An Integrated Framework for Competitive Multi-channel Marketing of Multi-featured Products

Swapnil Dhamal

An Integrated Framework for Competitive Multi-channel Marketing of Multi-featured Products

Swapnil Dhamal

INRIA Sophia Antipolis Méditerranée, France

Email: swapnil.dhamal@inria.fr

Abstract—For any company, multiple channels are available for reaching a population in order to market its products. Some of the most well-known channels are (a) mass media advertisement, (b) recommendations using social advertisement, and (c) viral marketing using social networks. The company would want to maximize its reach while also accounting for simultaneous marketing of competing products, where the product marketings may not be independent. In this direction, we propose and analyze a multi-featured generalization of the classical linear threshold model. We hence develop a framework for integrating the considered marketing channels into the social network, and an approach for allocating budget among these channels.

Index Terms—Social networks, viral marketing, mass media, social advertisement, budget allocation

I. INTRODUCTION

Companies can market their respective products through several possible channels, the most prominent being mass media advertising (television, radio, newspapers, etc.), fixed Internet banner ads, banner ads based on browsing history, recommendations based on attributes (location, age group, field of work, etc.), recommendations based on friends' purchases, sponsored search ads, social media, sponsored Internet reviews, and so on. Potential customers or nodes also get indirectly influenced through their friends owing to word-of-mouth marketing. In order to make optimal use of these channels, a company would want to make the decision of how to invest in each channel, based on the investment strategy of competitors who also market their products simultaneously.

This paper aims to present a framework for competitive influence maximization in the presence of several marketing channels. We focus on modeling three channels, namely, viral marketing, mass media advertisement, and recommendations based on friends' purchases using social advertisement.

1) Viral Marketing: In our context, a social network can be represented as a weighted, directed graph, consisting of nodes which are potential customers. The model we propose for influence diffusion in social network is a generalization of the well-studied linear threshold (LT) model [1]. Given such a model, a company would want to select a certain number of seed nodes to trigger viral marketing so that maximum number of nodes get influenced (buy the product) [2].

2) Mass media advertisement: This is one of the most traditional way of marketing where a company advertises its product to the masses using well-accessible media such as television, radio, and newspaper. The timing of when to show the ads is critical to ensure optimal visibility and throughput.

3) Social advertisement based on friends' purchases: While making purchasing decisions, nodes rely not only on their own preferences but also on their friends', owing to social correlation due to homophily (bias in friendships towards similar individuals) [3]. This, in effect, can be harnessed to suggest products to a node based on its friends' purchasing behaviors. If a node has high influence on its friend (which is accounted for in diffusion models like LT), it is likely that the two nodes are similar. However, if the influence is low, it is not conclusive whether the nodes are dissimilar. So in addition to the influence parameter considered in LT-like models, marketing in practice requires another parameter that quantifies similarity between nodes. Note that since diffusion models do not consider this similarity, they alone cannot justify why two nodes having negligible influence on each other, display similar behaviors. It has also been observed in Twitter that almost 30% of information is attributed to factors other than network diffusion [4]. The effect of such factors could hence be captured using the similarity parameter.

Related Work. The problem of influence maximization is well-studied in literature on social network analysis. It is known that computing the value of the objective function for a given seed set (the expected number of influenced nodes at the end of diffusion that was triggered at that set), is #P-hard under LT model [5]. However, the value can be well approximated using sufficiently large number of Monte-Carlo simulations. Though the influence maximization problem under LT model is NP-hard, the objective function is non-negative, monotone, and submodular; so greedy hill-climbing algorithm provides an approximation guarantee for its maximization [2]. There exist generalizations of LT model, e.g., general threshold model [2], extensions to account for time [6], and extensions to account for competition [7], [8]. State-of-the-art heuristics such as LDAG [5] and Simpath [9] perform close to greedy algorithm while running several orders of magnitude faster. There also exist algorithms that provide good performance irrespective of the objective function being submodular [10].

