Maximizing Cascades in Social Networks: An Overview

Aron Culotta culotta@cs.umass.edu
Computer Science Department
University of Massachusetts
Amherst, MA, 01002 USA

Abstract

Cascades are a network phenomenon in which small local shocks can result in wide-spread fads, strikes, innovations, or power failures. Determining which initial shocks will result in the largest cascades is of interest to marketing, epidemiology, and computer networking. This paper surveys work in maximizing the size of cascades in social networks.

1. Introduction

In his 1903 opus entitled "The Laws of Imitation," pioneering sociologist Gabriel Tarde set out "to learn why, given one hundred different innovations conceived of at the same time – innovations in the form of words, in mythological ideas, in industrial processes, etc. – ten will spread abroad, while ninety will be forgotten" [11]. The optimist's first answer is that a good innovation will spread by its own merit. But, as is apparent to anyone who has ever appreciated watching a movie on Betamax or has witnessed the illogical outcomes of rioting, merit is not the only catalyst of the spread of an idea. Tarde attributes these irrational occurrences to "extra-logical influences," resulting in situations where "the poorest innovations, from the point of view of logic, are selected because of their place, or even date of birth."

Current research in social network analysis asserts that these "extra-logical" influences can be explained by examining the dynamics of the network through which influence is transmitted between individuals. In other words, if we view individuals as nodes in a social network, where a directed edge indicates that one node influences another, then some graph configurations make it more likely that an innovation will be widely adopted. Moreover, in some graphs, if only a small proportion of individuals adopt an innovation, the idea will spread quickly to a large component of the graph. This is the phenomenon of "global cascades" [12], also known as "information cascades" [13], "diffusion of innovations" [10], or "spread of influence" [6].

Given such a system, a natural question emerges: If we wish to maximize the size of a cascade, which subset of nodes should we target? [3] Answers to this question have applications in marketing, politics, economics, epidemiology, and computer networking.

2. Modeling Cascades

2.1. Local Influence

Let G be a directed graph representing a social network, where each node can be *active* (adopter of an idea) or *inactive*, and a directed edge represents the influence of one node upon another. Cascade models attempt to explain how nodes become active. The underlying assumption in all cascade models is that the probability that a node v will become active increases monotonically with the number of active neighbors of v.

Sociologist Mark Granovetter [5] proposed that each node v has a threshold such that if the fraction of v's neighbors that are active exceeds v's threshold, then v will become active. This model has been generalized by [12] and others to the following *Linear Threshold Model*: Let $0 \leq \theta_v \leq 1$ be the threshold of node v. Let A_v be the subset of nodes that are neighbors of v and are active. For every node $w \in A_v$, $b_{w,v}$ defines the influence neighbor w has on v, where $\sum_{w \in A_v} b_{w,v} \leq 1$. Then v becomes active if $\sum_{w \in A_v} b_{w,v} \geq \theta_v$.

Thus, θ_v represents the tendency for v to be swayed by the actions of its neighbors – in the language of Tarde, the "coefficient of imitation" [11].

Kempe et al.[6] describe an alternate cascade model called the *Independent Cascade Model*. Imagine simulating the spread of an innovation. At time t, let node v become active. Then for each w that neighbors v, let $p_{v,w}$ be the probability that v activates w. If

v succeeds, then w will be active for the duration of the simulation. The model also specifies that v cannot attempt to activate w at any other time step in the simulation. The system is parameterized by $p_{v,w}$, and the assumption is that these probabilities are independent of time.

2.2. Global Cascades

Both of these cascade models are undoubtedly oversimplifications of the actual interactions occurring at the local level, but the hope is that they will provide insight into global phenomena. Specifically, global cascades in the real world (e.g. power failures and stock market crashes) possess two notable traits: (a) they occur infrequently, and (b) when they do occur, they affect a large portion of the network (by definition) [12].

Watts examined cascades in random graphs using the Linear Threshold Model [12]. He characterizes two constraints on the emergence of global cascades. First, when the network is sparsely connected, "the propagation of cascades is limited by the global connectivity of the network." In other words, even if θ_v is very small, most innovations do not spread because nodes do not interact enough to influence each other. On the other hand, when the network is dense, θ_v is the limiting factor. For a given θ_v, the more neighbors v has, the less likely it is that v will become active.

3. Viral Marketing

Businesses strive to make advertising more cost effective by judiciously choosing which potential customers to target. Recent developments in data mining have allowed companies to identify likely consumers based on demographic information [8]. This idea is extended in [4] to consider the "lifetime value" of a customer. Thus, decisions to target a customer are based not only on how likely that customer is to purchase the product, but also the expected revenue that customer will bring the company over his entire life. Also, Chickering and Heckerman differentiate potential customers based on how they will respond to advertising. Some customers will buy the product with no advertising (*always-buy*), others will buy only if given a discount (*persuadable*), and still others will buy only if they are *not* marketed to (*anti-persuadable*) [2].

