A Generalized Fractional Program for Maximizing Content Popularity in Online Social Networks

Alexandre Reiffers-Masson[1,2], Yezekael Hayel[2], Eitan Altman[3,2] and Guillaume Marrel[2]

Abstract—In this paper, we consider a "generalized" fractional program in order to solve a popularity optimization problem in which a source of contents controls the topics of her contents and the rate with which posts are sent to a time line. The objective of the source is to maximize its overall popularity in an Online Social Network (OSN). We propose an efficient algorithm that converges to the optimal solution of the Popularity maximization problem.

I. INTRODUCTION

As noticed in [8], one key challenge on Online Social Networks (OSNs) is to find an answer to the following question: how to become popular on an OSN? In the context of marketing and sales, the importance for a company to be popular on OSNs is of main interest. Indeed, in a recent report (http ://www.socialmediaexaminer.com), the number-one challenge marketers want to see resolved (92%) is: **"What social tactics are most effective?"**. Many empirical studies have been developed to answer this issue [3], [12], [18], [19]. An optimization approach has been proposed in the case where a marketer, also called a source in our paper, will pay the OSNs to get his content promoted [10], [5], [9], [13]. In this paper, we propose a novel approach, where a source controls the flow of posted contents and the topics associated to each content.

In order to provide an answer to this challenge of popularity in OSNs, our methodology pursues the following steps. We start by understanding how content popularity behaves on OSNs. A vast amount of literature about this question exists (for instance [19], [12] and references therein). In this paper, we assume that the popularity of a content is highly related to its position in a News Feed. This assumption is satisfied in Facebook/Twitter-style content posting system. In Facebook for example, a *News Feed* is a feed where the user's friends and subscribed pages' posts are published. These posts can be displayed in a top-down anti-chronological order from the newest to the oldest if you use the most recent option. More explanations about the News Feed dynamics and models are provided in [1]. Then, we assume that the maximization of content popularity is similar to the oil producer problem [17]. This problem assumes that an oil producer would pump oil from a single well and after some time, when the return on the well decreases, he would abandon that source and start exploiting another one. In the case of OSNs, we assume that a well pumping is equivalent to a content staying in the first position of the News Feed. At some point, the popularity return will decrease and it is recommended to send a new content, meaning to exploit a new source. Our model of popularity is expressed as a system of discrete differential equations. Finally, we study the maximization of content popularity. We assume that a company (or source) that tries to be popular will focus on two types of control: the number of posted contents and the topics of its contents. A continuous formulation of the popularity maximization problem is proposed and we prove its equivalence with a "generalized" fractional program.

The main contributions of this paper are as follows: First we formally describe the *popularity maximization problem*, an advertising optimization problem on online social networks. We also define pertinent popularity functions. We use the data to estimate the parameters of these popularity functions and assesses the quality of their correlation level with real data. Then we prove that for standard popularity functions, we can use convex optimization to design efficient algorithms. Finally, we list some of the constraints that generalized the popularity maximization problem and that can be expressed as convex inequalities.

Related works: Several publications have appeared in recent years documenting the maximization of popularity in online social networks. Usually there are two approaches. The first one is advertising allocation where a source decides the amount of money he is willing to pay to the OSN in order to get its content promoted. Papers about this subject are often linked to optimal control [9], [10]. The second approach adopted by researchers is to model the propagation of information through a cascade model in a social network. Then the question would be to find which node a source would need to target in order to maximize the spread of its contents [11], [4].

We propose in this paper to define a new approach based on flow control. In this framework the amount of contents is sent by the source affects their popularity. To the authors' best knowledge, very few publications can be available in the literature that address this issue. In [1], [14] the authors have used a Poisson process to model the flow of contents on a unique news feed and define a visibility measure of a content. The major drawback of this approach is that the authors are concerned with visibility and not popularity, so the measure is not realistic. Our visibility measure is more general. Moreover, they do not list all the possible constraints that can occur in a popularity maximization problem and only deals with a simple optimization problem.

Organization of the paper. In section III, we propose a mathematical model of the popularity. The section IV is

[1]Indian Institute of Science, India
[2]University of Avignon, France
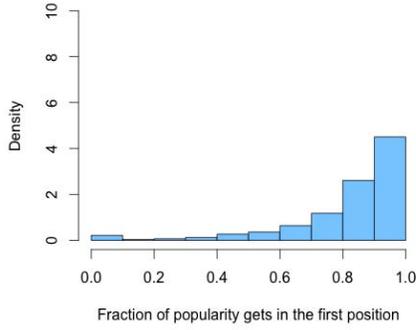[3] INRIA Sophia Antipolis, France

Fig. 1: Histogram of first position fraction of popularity

about the Popularity maximization problem. First, we provide its definition, and then we prove, for standard popularity functions, its equivalence with a pseudoconcave optimization problem. Finally, an algorithm that converges to the optimal solution is provided.

