

Network Seeding

Arun Sundararajan¹

NYU Stern School of Business

Extended Abstract for WISE 2006

1 Summary

This research studies the optimal seeding of a network of consumers who are potential adopters of a technology. This problem has a wide range of applications which include launching a network good, inducing network-based marketing, igniting collaboration in organizations, and stimulating R&D with local spillovers. I have two main results so far. The first result shows that when a firm is unaware of how connected its consumers' are, randomly seeding a network is almost always optimal. The second result provides a simple and plausible set of conditions under which, when a firm that is aware of how connected its consumers are, it is optimal for the firm to "seed the fringes rather than the hubs", that is, focus on seeding the *least connected* consumers, and not the highly connected consumers. These results hold for any distribution over any arbitrary set of asymmetric networks, and are also independent of the strength of the network effects. My ongoing research focuses primarily on extending these results to include seeding with probabilistic adoption, costly and constrained seeding, myopic consumers, and expanding the model to permit the social network to be a basis for the diffusion of product information.

2 Overview and outline of model

When launching a network good, it often seems sensible for a seller to "seed" its network by giving away the product for free to a subset of its potential customers. If the good displays local network effects – that is, rather than being influenced by an increase in the size of a product's user base or network in general, if each consumer is influenced directly by the decisions of only a typically small subset of other consumers, for instance those he or she is "connected" to via an underlying social or business network – then the seller may wish to seed consumers based on how connected they are. Targeting more highly connected consumers seems like a sensible strategy, since these consumers are more influential – they influence and increase the willingness-to-pay of a larger subset of others. However, there is a trade-off: consumers who are more connected are also likely to have a higher eventual (post-seeding) willingness to pay for the network good, and are also more likely to be influenced by the seeding of others, on account of having a higher number of neighbors.

¹I thank seminar participants at Carnegie-Mellon University for their feedback.

Although I focus on the specific problem of launching a new network good, the framework I introduce for modeling how to seed a network has a wide range of applications. Many economic outcomes are influenced by the structure of an underlying social, business, employee or technological network. This has been recognized by many recent papers in IS economics (examples include Aral, Brynjolfsson and Van Alstyne, 2005 and Sundararajan, 2006; a more detailed survey is available on request). Firms and policy makers can influence outcomes by "seeding" this network: most commonly, to alter incentives by subsidizing a subset of agents connected by the underlying network. Contexts include building awareness about a product or topic using network-based marketing, optimizing the performance of a technology based on a peer-to-peer network, and igniting collaborative innovation or R&D within an organization or a specific geographic region. Seeding can either be targeted, based on knowledge the external entity (firm, policy maker) has about the connectedness or network position of an agent, or can be random.

I investigate the problem of optimal network seeding using the following model. There is a monopoly seller of a good with local network effects. There is a set of n consumers $N = \{1, 2, \dots, n\}$. Each consumer's willingness to pay for the good is influenced by whether or not a subset of others (its neighbors) adopt the good. The structure of adoption complementarities is described by an underlying social network, modeled as an undirected graph. This graph is represented using a symmetric matrix $g \in \{0, 1\}^{n \times n}$, with $g(i, j) = 1$ signifying that consumers i and j are connected. The set of neighbors of i is $G_i = \{j | g(i, j) = 1\}$, and the degree of consumer i is $d_i = |G_i|$. Information about the underlying network is represented by its degree distribution q , where $q(x)$ is the expected fraction of consumers who have degree x . The related neighbor degree distribution,

$$\hat{q}(x) = \frac{xq(x)}{\sum_{i=1}^m iq(i)},$$

which specifies the probability that an arbitrarily chosen neighbor of a consumer has degree x .

There are two periods. In the first period, the seller chooses a subset of consumers to seed. These consumers are given a copy of the good for free, and seeding of this kind imposes no direct variable costs on the seller. If the seller is unaware of the degree of its consumers, then its seeding strategy is a scalar value $z \in [0, 1]$, the fraction of consumers it will give away the product to. If the seller is aware of the degree of each of its consumers, then its seeding strategy is the vector $z = [z(1), z(2), \dots, z(m)]$, where $z(i)$ is the fraction of consumers of degree i who are seeded, and m is the highest possible degree. I assume that even if the seller knows the degree of each of its consumers, it does not know the exact structure of the underlying network.

In the second period, the seller chooses a uniform price, those consumers who have not been seeded become aware of whether each of their neighbors has been seeded or not, and these unseeded consumers play a network game with incomplete information in which each consumer chooses

whether or not to adopt the product. Denoting the action of consumer i as $a_i \in \{0, 1\}$ where $a_i = 1$ signifies adoption, the payoff to consumer i from adopting is

$$\pi(a_i, a_{-i}) = a_i[v(\sum_{j \in G_i} a_j) - p],$$

where $v(x)$ is the value obtained from the product if x of a consumer's neighbors adopt. Demand is measured by the number of consumers who adopt in the greatest symmetric Bayes-Nash equilibrium of the second-period game (which I subsequently refer to as the *adoption game with seeding*).

3 Baseline results

Once the seeding strategy is chosen and implemented, each consumer i knows their own degree d_i , as well as the realized number k_i of these neighbors who have been seeded. The strategy chosen by a consumer in any symmetric Bayes-Nash equilibrium in pure strategies is therefore a function of both d_i . The following baseline results characterize properties of such equilibria.

