Posting behavior in Social Networks and Content Active Filtering

Alexandre Reiffers-Masson, Eitan Altman, Yezekael Hayel

Posting behavior in Social Networks and Content Active Filtering

Alexandre Reiffers Masson
INRIA/LIA
INRIA Sophia-Antipolis
2004 Route des Lucioles
06902 Sophia-Antipolis Cedex, France
Email: alexandre.reiffers@inria.fr

Eitan Altman
INRIA/LIA
INRIA Sophia-Antipolis
2004 Route des Lucioles
06902 Sophia-Antipolis Cedex, France
Email: eitan.altman@inria.fr

Yezekael Hayel
LIA
University of Avignon
39 chemin des Meinajaries
Agroparc BP 1228
84911 Avignon cedex 9, France
Email: yezekael.hayel@univ-avignon.fr

***Abstract*—In this paper, we have two objectives: First we model the posting behavior in Social Networks in topics which have negative externalities, and the second objective is to propose content active filtering in order to increase content diversity. By negative externalities, we mean that when the quantity of posted contents about some topic increases the popularity of posted contents decreases. We introduce a dynamical model to describe the posting behavior of users taking into account these externalities. Our model is based on stochastic approximations and sufficient conditions are provided to ensure its convergence to a unique rest point. We provide a close form of this rest point. Content Active Filtering (CAF) are actions taken by the administrator of the Social Network in order to promote some objectives related to the quantity of contents posted in various topics. As objective of the CAF we shall consider maximizing the diversity of posted contents.**

I. INTRODUCTION

Nowadays, Online Social Networks (like Facebook or Twitter) has allowed a large population of users to post contents about different topics and to express different opinions. And with this amazing liberty comes new problematics: Why users of Social Network post? Why they decide to post on a particular topic? etc. Some works in Sociological Science try to find an answer to this question [1] and [2].

The first goal of this paper is to model the posting behavior of users. In our model, users with content in some given topic, connect to the Social Network according to a Poisson Process. Following [1] we assume that there is some self-censorship exercised by the user: only contents who he estimates to be potentially sufficiently popular[1] will be posted. To the best of our knowledge, mathematical models for posting behavior do not involve popularity of contents [2], [3], [1]. We also consider negative externalities between different topics, i.e., the larger the number of contents about topic c is, the smaller is the popularity of a content about any other topics including topic c. Negative externalities may occur for example when content consumers have limited budget of attention [4]. Other typical scenarios are described in the introduction of [5]. In previous works the posting behavior of a publisher in a Social Network was modelled using three main factors: trends, interest of the publisher and of its neighbours [3]. We introduce a model based on stochastic approximations [6] in order to study the evolution of the number of posted contents in various topics taking into account the negative externalities. Sufficient conditions are provided to ensure its convergence to a unique rest point. A close form of this rest point is given and we show that it can be obtained as the unique equilibrium of some non-cooperative game.

The second goal of our paper is to propose content active filtering in order to increase content diversity at the rest point of the stochastic approximation. Content Active Filtering (CAF) are actions taken by the administrator of the Social Network in order to promote some objectives related to the quantity of contents posted in various topics. There are many ways to implement CAF in practice. For example, when a user posts some content in Facebook, the administrator can decide on which wall the content will appear and who will be notified on this post [7]. As objective of the CAF we shall consider maximizing the diversity of posted contents for the following reasons.

We find it desirable that Social Networks provide an access to content of different topics in a diverse way. We call this property content diversity. Such diversity is critical in Political News [8] and [9]. Websites Initiatives (like *politifact.com* or *FactCheck.org*) that try to introduce the same visibility to different opinions. Another context where content diversity is an important issue in Health Information [10] and [11]. In these papers, the authors show that high quality Health information is not easily accessible. If we consider that we are not able to measure the quality of information, a simple solution to solve this quality/accessibility trade off is to increase diversity between information. A similar issue arises when we are interested in decreasing misinformation propagation and rumours in Social Network, [12], [13] and [14].