The problem of competitive influence maximization wherein multiple companies market competing products using viral marketing has also been studied [11], [12]. Also, more realistic models have been developed, where influences not only diffuse simultaneously but also interact with each other [13], [14].

The impact of recommendations and word-of-mouth marketing on product sales revenue is well studied in marketing

b_{uv}	influence weight of node u on node v
$\mathcal{N}(v)$	set of influencing neighbors of v
h_{uv}	similarity between nodes u and v
χ_v	threshold of node v
γ^p	total budget for the marketing of product p
k^p	budget for seed nodes for viral marketing of product p
β_t^p	mass media advertising weight of product p in time step t
α^p	social advertising weight of product p
$\mathcal{A}_v$	final aggregate preference of node v
$\mathcal{P}_v$	product bought by node v
$\mathcal{B}(t)$	set of nodes influenced by time t

Fig. 1. Geometric interpretation of the proposed model

literature [15]–[17]. Biases in product valuation and usage decisions when agents consider a product that offers new features of uncertain value, have been investigated [18]. It has also been discussed how marketers can apply latent similarities of customers for segmentation and targeting [19].

II. THE PROPOSED FRAMEWORK

We propose a framework to facilitate study of different marketing aspects using a single model, capturing several factors:
1) Companies market their products using multiple channels;
2) Diffusions of different products are mutually dependent;
3) Each node aggregates the mass media advertisements, recommendations, and neighbors' purchasing decisions.

We first describe LT model, followed by our competitive multi-feature generalization, and then integration of other channels into this generalized model. Table I presents notation.

In LT model, every directed edge (u, v) has weight $b_{uv} \geq 0$, which is the degree of influence that node u has on node v, and every node v has an influence threshold χ_v. The weights b_{uv} are such that $\sum_u b_{uv} \leq 1$. Owing to thresholds being private information to nodes, they are assumed to be chosen uniformly at random from $[0, 1]$. The diffusion process starts at time step 0 with the initially activated set of seed nodes, and proceeds in discrete time steps. In each time step, a node gets influenced if and only if the sum of influence degrees of the edges incoming from its already influenced neighbors crosses its influence threshold, that is, $\sum_u b_{uv} \geq \chi_v$. The process stops when no further nodes can be influenced. Formally, let $u \in \mathcal{N}(v)$ if and only if $b_{uv} \neq 0$. Let $\mathcal{B}(t)$ be the set of nodes influenced by time t. Then

$$v \in \mathcal{B}(t) \backslash \mathcal{B}(t-1) \text{ iff } \sum_{\substack{u \in \mathcal{N}(v) \\ u \in \mathcal{B}(t-2)}} b_{uv} < \chi_v \text{ and } \sum_{\substack{u \in \mathcal{N}(v) \\ u \in \mathcal{B}(t-1)}} b_{uv} \geq \chi_v \quad (1)$$

A. Competitive Multi-feature Generalization of LT Model

Products these days, be they toothpastes or mobile phones, come with several features with different emphases on different features. Let such an emphasis be quantified by a real number between 0 and 1. That is, a product can be represented by a vector of mutually independent features $p = (p_1, \cdots, p_f)$, where $p_i \in [0, 1]$. Let the features of each product be suitably scaled such that $||p|| = 1$. Note that such scaling may not be feas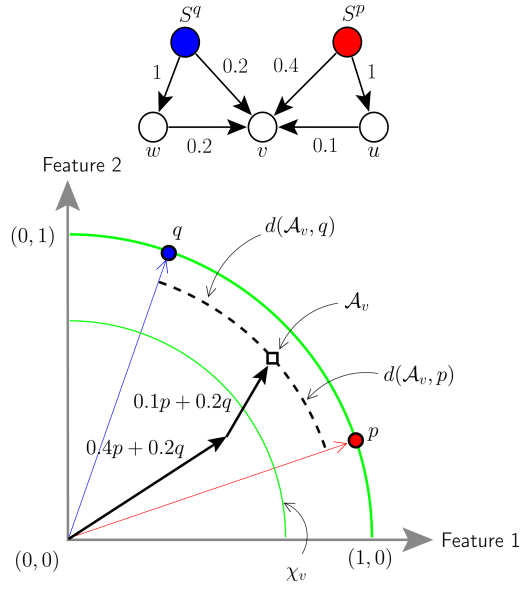ible when there is a product p which offers strictly better features than product q ($\forall i : p_i > q_i$); so let the products be such that one of the features corresponds to 'null'. So p would have a lower null component as compared to q, thus making the scaling feasible (a higher null component would imply that the product has a poorer feature set).