II. DATA EXPERIMENT: IMPORTANCE OF THE FIRST POSITION

In this section, we develop an empirical experiment to illustrate how the popularity of a content is related to the position of a content in the News Feed. We inspect specifically the relationship between the popularity (in terms of number of comments) and the duration a content spends in the first position in the News Feed. Our dataset is composed of contents posted on a specific Facebook page associated to a french political personality (Anne Hidalgo https://www.facebook.com/HidalgoAnne Paris mayor). We have extracted 892 contents and 39083 associated comments with timestamps using the Netvizz app [15]. The dataset was collected from June 15 2013 to December 01 2014. The histogram depicted on fig. 1 represents the proportion (or density) of contents and the relative proportion of popularity obtained in the first position in the News Feed. For example, this histogram shows that close to 50% of the contents get 90% of their popularity while being in the first position in the News Feed. Also, it shows that more than 70% of the posted contents reached 80% of their total popularity while the content is in the first position in the News Feed. Therefore, it makes sense to consider a popularity model, described in detail in the next section, by assuming that almost all the popularity of a content is obtained while the content is in the first position in the News Feed.

III. POPULARITY MODEL

The level of popularity associated with a topic can be defined as the total number of comments, given by OSN subscribers, of all contents related to this topic. We model the dynamic of the popularity of each topic through a discrete time dynamical system.

Let $c \in \mathscr{C} := \{1, \dots C\}$ be a topic index. Topics can be "Sport", "Arts", "Entertainment", etc. Note that our model is scalable and can be applied to any number of topics studied.

The random state variable $X_c(t) \in \mathbb{N}_+$ is the total number of comments that gets all content related to topic c in a News Feed, at a given time $t \in \mathbb{R}_+$. For each topic $c \in \mathscr{C}$, the arrival process of a type c content in the News Feed is a Poisson Point Process with intensity $0 < \lambda_c < \infty$. All arrival processes are independent. We also consider that there exists a flow, with intensity σ of posts which are not related to any topics. Let $t_m \in \mathbb{R}_+$ be the random variable denoting the arrival instant of the mth content, with $m \in \mathbb{N}_+^*$. As the arrival processes are independent Poisson Point Processes, the inter-arrival time denoted by Δ_m between consecutive contents arrived at t_m and t_{m-1}, follows an exponential distribution with rate $\Lambda := \sum_{d=1}^{C} \lambda_d$. Therefore, the probability associated with the inter arrival time Δ_m of two consecutive contents is given by:

$$\forall m, \, \delta \in \mathbb{R}_+, \, P(\Delta_m < \delta) = \int_0^\delta \Lambda e^{-\Lambda t} dt. \quad (1)$$

A content can also be called a post on an OSNs. We define by $\eta(m) \in \mathscr{C}$ the topic of the mth content. According to [7], for each content m, $\eta(m)$ is a random variable with the following distribution:

$$\forall m, \; P(\eta(m) = c) = \frac{\lambda_c}{\Lambda}. \quad (2)$$

The evolution of the cumulative number of comments associated with topic c in a News Feed, is modeled by the following dynamical system:

$$X_c(t_m) = X_c(t_{m-1}) + 1_{\{\eta(m-1)=c\}} Y_c(\Delta_m), \quad (3)$$
$$X_c(t_1) = 0, \quad (4)$$

where the random variable $Y_c(\delta) \in \mathbb{N}_+$, with $\delta \in \mathbb{R}_+$, describes the cumulative popularity obtained by a content of type c that stays δ units of time on the first position of the News Feed. The indicator $1_{\{\eta(m-1)=c\}} = 1$ if the $(m-1)^{th}$ content is of type c, it is equal to 0 otherwise. We assume that for each c, the random variable $Y_c(\delta)$ is described by a non-homogeneous Poisson process with intensity $E[Y_c(\delta)] := \int_0^\delta f_c(\tau) d\tau \geq 0$. The density function $f_c(.)$ are called popularity functions. In point of fact, in recent studies on analyzing social networks, non-homogeneous Poisson processes have been used to model and predict the popularity of contents. For instance, in [6], [16], the authors suggest to use Hawks processes in order to capture the evolution of the number of likes, comments of contents posted in online social networks. Hawks processes are composed of an exogenous and an endogenous part. The exogenous part corresponds to $E[Y_c(\delta)]$ and the usual assumption about the shape of popularity functions is a uniform, an exponential or a power law distribution. Therefore, our model can be seen as the worst-case of the hawks model since the endogenous part is equal to zero.