Organization of the paper. After a short state-of-the-art section, we introduce in section III the posting behavior model which include the negative externalities. We formulate the limit regime for appropriate scaling obtained through stochastic approximations. In the section IV, sufficient conditions are provided to ensure the convergence of the stochastic approximation to a unique rest point, and we provide a closed form expression of this rest point. In section V, we define content diversity and we propose a Content Active Filtering with a closed form. Finally, in section VII we give some conclusions and perspectives of the paper.

II. RELATED WORKS

Many recent works study empirically the behaviour of users in Social Networks, for example [15] and [16]. In [15], empirical analysis are proposed to the study connectivity properties and users' activities on different Social Networks,

[1]Standard measures of popularity are the number likes, views, comments, shares, etc...

in particular on Facebook, Twitter and Google+. The authors notice that in Facebook and Google+, there are more creation of messages than reshares (to relay a received post), and that the opposite scenario occurs in Twitter. In [16], the authors consider a temporal context-aware mixture model to describe the behaviour of users. They assume that topics of interest for each user are related to their own *intrinsic interest* and a global *temporal context*. The authors show, considering a Social Network dataset, that their model is more realistic than other models from the literature.

The question of visibility and popularity of contents in Social Networks are recently subjects of interest from a theoretical perspective point of view. For example, in [17] the authors consider a dynamical model to describe the popularity of various contents in Youtube. Each publisher can increase the popularity of its content in Youtube to get views by paying a cost for advertisement. The authors provide an exact optimal policy of investment for publishers. In [18], the authors propose a model in which several publishers maximize their visibility in a News Feed. This work assumes that more a message stays on the top first messages of a News Feed, more the popularity of this message increases. The goal of this work is to understand how to model the dynamics of the message visibility and to understand what are the consequences of a competition over visibility between different publishers.

Negative externalities are well-known phenomena in diffusion process [19] and [20]. The model proposed in [19] describes a situation in which several users of a Social Network have to decide to adopt one technology among several. In a marketing context, companies by using targeting strategies decide which users receive ads and therefore optimize their profit. In this paper, the authors provide a mathematical framework in a competition scenario between several companies. In [20], the authors discuss about seeding competition, where firms need to decide which users will be the first to be contaminated. The goal of this paper is to provide solutions of the seeding competition and also to quantify the inefficiency of the competition over firms.

To the best of our knowledge, there is no theoretical works related to improvement of content diversity in Social Networks. In social sciences, there exist some studies that try to design tools in order to increase diversity in political opinions [21] or to improve civil discourses [22]. In [21] the authors propose to use a widget[2] that shows to readers their reading diversity. A search engine algorithm is proposed in [21] to increase content diversity.

III. EVOLUTION OF POSTING BEHAVIOR

The Social Network consists of large number of users. We denote by $t_n \in \mathbb{R}_+$ the arrival instant of the n^{th} user. The n^{th} new incoming user is associated with a content of a particular topic $c \in \mathscr{C} := \{1,\dots,C\}$ and with a potential level of popularity $Z_c(n) \in \mathbb{R}_+$. We assume that there is no distinction between users that are interested in the same topic c. $\mathscr{C}$ is the set of all possible topics in the social network. For any content topic c, the arrival rate of users is an independent Poisson point process with intensity $\lambda_c \in \mathbb{R}_+$. Let $\xi(n) \in \mathscr{C}$ the random variable that determines the topic of the n^{th} new arrival user. The probability that the n^{th} arrival user has topic c content is given by [23]:

$$P(\xi(n) = c) = \frac{\lambda_c}{\sum_{c'} \lambda_{c'}}.$$

For any topic c, let $y_c(n)$ denote the total number of posts on topic c in the Social Network during $[0,t_n]$. For each topic c, let $x_c(n) := \frac{y_c(n)}{n}$ be the average number of posts about topic c. We denote the vector $\mathbf{x}(n) = (x_1(n),\dots,x_C(n))$. The scalar product of two vectors $\mathbf{x}$ and $\mathbf{y}$ is denoted by $< x,y >$. The transpose of a matrix A is denoted A^T.