Our model for a node getting influenced is analogous to that used in classical mechanics to study the initial motion of a body placed on a rough horizontal surface, as a result of several forces acting on it. In our context, a node is analogous to the body, and its threshold is analogous to static frictional force stopping it from moving. Such a force is equal to $\mu_s mg$, where m is mass of the body, g is acceleration due to gravity, and μ_s is the coefficient of static friction between the body and surface. For $\mu_s g = 1$ unit for all nodes, the frictional force and analogously, the threshold equals mass, which is chosen uniformly at random from $[0, 1]$ (as assumed in the LT model).

A node v gets influenced in time step t when the net force on it crosses its threshold value χ_v; let the net force correspond to aggregate vector (say $\mathcal{A}_v$). Let $d(\mathcal{A}_v, p)$ be the angular distance between $\mathcal{A}_v$ and the product vector p. Since $||p|| = 1$,

$$d(\mathcal{A}_v, p) = \arccos\left(\frac{\mathcal{A}_v \cdot p}{||\mathcal{A}_v||}\right)$$

A node buys a product whose angular distance from its aggregate vector is the least (it can be easily shown that such a product would have the least Euclidean distance as well). If there exist multiple such products, one of them is chosen uniformly at random.

Hence the competitive multi-feature version of (1) is

$$v \in \mathcal{B}(t) \backslash \mathcal{B}(t-1) \, , \, \mathcal{A}_v = \sum_{u \in \mathcal{N}(v)} b_{uv} \mathcal{P}_u \, , \, \mathcal{P}_v = \arg \min_p d(\mathcal{A}_v, p)$$

$$\text{iff } \left\| \sum_{\substack{u \in \mathcal{N}(v) \\ u \in \mathcal{B}(t-2)}} b_{uv} \mathcal{P}_u \right\| < \chi_v \text{ and } \left\| \sum_{\substack{u \in \mathcal{N}(v) \\ u \in \mathcal{B}(t-1)}} b_{uv} \mathcal{P}_u \right\| \geq \chi_v$$

A geometric interpretation of the proposed model is presented in Figure 1. Consider 2 competing products having 2 features, say $p = (p_1, p_2), q = (q_1, q_2)$. In time step 0, S_p and S_q are selected for seeding by products p and q, respectively. In time step 1, node v aggregates the purchasing decisions of its neighbors S^p and S^q, hence obtaining the aggregate vector $0.4p + 0.2q$. Say $\sqrt{(0.4p_1 + 0.2q_1)^2 + (0.4p_2 + 0.2q_2)^2} < \chi_v$, so v is not influenced yet. However, u and w purchase products p and q respectively (since the influence weights from S^p to u and S^q to w are 1). Hence in time step 2, node v aggregates the purchasing decisions of u and w, hence obtaining the aggregate vector $\mathcal{A}_v = (0.4p + 0.2q) + (0.1p + 0.2q)$. Say $||\mathcal{A}_v|| = \sqrt{(0.5p_1 + 0.4q_1)^2 + (0.5p_2 + 0.4q_2)^2} \geq \chi_v$, so v is now influenced and it purchases product p if $d(\mathcal{A}_v, p) < d(\mathcal{A}_v, q)$, q if $d(\mathcal{A}_v, p) > d(\mathcal{A}_v, q)$, else it chooses randomly.

