overambitious seeding only in a simplified model of social-product usage, we feel that the core assumptions of our model are very plausibly relevant to real-world scenarios. The "need for social support" (assumption 1 in the Introduction) is, almost by definition, a feature of social products. It also seems reasonable that people will adjust their rate of usage up or down in some way depending upon their satisfaction (assumption 2) and that people will churn if they are consistently dissatisfied (assumption 3)[8].

Future work can probe the generality of our results by relaxing or changing some of the modeling assumptions. For example, in our model, new people are given access through the hard-coded gradual access expansion rule, triggered by the activity of their friends in the social network. A different modeling choice, which is well-motivated by real-world scenarios, is for access to expand via invitations sent by current users to their friends in the social network. Another fruitful direction would be to study what happens when the *objective* changes: although IM usually focuses on maximizing long-term usage, the goals of real-life product deployment scenarios can be considerably more nuanced (e.g., to maximize another notion of consumer satisfaction). Finally, we have also assumed that people care about the *number* of their friends who are active. What would happen if people care about having a structurally diverse set of active friends [17]? Or, if we consider products that enable *asynchronous* communication, what would happen if people care about the amount of content produced by their friends in the last n time steps? Would there still be costs to overambitious seeding in all of these scenarios?

Although we leave a definitive answer to that question for future research, we conclude by connecting our study to recent work by Sela et al. that may show a similar phenomena in a rather different model. These authors study product adoption through an SIR model, where a person transitions from the influential (I) to non-influential (R) state a fixed time after adoption. When there is a seeding budget b and people are prioritized for seeding by eigenvector centrality, the final adoption rate is non-monotonic in the budget b (a phenomenon Sela et al. call the "flip anomaly") [14]. A plausible explanation for the "flip anomaly" is the context in which the seeded people adopt. If a seeded person is the only adopting friend in a non-seeded person's local network, then the non-seeded person may not adopt before the seeded person becomes non-influential. In this way, overambitious seeding can incur downstream costs, as in our model. However, Sela et al. note that their "flip anomaly" must reverse as the budget grows, because adoption is universal [14]. This is also true of other models with similar properties that have recently been reviewed by Centola [3]. Our results show how an analogue of the flip anomaly of Sela et al. can still persist with universal adoption and no seeding budget.

[7] For other recent explorations of the advantages of gradual seeding, see references [6] and [14].

[8] Although it is definitely questionable whether churn is ever completely irreversible.

Acknowledgements: We thank Udi Weinsberg and Israel Nir for helpful discussions, Shuyang Lin for development of the original simulation infrastructure, and Justin Cheng for reviewing code.

References

1. Abebe, R., Adamic, L., Kleinberg, J.: Mitigating overexposure in viral marketing. In: Proceedings of the 32nd Conference on Artificial Intelligence. AAAI (2018)
2. Alkemade, F., Castaldi, C.: Strategies for the diffusion of innovations on social networks. Computational Economics **25**(1-2), 3–23 (2005)
3. Centola, D.: How Behavior Spreads: The Science of Complex Contagions, vol. 3. Princeton University Press (2018)
4. Fortunato, S., Hric, D.: Community detection in networks: A user guide. Physics Reports **659**, 1–44 (2016)
5. Hagberg, A., Swart, P., S Chult, D.: Exploring network structure, dynamics, and function using networkx. Tech. rep., Los Alamos National Lab.(LANL), Los Alamos, NM (United States) (2008)
6. Jankowski, J., Bródka, P., Kazienko, P., Szymanski, B.K., Michalski, R., Kajdanowicz, T.: Balancing speed and coverage by sequential seeding in complex networks. Scientific reports **7**(1), 891 (2017)
7. Juul, J.S., Porter, M.A.: Hipsters on networks: How a small group of individuals can lead to an anti-establishment majority. arXiv preprint arXiv:1707.07187 (2017)
8. Kabiljo, I., Karrer, B., Pundir, M., Pupyrev, S., Shalita, A.: Social hash partitioner: a scalable distributed hypergraph partitioner. Proceedings of the VLDB Endowment **10**(11), 1418–1429 (2017)
9. Kempe, D., Kleinberg, J., Tardos, É.: Maximizing the spread of influence through a social network. In: Proceedings of the ninth ACM SIGKDD international conference on Knowledge discovery and data mining, pp. 137–146. ACM (2003)
10. Kiesling, E., Günther, M., Stummer, C., Wakolbinger, L.M.: Agent-based simulation of innovation diffusion: a review. Central European Journal of Operations Research **20**(2), 183–230 (2012)
11. Kim, J.H., Vu, V.H.: Generating random regular graphs. In: Proceedings of the 35th Annual ACM symposium on Theory of Computing, pp. 213–222. ACM (2003)
12. Li, Y., Fan, J., Wang, Y., Tan, K.L.: Influence maximization on social graphs: A survey. IEEE Transactions on Knowledge and Data Engineering (2018)
13. Moore, C., et al.: The computer science and physics of community detection: Landscapes, phase transitions, and hardness. Bulletin of EATCS **1**(121) (2017)
14. Sela, A., Shmueli, E., Goldenberg, D., Ben-Gal, I.: Why spending more might get you less, dynamic selection of influencers in social networks. In: Science of Electrical Engineering (ICSEE), IEEE International Conference on the, pp. 1–4. IEEE (2016)
15. Shalita, A., Karrer, B., Kabiljo, I., Sharma, A., Presta, A., Adcock, A., Kllapi, H., Stumm, M.: Social hash: An assignment framework for optimizing distributed systems operations on social networks. In: NSDI, pp. 455–468 (2016)
16. Steger, A., Wormald, N.C.: Generating random regular graphs quickly. Combinatorics, Probability and Computing **8**(4), 377–396 (1999)
17. Ugander, J., Backstrom, L., Marlow, C., Kleinberg, J.: Structural diversity in social contagion. Proceedings of the National Academy of Sciences **109**(16), 5962–5966 (2012)