Popularity and decision to post Each arrival user decides to post his content of topic c if the potential level of popularity of his content (i.e. number of likes, comments, etc.) exceeds a threshold $\theta_c \in \mathbb{R}_+$. We assume that each user has enough experience to predict the popularity of his content. The random variables $Z_c(n)$ are independent. The probability that the n^{th} user posts his content given that it is of topic c is given by

$$P(Z_c(n) \geq \theta_c \mid \xi(n) = c, \mathbf{x}(n)) := f_c(x_1(n),\dots,x_C(n),\theta_c).$$

Note that for any topic c, the popularity distribution of $Z_c(n)$ depends on the average number of posts of each topic so far. This type of dependence has been proposed in [5]. Then the probability that the n^{th} content is posted and is about topic c is given by:

$$\begin{aligned} & P(Z_c(n) \geq \theta_c, \xi(n) = c \mid \mathbf{x}(n)) \\ = \;& P(\xi(n) = c) \times P(Z_c(n) \geq \theta_c \mid \xi(n) = c, \mathbf{x}(n)) \\ = \;& \frac{\lambda_c}{\sum_{c'} \lambda_{c'}} f_c(x_1(n),\dots,x_C(n),\theta_c). \end{aligned}$$

Example An example of an arrival process of users in a Social Network is depicted in Fig. 1. The different topics are sport and culture. The couple (t_n, c_n) describes the arrival time t_n of the n^{th} user, and his content topic c. The first horizontal axis represents time arrival of users. The second one is about posting times of contents. We observe that some users not post, for example the first one. The third horizontal axis represents non-posted contents. In our example, there are 3 users connected to the Social Network, and only the second user posts a message on the topic culture.

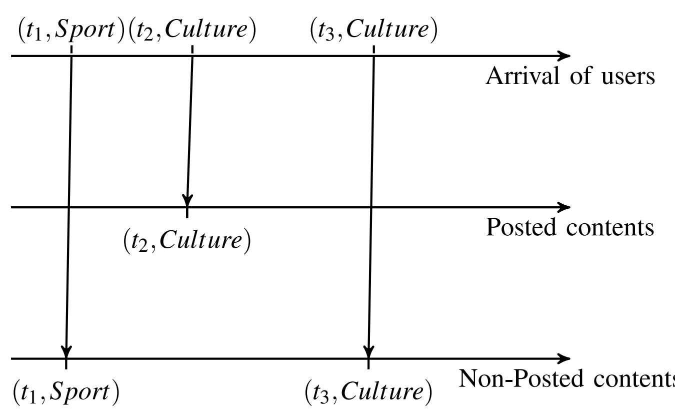

Fig. 1: Arrival time of users and their posting decision.

[2] http://balancestudy.org/

Negative externalities There is an empirical evidence that the level of popularity of a contents is an important metric for users. Users do not post (called the *Selfcensorship behavior* [1]) mainly because there is no response from the social network users [2] and [24]. By negative externalities, we mean that the average number of posts in every topic negatively influences the popularity of each topic. The following assumption presents a linear model for such externalities.

Assumption 1 Let $A \in [0,1]^{C\times C}$ a non-negative, substochastic matrix that describes the negative impacts or externalities between topics. The cc'-entry of A models the negative impact of topic c' on topic c. For each $c \in \mathscr{C}$, we consider the following posting probability:

$$f_c(x_1,\dots,x_C,\theta_c) = g_c(\theta_c)(1-\sum_{c'\in\mathscr{C}} a_{cc'}x_{c'}),$$

such that, for each $c \in \mathscr{C}$, $\sum_{c'} a_{cc'} \le 1$ and $g_c : \mathbb{R}^+ \to [0,1]$ is decreasing in θ_c.