From data experiments, we have been able to show that three types of popularity functions $f_c(.)$ have significant correlation levels with data. The first function is the constant function $f_c(\tau) = M_c$, the second function is the linear function $f_c(\tau) = a_c \tau + b_c$ and the third function is the Michaelis-Menten model which is given by $f_c(\tau) = \frac{a_c b_c}{(b_c+\tau)^2}$. M_c is the parameter of the
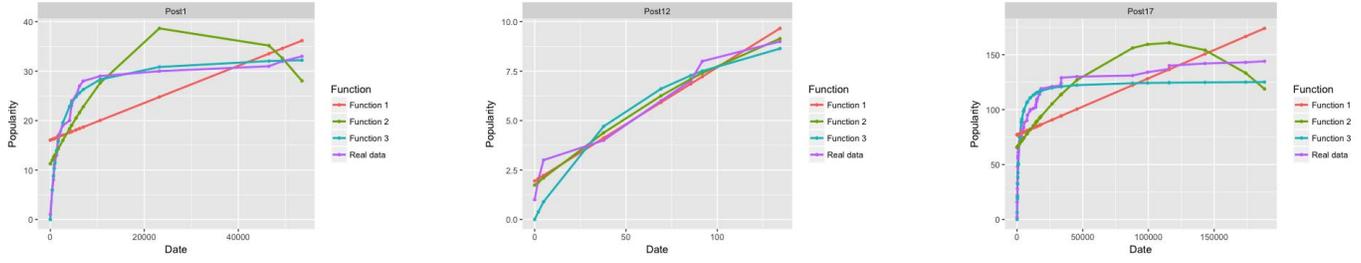
Fig. 2: Estimation of popularity functions for three posts.

first function and a_c and b_c are the parameters of the second and third functions. Examples of estimations are depicted in Fig. 2. The levels of correlation are described in Fig. 3 for three different posts. We observe that for the type 3 function the average level of correlation is 0.95 which is very significant. And even for the simplest constant function 1 $f_c(\tau) = M_c$, the level of correlation has an average of 0.7. Moreover the median of the p-value associated to each curve is below 1e-05 (Function 1: 4.25e-06, Function 2: 1.09e-10, Function 3: 1.07e-17).
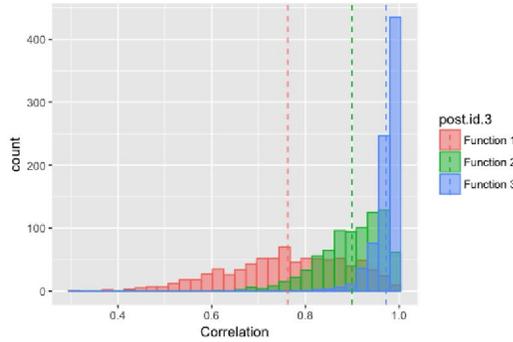

Fig. 3: Level of correlation for the three functions: function 1 $f_c(\tau) = M_c$, function 2 $f_c(\tau) = a_c\tau + b_c$ and function 3 $f_c(\tau) = \frac{a_c b_c}{(b_c + \tau)^2}$.

A closed form expression of the expected cumulative popularity for each topic c at any given time, is given in the following proposition.

Proposition 1: The expected cumulative popularity for each topic c at the time t_m when the mth content is posted is given by:

$$\forall c \in \mathscr{C}, t_m \in \mathbb{R}_+^*, \quad E[X_c(t_m)] = m \int_0^\infty \lambda_c e^{-\Lambda t} \int_0^t f_c(\tau) d\tau dt. \tag{5}$$

In table I we present mathematical closed form expression of the expected cumulative popularity $E[X_c(t_m)]$ for different shapes of popularity functions of the non-homogeneous Poisson process.

IV. POPULARITY OPTIMIZATION

We now study the case where a single source maximizes the popularity of his content. The source determines the topics of his content and also the rate of sending contents per topic.