B. Properties of the Generalized LT Model

The standard LT model is a special case of the proposed model, where there is a single product with one feature, i.e., $p = (1)$. As the problem of influence maximization in the standard LT model is NP-hard, we have that the problem of influence maximization in the proposed model is also NP-hard.

We now explain the multi-feature (vector-based) model with an illustrative example, which will also throw light on the properties of the objective function under the proposed model. Recollect that the threshold for any node is chosen uniformly at random from $[0, 1]$. Let $\mathbb{P}_v^p(S^p, S^q)$ be the probability that node v gets influenced by product p when S^p and S^q are selected for seeding by p and q respectively. In Figure 2, let the 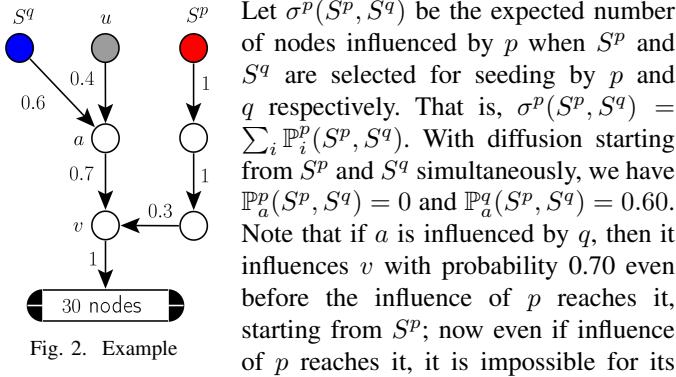two products be $p = (1, 0)$ and $q = (0, 1)$. Let $\sigma^p(S^p, S^q)$ be the expected number of nodes influenced by p when S^p and S^q are selected for seeding by p and q respectively. That is, $\sigma^p(S^p, S^q) = \sum_i \mathbb{P}_i^p(S^p, S^q)$. With diffusion starting from S^p and S^q simultaneously, we have $\mathbb{P}_a^p(S^p, S^q) = 0$ and $\mathbb{P}_a^q(S^p, S^q) = 0.60$. Note that if a is influenced by q, then it influences v with probability 0.70 even before the influence of p reaches it, starting from S^p; now even if influence of p reaches it, it is impossible for its aggregate preference to be closer to p than to q. So node v can get influenced by p only if a is not influenced by q. So $\mathbb{P}_v^p(S^p, S^q) = 0.3(1 - \mathbb{P}_a^q(S^p, S^q)) = 0.12$. So all 30 nodes which have v as sure influencer get influenced by p with probability 0.12. Hence $\sigma^p(S^p, S^q) = 1 + 2 + 0.12(1 + 30) = 6.72$.

Now if the seed set for p is $T^p = S^p \cup u$, $\mathbb{P}_a^p(T^p, S^q) = 0$ due to an incoming edge of 0.6 from S^q (it is impossible for the aggregate preference of node a to be closer to p than to q). However, $\mathbb{P}_a^q(T^p, S^q) = \sqrt{0.6^2 + 0.4^2} \approx 0.72$. From the argument similar as above, $\mathbb{P}_v^p(T^p, S^q) = 0.3(1 - \mathbb{P}_a^q(T^p, S^q)) < 0.084$. So the 30 nodes get influenced by p with probability less than 0.084. Hence $\sigma^p(T^p, S^q) < 2 + 2 + 0.084(1 + 30) < 6.61$. That is, $\sigma^p(T^p, S^q) < \sigma^p(S^p, S^q)$.
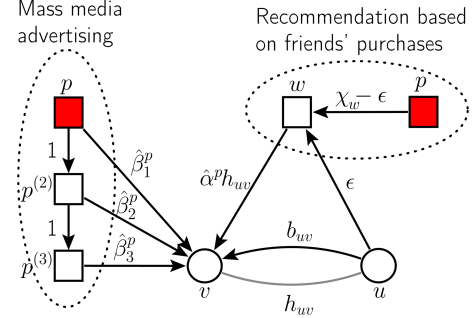

Fig. 2. Example

Fig. 3. Integration of mass media and social advertisements into the network

Thus while the objective function in the standard LT model follows monotone increasing property, adding a node to a set in the generalized model could decrease its value; this proves non-monotonicity. It can also be shown using counterexamples that $\sigma^p(\cdot)$ is neither submodular nor supermodular.