None of the above methods considers the network effects of a potential customer. Yet, as long ago as 1969, Frank Bass modeled a customer's initial purchase of an invention (e.g. the refrigerator, air conditioner, or lawn-mower) as a function of the number of previous buyers [1]. To estimate the number of previous buyers, he separates customers into two types, innovators and imitators, where an imitator's decision to purchase an invention is highly influenced by the number of previous buyers.

Extending this to social networks, Krackhardt [7] considers the local influence each customer exerts on her neighbors. In what he calls the "friendship ripple effect," a customer who receives a free trial of a product coaxes her friends into purchasing the product. But Krackhardt confines the ripple to friends one edge away from the targeted customer.

Viral marketing ([3],[9]) generalizes Krackhardt's model by considering the entire network effect of targeting a customer. Here, we look not only at the number of friends a targeted customer will coax into purchasing a product, but also at the friends of her friends, and so on. Businesses must now consider two factors when targeting a consumer: (1) How likely is she to purchase a product (an intrinsic property), and (2) how many additional people will she likely coax into buying the product (a network property). In the language of cascade models, which nodes should businesses attempt to activate to maximize the size of the resulting cascade?

4. Maximizing Cascade Size

4.1. Case 1: General Threshold model

We follow the notation of Richardson and Domingos to setup the problem [9]. For a set of n potential customers, let X_i be a Boolean variable that is 1 if customer i buys the product being marketed, and is 0 otherwise. Assume the product being sold has a set of attributes $\mathbf{Y} = \{Y_1, \ldots, Y_m\}$. Let the neighbors of X_i be those nodes that directly influence X_i, which we denote $\mathbf{N_i}$. Let M_i be a Boolean variable that is 1 if we market to customer X_i, and is 0 otherwise. Previous research has ignored network effects and simply targeted customers most likely to purchase a product, as defined by $P(X_i = 1|\mathbf{Y}, M_i)$ [8]. But, taking a customer's network value into account, we instead model $P(X_i = 1|\mathbf{Y}, M_i, \mathbf{N_i})$. An optimal marketing plan can be viewed as the solution to $\operatorname*{\mathbf{argmax}}_{\mathbf{M}} \sum_i P(X_i = 1|\mathbf{Y}, M_i, \mathbf{N_i})$.

However, to maximize profits, we must also factor in the cost of $\mathbf{M}$. Let r_0 be the revenue from selling a product to a customer with no marketing, and r_1 be the revenue if marketing is used. (E.g. if marketing includes giving the customer a discount, then $r_1 < r_0$.) Let the constant c be the cost of marketing to any customer. Also let $f_i^1(\mathbf{M})$ be the result of setting M_i to 1 and leaving the rest of $\mathbf{M}$ unchanged, defined similarly for $f_i^0(\mathbf{M})$. Then we can define the *expected lift in profit* from marketing to customer i in isolation as

$$\begin{aligned} ELP_i^1(\mathbf{Y},\mathbf{M}) &= r_1 P(X_i = 1|\mathbf{Y}, f_i^1(\mathbf{M}), \mathbf{N_i}) \\ &\quad - r_0 P(X_i = 1|\mathbf{Y}, f_i^0(\mathbf{M}), \mathbf{N_i}) - c \end{aligned}$$

Domingos calls this a customer's *intrinsic value*. If we let $\mathbf{M_0}$ be a vector of zeros, then the global lift in profit from marketing plan $\mathbf{M}$ is

$$\begin{aligned} ELP(\mathbf{Y},\mathbf{M}) &= \sum_{i=1}^{n} r_i P(X_i = 1|\mathbf{Y},\mathbf{M},\mathbf{N_i}) \\ &\quad - r_0 \sum_{i=1}^{n} P(X_i = 1|\mathbf{Y},\mathbf{M_0},\mathbf{N_i}) \\ &\quad - |\mathbf{M}|c \end{aligned}$$

A customer's *total value* is the ELP obtained by marketing to her. Therefore, a customer's *network value* is the difference between her total and intrinsic values.

We wish to find an assignment of $\mathbf{M}$ that maximizes ELP. Consider a simulation of the network. Let A be the initial set of active nodes, and define $\sigma(A)$ to be the set of active nodes at the end of simulation. If we temporarily ignore discounts (i.e. $r_0 = r_1$), the problem of maximizing ELP for a given marketing budget of size $c|A|$ is equivalent to finding the set of nodes of size $|A|$ that maximizes $\sigma(A)$, i.e. for a fixed size A, find $\mathbf{argmax}_A\, \sigma(A)$ [6].

First, however, it is important to note that the model defined above is more general than the Linear Threshold model defined in section 2. Here, we simply condition X_i on $\mathbf{N_i}$, but we do not place linear constraints on the relationship between X_i and $\mathbf{N_i}$. Thus, we have assumed that X_i's decision to become active is based on some monotonic threshold function $f(\mathbf{N_i})$ that is a mapping from $\mathbf{N_i}$ to a real number in [0,1]. The Linear Threshold model is a special case where $f(\mathbf{N_i}) = \sum_{X_j \in \mathbf{N_i}} b_{j,i} X_j$, and $\sum_{X_j \in \mathbf{N_i}} b_{j,i} \le 1$.