$f_c(\tau)$	M_c	$a_c\tau + b_c$	$\frac{a_c b_c}{(b_c+\tau)^2}$
$E[X_c(t_m)]$	$\frac{M_c\lambda_c}{\Lambda}$	$\lambda_c\left(\frac{-2a_c}{\Lambda^3} - \frac{b_c}{\Lambda}\right)$	$a_c\lambda_c\left(\frac{1}{\Lambda} - b_c e^{b_c\Lambda}\Gamma(0,b_c\Lambda)\right)$

TABLE I: Examples of expected cumulative popularity and their closed form for the three different popularity functions.

Therefore, the source parameters are described by a vector $(\lambda_1,\ldots,\lambda_C)$ where $\lambda_c \in [0,\overline{\lambda}]$ is the rate of contents related with topic c and $\overline{\lambda}$ is the maximum rate for each topic. Based on proposition 1, the average expected cumulative popularity X_c^∞ associated with the content related to topic c posted by the source is equal to:

$$\forall c \in \mathscr{C}, \qquad X_c^\infty := \lim_{m\to\infty} \frac{1}{m}E[X_c(t_m)] = \int_0^\infty \lambda_c e^{-\Lambda t} \int_0^t f_c(\tau)d\tau dt. \tag{6}$$

Let $\lambda_{cost} := \sum_{c=1}^C \gamma_c\lambda_c$ be the cost perceived by the source for sending all his contents. This cost is considered to be in monetary units and we assume that the source wants to maximize the sum of its total average expected cumulative popularity minus the cost. Then, the net utility of the source is defined as follows:

$$U(\lambda_1,\ldots,\lambda_C) := \sum_{c=1}^C \int_0^\infty \lambda_c e^{-\Lambda t} \int_0^t f_c(\tau)d\tau dt - \sum_{c=1}^C \gamma_c\lambda_c. \tag{7}$$

The *Popularity Optimization Problem* (PoP) is the following optimization problem defined below:

$$\max_{\lambda_c\in[0,\overline{\lambda}],\,\forall c} U(\lambda_1,\ldots,\lambda_C) \tag{8}$$

A solution of this problem is denoted by a vector $\lambda^{OA} := (\lambda_1^{OA},\ldots,\lambda_C^{OA}) \in \arg\max_{\lambda_c\in[0,\overline{\lambda}],\,\forall c} U(\lambda_1,\ldots,\lambda_C)$. This optimization problem is note straightforward, and the purpose of the rest of this section is to find an equivalent form of the Popularity optimization problem that can be solved efficiently.

The next theorem proves that the optimal vector rate λ^{OA} has a particular form. In fact, optimally, the source has to send content only related to a single and unique topic. Therefore, the Popularity optimization problem can be reduced to a simpler optimization problem.

Proposition 2: If the $f_c(\cdot)$ is one of the standard functions used to model the popularity of a content (see table I), the

Popularity optimization problem is equivalent to the following generalized fractional problem:

$$\max_{c}\left[\max_{\lambda_c\in[0,\bar{\lambda}]} U_c(\lambda_c)\right]. \tag{9}$$

According to [2], optimization problems like the one defined in equation (9) is a pseudo-concave optimization problem. Then, for each $c \in \mathscr{C}$, a projected gradient algorithm, converges to the optimal solution of (9).

We have observed that under verified conditions, the Popularity optimization problem is a generalized fractional problem. Considering a gradient scheme, we are able to propose a global algorithm that first looks for the optimal rate for each topic c and then determines, by comparing all the outputs of the C gradient algorithms, on which topic the source has to send his contents.

We illustrate the convergence of the gradient scheme for the Popularity optimization problem on a numerical example, with $f_c(\tau) = e^{-\alpha_c\tau}$, where the parameters are $C = 3$ and $(\alpha_1, \alpha_2, \alpha_3) = (10, 5, 50)$. On Figure 4, each curve draws the iteration of the gradient algorithm for each type c. We observe the convergence of the gradient in each case as stated previously. Each dashed line shows the optimal flow λ_c^{OA} for each type c. The last iteration of the algorithm implies that the source has to create contents about the second topic, in order to become popular.
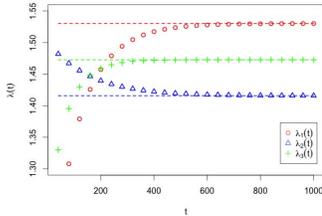

Fig. 4: Convergence of the gradient algorithm.