C. Integrating Mass Media & Social Advertising into Network

Let β_t^p be the investment for mass media advertising of product p in time step t and β^p be the total investment over time T, that is, $\beta^p = \sum_{t=1}^T \beta_t^p$. For social advertising, we consider that a company would recommend or advertise product p to a node when any of its friends u has bought the product. Let α^p be the effort invested in social advertising. Let h_{uv} (or h_{vu}) be the parameter that quantifies the similarity between nodes u and v. So $\alpha^p h_{vu}$ could be viewed as the influence of such a recommendation on v owing to the purchase of product p by u. Since the total influence weight allotted by node v for viral marketing is $\sum_{u \in \mathcal{N}(v)} b_{uv}$, the total weight that it can allot for other channels is $1 - \sum_{u \in \mathcal{N}(v)} b_{uv}$. Hence the weights allotted for other channels ($(\beta_t^p)_{t=1}^T$ and α^p) would be scaled accordingly to obtain the values of $(\hat{\beta}_t^p)_{t=1}^T$ and $\hat{\alpha}^p$ specific to node v. A simple scaling rule is:

$$\frac{\hat{\beta}_t^p}{\beta_t^p} = \frac{\hat{\alpha}^p}{\alpha^p} = \frac{1 - \sum_{u \in \mathcal{N}(v)} b_{uv}}{\sum_p \left(\sum_{u \in \mathcal{N}(v)} \alpha^p h_{uv} + \sum_{t=1}^T \beta_t^p \right)}$$

In order to integrate mass media and social advertisements into the network, we add pseudonodes and pseudoedges corresponding to them, as illustrated in Figure 3. Pseudonode p corresponds to the product company itself (the figure shows two separate copies of pseudonode p for the two channels for better visualization; they are the same pseudonode). Pseudonode p and all seed nodes selected for viral marketing, are influenced in time step 0.

For integrating mass media advertisement, we create a set of pseudonodes $\{p^{(t)}\}_{t=1}^T$ (where $p^{(1)}$ corresponds to pseudonode p), and pseudoedges $\{(p^{(t-1)}, p^{(t)})\}_{t=2}^T$ of weight 1. Hence $p^{(t)}$ gets influenced with probability 1 in time step $t - 1$ (see Figure 3). We further create pseudoedges $\{(p^{(t)}, v)\}_{t=1}^T$ for node v such that $b_{p^{(t)}v} = \hat{\beta}_t^p$. Since $p^{(t)}$ gets influenced in time $t - 1$, node v receives influence of $\hat{\beta}_t^p$ from pseudonode $p^{(t)}$ in time step t; this is equivalent to mass media advertisement.

For integrating social advertisement, corresponding to edge (u, v), we create an intermediary pseudonode w with a fixed threshold $\chi_w > 0$, and pseudoedges such that $b_{uw} = \epsilon \in$

$(0, \chi_w), b_{pw} = \chi_w - \epsilon, b_{wv} = \hat{\alpha}^p h_{uv}$ (see Figure 3). Now if the reference friend u is influenced by some product q, where the angle between products p and q be θ, the intermediary pseudonode w gets influenced if and only if $||(\chi_w - \epsilon)p + \epsilon q|| \geq \chi_w$. Since $||p|| = ||q|| = 1$, this is equivalent to

$$(\chi_w - \epsilon)^2 + \epsilon^2 + 2\epsilon(\chi_w - \epsilon)\cos\theta \geq \chi_w^2$$
$$\Longleftrightarrow 2\epsilon(\chi_w - \epsilon)(\cos\theta - 1) \geq 0$$
$$\Longleftrightarrow \theta = 0 \quad (\because \epsilon < \chi_w)$$
$$\Longleftrightarrow q = p \quad (\because ||p|| = ||q||)$$