This assumption implies that the posting probability is decreasing in any component of $\mathbf{x}$. The linearity property is similar to the one proposed in [5]. The authors consider a linear relation between the probability for a user to get some content of a given topic and the previous topics he received. The advantage of using linear function is that we can have matrix representation of the externality between topics. Therefore, the posting probability can also be expressed linearly in $\mathbf{x}$. The posting probability function $f_c(x_1,\dots,x_C)$ considered here, can be related to demand functions in Cournot competition [25] or delay function in routing games [26]. In fact, in an economic framework, the inverse linear demand is a classical assumption [25] and [27]. It is also the case in network economic models [28].

IV. REST POINT ANALYSIS

We first propose to study the asymptotic behavior of $\mathbf{x}(n)$. We provide an assumption such that it exists a unique $\mathbf{x}^*$ where $\lim_{n\to\infty}\mathbf{x}(n) = \mathbf{x}^*$. For each topic c, the evolution of the total number of posts y_c is described by:

$$y_c(n+1) = y_c(n) + \zeta_c(n),$$

where the update of the number of topic c posts $\zeta_c(n)$, is given by:

$$\zeta_c(n) := \begin{cases} 1 & \text{w.p} \quad \dfrac{\lambda_c}{\sum_{c'}\lambda_{c'}} f_c(x_1(n),\dots,x_C(n),\theta_c), \\ 0 & \text{w.p} \quad 1-\dfrac{\lambda_c}{\sum_{c'}\lambda_{c'}} f_c(x_1(n),\dots,x_C(n),\theta_c). \end{cases} \tag{1}$$

Moreover according to the previous equation, for each topic c, the evolution of the average posts $x_c(n)$ is described by the following stochastic approximation:

$$x_c(n+1) = x_c(n) + \frac{1}{n+1}\left(\zeta_c(n) - x_c(n)\right). \tag{2}$$

$\mathbf{x}^*$ is called the rest point associated to (2). Intuitively, (2) can be seen as a finite difference Euler scheme of the following system of differential equations:

$$\dot{x}_c(t) = \frac{\lambda_c}{\sum_{c'\in\mathscr{C}}\lambda_{c'}} f_c(x_1(t),\dots,x_C(t),\theta_c) - x_c(t),\ x_c(0) = x_c^0. \tag{3}$$

However, equation (2) is a stochastic difference equation, which is not the case for the classical Euler scheme. The theory of stochastic approximations [6] links the asymptotic behavior of $\mathbf{x}(n)$ and $\mathbf{x}(t)$. Introduce the following matrix $B = (B_{cc'})$ and vector $\mathbf{D}$ as:

$$\forall c,c', \quad B_{cc'} = -\frac{\lambda_c}{\sum_{c'\in\mathscr{C}}\lambda_{c'}} g_c(\theta_c)a_{cc'} - 1_{c=c'}, \tag{4}$$

(with $1_{c=c'} = 1$ if $c = c'$ otherwise it is 0) and

$$\mathbf{D} = \left[\frac{\lambda_1}{\sum_{c'\in\mathscr{C}}\lambda_{c'}} g_1(\theta_1),\dots,\frac{\lambda_C}{\sum_{c'\in\mathscr{C}}\lambda_{c'}} g_C(\theta_C)\right]. \tag{5}$$

We show in the following theorem the convergence of the dynamical process $\mathbf{x}(n)$ to the unique rest point $\mathbf{x}^*$.

Theorem 1: Consider assumption 1. The sequence $\{\mathbf{x}(n)\}$ converges almost surely to $\mathbf{x}^*$ which is

$$\forall c, \quad x_c^* = -\frac{1}{\sum_{c''}\lambda_{c''}}\sum_{c'}[B^{-1}]_{cc'}\lambda_{c'}g_{c'}(\theta_{c'}).$$

which is the unique positive solution of:

$$B\mathbf{x}^* + \mathbf{D} = 0. \tag{6}$$

V. CONTENT DIVERSITY

The second goal of our paper is to increase the content diversity of topics in a Social Network. In this section, we first define the diversity of topics in a Social Network in terms of relationship between the average number of posts of all topics. Then, we provide a necessary and sufficient condition such that the rest point $\mathbf{x}^*$ satisfies this diversity property. Using a content active filtering (CAF) approach, we provide a closed form optimal control of the posting rates that ensures this diversity property.