Kempe et al. show that it is NP-hard to approximate the influence maximization problem under the General Threshold model to within a factor of $n^{1-\epsilon}$, for any $\epsilon > 0$ [6]. Despite this limitation, Domingos and Richardson achieve reasonable results by performing the following greedy hill-climbing algorithm to find the optimal $\mathbf{M}$: while there exists a marketing action that increases ELP, set $M_j = 1$, where marketing to customer j maximizes ELP.

4.2. Case 2: Linear Threshold model

Realizing the intractability of the General Threshold model, Richardson and Domingos reformulated the cascade maximization problem in terms of the Linear Threshold model. Here, we simplify by letting $P(X_i|\mathbf{Y}, M_i, \mathbf{N_i}) =$

$$\beta_i P_0(X_i|\mathbf{Y}, M_i) + (1-\beta_i) P_N(X_i|\mathbf{Y},\mathbf{M},\mathbf{N_i})$$

where

$$P_N(X_i = 1|\mathbf{Y},\mathbf{M},\mathbf{N_i}) = \sum_{X_j \in \mathbf{N_i}} b_{j,i} X_j$$

Here, $0 \le \beta_i \le 1$ measures how "self-reliant" X_i is, P_0 is X_i's *internal* probability of buying the product, and P_N defines the influence of X_i's neighbors on his decision [9].

With this simplification, the optimal setting of M can be found analytically by solving a system of linear equations [9].

4.3. Case 3: Triggering Model

Kempe et al. specify a model that is more general than the Linear Threshold model but less general than the General Threshold model [6]. In the *Triggering Model*, each node X_i independently chooses a random "triggering set" T_i according to some distribution over $\mathbf{N_i}$. At the start of simulation, set A is targeted for initial activation. At each step, X_i is set to 1 if there exists an $X_j \in T_i$ such that $X_j = 1$.

We can think of T_i as defining "live" and "blocked" edges. Edge (i, j) is live if $X_j \in T_i$, and is blocked otherwise. Finding $\sigma(A)$ is therefore equivalent to the graph reachability problem over live edges. The set of nodes activated by A is the set of all nodes reachable from nodes in A using "live" edges. By iteratively simulating this process, we can estimate $\sigma(A)$ for a given $\mathbf{M}$. Under this model, it can be shown that a greedy hill-climbing algorithm for setting $\mathbf{M}$ is within 63% of optimal [6].

5. Extensions and Future Work

Richardson and Domingos extend their models by letting $\mathbf{M}$ be a continuous variable, so that businesses can allocate marketing funds more effectively [9]. We can also extend the simulations to the *non-progressive* case, in which nodes can go from *active* to *inactive* as well as vice versa. Kempe et al. have shown the progressive case to be equivalent to the non-progressive case [6].

Future work could include examining (1) the effects of network evolution on cascades, (2) the implications of allowing marketing interventions at different time steps, and (3) the impact of adversarial markets (where a competitor's marketing strategy must be considered).

References

[1] F. Bass. A new product growth for model consumer durables. *Management Science*, 15(5):215–227, 1969.

[2] D. M. Chickering and D. Heckerman. A decision theoretic approach to targeted advertising. In *Proceedings of the Sixteenth Annual Conference on Uncertainty in Artificial Intelligence*, Stanford, CA, 2000.

[3] P. Domingos and M. Richardson. Mining the network value of customers. In *Proceedings of the Seventh International Conference on Knowledge Discovery and Data Mining*, pages 57–66, San Francisco, CA, 2001.

[4] K. Gelbrich and R. Nakhaeizadeh. Value miner: A data mining environment for the calculation of the customer lifetime value with application to the automotive industry. In *Proceedings of the Eleventh European Conference on Machine Learning*, pages 154–161, Barcelona, Spain, 2000.

[5] M. Granovetter. Threshold models of collective behavior. *American Journal of Sociology*, 83(6):1420–1433, 1978.

[6] D. Kempe, J. Kleinberg, and E. Tardos. Maximizing the spread of influence through a social network. In *Proc. 9th ACM SIGKDD Intl. Conf. on Knowledge Discovery and Data Mining*, 2003.

[7] D. Krackhardt. Structural leverage in marketing. In D. Iacobucci, editor, *Networks in Marketing*, pages 50–59, Thousand Oaks, CA, 1996. Sage.

[8] Charles X. Ling and Chenghui Li. Data mining for direct marketing: Problems and solutions. In *Knowledge Discovery and Data Mining*, pages 73–79, 1998.

[9] M. Richardson and P. Domingos. Mining knowledge-sharing sites for viral marketing. In *Eighth Intl. Conf. on Knowledge Discovery and Data Mining*, 2002.

[10] E. Rogers. *The Diffusion of Innovation*. Free Press, New York, 1995.

[11] G. Tarde. *The Laws of Imitation*. Henry Holt, New York, 1903.

[12] D. Watts. A simple model of global cascades in random networks. In *Proc. Natl. Acad. Sci*, pages 5766–71, 2002.

[13] D. Watts. *Six Degrees: The Science of a Connected Age*. W.W. Norton, New York, 2003.