So w gets influenced if and only if u buys product p, after which v is recommended to buy p with influence weight $\hat{\alpha}^p h_{uv}$. Also note the time lapse of one step between the reference friend u buying the product and the target node v receiving the recommendation. Hence the latency in recommendation using social advertising is implicitly accounted for.

III. THE UNDERLYING PROBLEM

The fundamental problem here is to distribute the total available budget among the three marketing channels under study. Let k^p be the number of free samples of product p that the company would be willing to distribute. Let S^p be the corresponding set of nodes in the social network to whom free samples would be provided ($|S^p| = k^p$). Let $c_p(\cdot)$ be the cost function for allotting effort of activating set S^p to trigger viral marketing, α^p for social advertising, and $(\beta_t^q)_{t=1}^T$ corresponding to each step of mass media advertising. In general, $c_p(\cdot)$ would be a weighted sum of these parameters since the costs for adjusting parameters corresponding to different channels would be different. Let γ^p be the total budget for marketing of product p. Let $\nu^p(\cdot)$ be the expected number of nodes (excluding pseudonodes) influenced by product p, accounting for the marketing strategies of p and its competitors. Hence, the optimization problem for the marketing of product p is

Find $S^p, \alpha^p, (\beta_t^p)_{t=1}^T$ to maximize
$\nu^p(S^p, \alpha^p, (\beta_t^p)_{t=1}^T, (S^q, \alpha^q, (\beta_t^q)_{t=1}^T)_{q \neq p})$
such that $c^p(S^p, \alpha^p, (\beta_t^p)_{t=1}^T) \leq \gamma^p$

In the above optimization problem, we not only need to determine the optimal allocation among channels (k^p, β^p, α^p), but the best k^p nodes to trigger viral marketing S^p such that $|S^p| = k^p$, and the optimal allocation over time for mass media advertising $(\beta_t^p)_{t=1}^T$ such that $\sum_{t=1}^T \beta_t^p = \beta^p$. The problem hence demands a method for multi-parametric optimization.

Methods such as Fully Adaptive Cross Entropy (FACE) provide a simple, efficient, and general approach for simultaneous optimization over several parameters [20]. In our context, the FACE method involves an iterative procedure where each iteration consists of two steps, namely, (a) generating data samples according to a specified distribution and (b) updating the distribution based on the sampled data to produce better samples in the next iteration. Here, our sample is a vector consisting of whether a node should be included in S^p, budget allotted for each time step of mass media advertising $(\beta_t^p)_{t=1}^T$,

and budget allotted for social advertising α^p; each data sample satisfies the cost constraint $c^p(S^p, \alpha^p, (\beta_t^p)_{t=1}^T) \leq \gamma^p$. Initially, the data samples could be generated based on a random distribution. The value of the objective function $\nu^p(\cdot)$ is computed for each data sample as per the proposed model using a sufficiently large number of Monte Carlo simulations. The distribution is then updated by considering data samples which provide value of the objective function better than a certain percentile. This iterative updating continues until convergence or for a fixed number of iterations. The obtained terminal data sample would act as a best response allocation strategy for product p, in response to the strategies of competitors. Using such best response dynamics, the possible equilibria of this generalized marketing game could be determined.

A Note for Practical Implementation: As a company would need to map its customers to the corresponding nodes in social network, it would be useful to obtain its online social networking identity (say Facebook ID) when it buys the product. This could be done using a product registration website (say for activating warranty) where a customer, when it buys the product, is required to login using a social networking website, or provide its email address which could help in discovering its online social networking identity. The time step when the node has bought the product, can thus also be obtained.