Lemma 1 *Each symmetric Bayes-Nash equilibrium of the adoption game with seeding has a threshold strategy of the form:*

$$s(d, k) = \begin{cases} 0, & k < k^*(d) \\ 1, & k \geq k^*(d) \end{cases}$$

and is monotonic in both d and k .

Lemma 2 *These equilibria can be strictly Pareto-ranked based on the neighbor adoption probability associated with each equilibrium, which specifies the probability that an arbitrary unseeded neighbor will adopt, given that all unseeded consumers are playing the strategy $s(d, k)$ associated with the equilibrium.*

Lemmas 1 and 2 simplify the analysis of the "demand" phase of the model considerably, making the characterization of optimal seeding tractable. There are a number of other baseline results which describe other interesting properties of the adoption game with seeding, which are omitted due to constraints on space.

4 Optimal seeding strategies

My first result analyzes optimal seeding with an uninformed seller – that is, when the seller does not know the degree of any of the consumers, but knows the degree distribution q . Consider the greatest equilibrium of the unseeded adoption game, that is, when $z = 0$, and associated with the

price p_0^* that maximizes profits in the absence of seeding. Let's say that a network displays full adoption without seeding. If all consumers adopt under this equilibrium.

Proposition 1 *If the seller does not know the degree of its consumers, and the network does not display full adoption without seeding, then it is always optimal for the seller to randomly seed a positive fraction of consumers*

That is, suppose the seller chooses not to seed any consumer, and chooses its optimal price in the absence of seeding. Either all consumers adopt, or a subset of consumers do not adopt. The result shows that in the latter case, the seller can always increase its profits by choosing a random subset of consumers and giving them a free copy of the product *prior* to offering the product at a positive price. This is a sufficient rather than necessary condition for random seeding to be profit-improving. It is possible that random seeding benefits the seller even when all consumers would have adopted in its absence, but this is not true in general.

My second result shows that when the seller is aware of how connected each consumer is (that is, the seller knows the degree of each consumer), then its optimal seeding strategy is to seed all consumers of the *lowest* degree, upto a threshold.

Proposition 2 *If the seller knows the degree of each consumer, then, for every degree distribution $q(x)$ and value function $v(x)$, its optimal seeding strategy is of the form:*

$$z(d) = \begin{cases} 1, & d < d^* \\ 0, & d > d^* \end{cases}$$

This is a counter-intuitive finding: it is never profit-improving for the seller to target more highly connected consumers in favor of those who are less connected. The intuition for this result is based on two independent reasons – a direct effect and an informational effect. First, any seeding strategy *directly* affects the willingness to pay of a fraction of unseeded consumers, by making it certain that one or more of their neighbors owns the product, while simultaneously eliminating as potential buyers those consumers who are seeded. A higher-degree consumer is more likely to be an adopter in the second period because every symmetric equilibrium of the adoption game played by unseeded consumers has strategies that are monotone in degree. Thus, the fraction of potential adopters who are eliminated is lower when the seeding strategy is biased towards lower degree consumers. Further, if the degree of consumers who are unseeded is higher on average, the seller can charge a higher price. This effect must be traded off with the fact that seeding a higher degree consumer increases the willingness to pay of a larger group of unseeded consumers.

The second reason is related to the extent to which the seller's seeding strategy *informs* non-seeded consumers about the degree of their non-seeded neighbors. Suppose, for instance, the seller

seeds consumers randomly. This would not change the posterior a non-seeded consumer has about the degree of its non-seeded neighbors. However, as the seller biases its seeding strategy more towards lower degree consumers, it becomes more likely that a non-seeded neighbor has a higher degree. For any fixed number of seeded neighbors, this higher posterior increases the willingness to pay of each non-seeded consumer, since it increases their assessment of how likely each of their non-seeded neighbors is to adopt, thus increasing demand at any price.

5 Research in progress

My research in progress has two primary directions. The first is towards extending the current model and examining how sensitive its results are to changes in its assumptions. There are three such extensions in progress:

(1) Costly and/or constrained seeding: If the variable cost of seeding a consumer is high enough, the result of Proposition 2 may change. On the other hand, a high enough variable cost may cause seeding to become suboptimal in general. The focus of this extension is to determine whether there is a range of variable costs over which seeding the lowest degree consumers continues to be optimal.

(2) Probabilistic seeding: Often, a seeded consumer does not actually use the free product. The focus of this extension is towards understanding whether the results of Propositions 1 and 2 persist if seeding is imperfect in this way. My preliminary analysis suggests that Proposition 1 continues to hold, and progress on extending Proposition 2 remains in progress.

(3) Myopic consumers: The proof of the result of Proposition 2 relies on the assertion that consumers are rational enough to choose their actions according to a symmetric Bayes-Nash equilibrium of the adoption game with seeding. My third extension relaxes this assumption of perfect rationality, assuming instead that consumers are myopic and base their actions exclusively on how many of their neighbors have been seeded. My analysis thus far suggests that the result of Proposition 2 does not always hold under this extension.

The second direction is towards expanding the model to include information transmission via the underlying social network. That is, the current model assumes that all consumers are aware of the product's existence, and seeding merely affects their expected value from adopting it. However, in reality, many firms seed consumers with new products to increase awareness of the product in general. This is a significant expansion of the scope of the model described. My intuition is that it would lead to a wider range of cases under which the converse of Proposition 2 is true. I hope to report on further progress by the time WISE is held in December 2006.