Convex constraints: Some constraints can be introduced into the generalized the popularity optimization problem and that can be expressed as convex inequalities. For example, a minimum time T can be imposed on the interval between two contents in order to avoid spamming. This type of constraint is captured by $\frac{1}{\lambda} \geq T$. To go further in this direction, we can also assume that we are interested to bound the quantity of contents related to a specific topic sending per unit of time. In this case, the following constraint is modeled by $(\frac{\lambda_c}{\Lambda})^{k_c} \leq d_c$ where k_c is this bound for topic c. Both of these constraints are convex and can be easily integrated into our algorithm. Finally We can impose upper and lower bounds on the intensity ratio between two topics $l_c \leq \frac{\lambda_c}{\lambda_d} \leq u_c$, as a measure of fairness between topics.

V. DISCUSSION AND CONCLUSION

In this paper, we define a popularity optimization problem in online social networks. The model is based on the observation that most of the popularity of a content is obtained during the time a content spent in the first position in a News Feed. We propose an optimization problem where a source determines optimally the rate of his content and also the topics. An equivalent fractional problem is obtained and an efficient gradient-based algorithm is proposed to determine the optimal solution. Finally, we provide to generalize our problem by considering constraints that do not alter the algorithm proposed.

REFERENCES

[1] E. Altman, P. Kumar, S. Venkatramanan, and A. Kumar. Competition over timeline in social networks. In *Advances in Social Networks Analysis and Mining (ASONAM), 2013 IEEE/ACM International Conference on*, pages 1352–1357. IEEE, 2013.

[2] M. Avriel, W. E. Diewert, S. Schaible, and I. Zang. *Generalized concavity*, volume 63. SIAM, 1988.

[3] R. Bandari, S. Asur, and B. A. Huberman. The pulse of news in social media: Forecasting popularity. In *ICWSM*, pages 26–33, 2012.

[4] W. Chen, C. Wang, and Y. Wang. Scalable influence maximization for prevalent viral marketing in large-scale social networks. In *Proceedings of the 16th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 1029–1038. ACM, 2010.

[5] F. De Pellegrini, A. Reiffers, and E. Altman. Differential games of competition in online content diffusion. In *Networking Conference, 2014 IFIP*, pages 1–9. IEEE, 2014.

[6] W. Ding, Y. Shang, L. Guo, X. Hu, R. Yan, and T. He. Video popularity prediction by sentiment propagation via implicit network. In *Proceedings of the 24th ACM International on Conference on Information and Knowledge Management*, pages 1621–1630. ACM, 2015.

[7] R. Durrett. *Essentials of stochastic processes*. Springer Science & Business Media, 2012.

[8] J. Heidemann, M. Klier, and F. Probst. Online social networks: A survey of a global phenomenon. *Computer Networks*, 56(18):3866–3878, 2012.

[9] K. Helmes, R. Schlosser, and M. Weber. Optimal advertising and pricing in a class of general new-product adoption models. *European Journal of Operational Research*, 229(2):433–443, 2013.

[10] J. Huang, M. Leng, and L. Liang. Recent developments in dynamic advertising research. *European Journal of Operational Research*, 220(3):591–609, 2012.

[11] D. Kempe, J. Kleinberg, and É. Tardos. Maximizing the spread of influence through a social network. In *Proceedings of the ninth ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 137–146. ACM, 2003.

[12] H. Lakkaraju, J. J. McAuley, and J. Leskovec. What's in a name? understanding the interplay between titles, content, and communities in social media. *ICWSM*, 1(2):3, 2013.

[13] H. I. Mesak and T. S. Ellis. On the superiority of pulsing under a concave advertising market potential function. *European Journal of Operational Research*, 194(2):608–627, 2009.

[14] A. Reiffers-Masson, E. Altman, and Y. Hayel. A time and space routing game model applied to visibility competition on online social networks. In *International Conference on NETwork Games COntrol and oPtimization 2014 (NetGCoop'14)*, 2014.

[15] B. Rieder. Studying facebook via data extraction: the netvizz application. In *Proceedings of the 5th Annual ACM Web Science Conference*, pages 346–355. ACM, 2013.

[16] M.-A. Rizoiu, L. Xie, S. Sanner, M. Cebrian, H. Yu, and P. Van Hentenryck. Can this video be promoted?-endogenous and exogenous popularity processes in social media. *arXiv preprint arXiv:1602.06033*, 2016.

[17] S. Sethi and G. Thompson. *Optimal Control Theory–Applications to Management Science and Economics*. Kluwer, 2000.

[18] A. Susarla, J.-H. Oh, and Y. Tan. Social networks and the diffusion of user-generated content: Evidence from youtube. *Information Systems Research*, 23(1):23–41, 2012.

[19] G. Szabo and B. A. Huberman. Predicting the popularity of online content. *Communications of the ACM*, 53(8):80–88, 2010.