REFERENCES

[1] D. Easley and J. Kleinberg, *Networks, Crowds, and Markets: Reasoning About a Highly Connected World*. Cambridge University Press, 2010.

[2] D. Kempe, J. Kleinberg, and É. Tardos, "Maximizing spread of influence through a social network," in *SIGKDD*. ACM, 2003, pp. 137–146.

[3] F. Chua, H. Lauw, and E.-P. Lim, "Generative models for item adoptions using social correlation," *IEEE Transactions on Knowledge and Data Engineering*, vol. 25, no. 9, pp. 2036–2048, 2013.

[4] S. Myers, C. Zhu, and J. Leskovec, "Information diffusion and external influence in networks," in *SIGKDD*. ACM, 2012, pp. 33–41.

[5] W. Chen, Y. Yuan, and L. Zhang, "Scalable influence maximization in social networks under the linear threshold model," in *ICDM*. IEEE, 2010, pp. 88–97.

[6] W. Chen, W. Lu, and N. Zhang, "Time-critical influence maximization in social networks with time-delayed diffusion process." in *AAAI*. AAAI, 2012, pp. 592–598.

[7] A. Borodin, Y. Filmus, and J. Oren, "Threshold models for competitive influence in social networks," in *WINE*. Springer, 2010, pp. 539–550.

[8] N. Pathak, A. Banerjee, and J. Srivastava, "A generalized linear threshold model for multiple cascades," in *ICDM*. IEEE, 2010, pp. 965–970.

[9] A. Goyal, W. Lu, and L. V. S. Lakshmanan, "SIMPATH: An efficient algorithm for influence maximization under the linear threshold model," in *ICDM*. IEEE, 2011, pp. 211–220.

[10] R. Narayanam and Y. Narahari, "A Shapley value-based approach to discover influential nodes in social networks," *IEEE Transactions on Automation Science and Engineering*, vol. 8, no. 1, pp. 130–147, 2011.

[11] S. Bharathi, D. Kempe, and M. Salek, "Competitive influence maximization in social networks," in *WINE*. Springer, 2007, pp. 306–311.

[12] S. Goyal and M. Kearns, "Competitive contagion in networks," in *STOC*. ACM, 2012, pp. 759–774.

[13] S. Myers and J. Leskovec, "Clash of contagions: Cooperation & competition in information diffusion," in *ICDM*. IEEE, 2012, pp. 539–548.

[14] A.Zarezade, A.Khodadadi, M.Farajtabar, H.Rabiee, and H.Zha, "Correlated cascades: Compete or cooperate." in *AAAI*, 2017, pp. 238–244.

[15] D. Godes, "The strategic impact of references in business markets," *Marketing Science*, vol. 31, no. 2, pp. 257–276, 2012.

[16] R. Van der Lans, G. Van Bruggen, J. Eliashberg, and B. Wierenga, "A viral branching model for predicting the spread of electronic word of mouth," *Marketing Science*, vol. 29, no. 2, pp. 348–365, 2010.

[17] S. Aral and D. Walker, "Creating social contagion through viral product design: A randomized trial of peer influence in networks," *Management Science*, vol. 57, no. 9, pp. 1623–1639, 2011.

[18] R. Meyer, S. Zhao, and J. Han, "Biases in valuation vs. usage of innovative product features," *Marketing Science*, vol. 27, no. 6, pp. 1083–1096, 2008.

[19] M. Braun and A. Bonfrer, "Scalable inference of customer similarities from interactions data using dirichlet processes," *Marketing Science*, vol. 30, no. 3, pp. 513–531, 2011.

[20] P.-T. De Boer, D. P. Kroese, S. Mannor, and R. Y. Rubinstein, "A tutorial on the cross-entropy method," *Annals of operations research*, vol. 134, no. 1, pp. 19–67, 2005.