A. Diversity analysis

The diversity of topics into the contents published in a Social network is highly related to equal proportion of topics such that subscribers have access to all topics in an equal way. We then define the diversity property of topic rate vector $\mathbf{x} \in \mathbb{R}^C$.

Definition 1: Let $\mathbf{x} \in \mathbb{R}^C$ denotes a vector of the average number of posts for all topics in a Social Network. We say that $\mathbf{x}$ satisfies the *diversity property* if: for all $(c,c') \in \mathscr{C}$, we have:

$$x_c = x_{c'}. \tag{7}$$

The next proposition provides a necessary and sufficient condition for the rest point $\mathbf{x}^*$ to satisfy the *diversity property* given by equation (7).

Proposition 1: The rest point $\mathbf{x}^* \in [0,1]^C$ satisfies the *diversity property* if and only if: $\forall (c,c'') \in \mathscr{C}^2$

$$\frac{\lambda_c g_c(\theta_c)}{\lambda_c g_c(\theta_c)(\sum_{c'\in\mathscr{C}} a_{c,c'}) + \sum_{c'}\lambda_{c'}} = \frac{\lambda_{c''} g_{c''}(\theta_{c''})}{\lambda_{c''} g_{c''}(\theta_{c''})(\sum_{c'\in\mathscr{C}} a_{c'',c'}) + \sum_{c'}\lambda_{c'}}.$$

Diverse measures of diversity are already proposed in different scientific fields. But the most well-known are the Gini index defined in [29] and the Shannon entropy defined in [30]. Usually, a diversity measure is defined as a function $D:\mathbb{R}^C \to \mathbb{R}$ satisfying the following property:

$$\mathbf{x}^E = \underset{x}{\operatorname{argmax}}\ D(\mathbf{x}) \quad \text{iff} \quad x_c^E = x_{c'}^E,\ \forall (c,c') \in \mathscr{C}. \tag{8}$$

In this paper, we look for a diversity property and not a measure of the level of diversity in a Social Network. We study in next section how to control the posting behavior such that the rest point of the posts dynamics satisfy this diversity property.

B. Increase diversity: a content active filtering approach

What kind of control can be used to increase diversity in a Social Network? We answer this question by proposing a content active filtering (CAF) approach. For each new arrival n, if the message is posted, the Social Network may accept or not the message. Therefore the Social Network may refuse to post some messages. We consider a topic type control and we denote by $\mathbf{p} = [p_1, \dots, p_C]$ the control vector. The element $p_c \in [0,1]$ represents the probability that the content of topic c is accepted, and therefore posted on the Social Network. We assume that each user knows the Social Network CAF policy $\mathbf{p}$. Considering a given CAF $\mathbf{p}$, for all topic c the probability that a new arrival message is posted is then given by:

$$P(Z_c(n) \geq \theta_c \mid \xi(n) = c, p_c, \mathbf{x}(n)) = p_c g_c(\theta_c)(1 - \sum_{c'} a_{cc'} x_{c'}(n)).$$

Thus for each c, the evolution of $x_c(n)$ is described by the following stochastic approximation:

$$x_c(n+1) = x_c(n) + \frac{1}{n+1}\left(\zeta_c^1(n) - x_c(n)\right), \tag{9}$$

where the update of the number of topic c posts $\zeta_c^1(n)$, is given by:

$$\zeta_c^1(n) := \begin{cases} 1 & \text{w.p} \quad p_c \dfrac{\lambda_c}{\sum_{c'} \lambda_{c'}} g_c(\theta_c)(1 - \sum_{c'} a_{cc'} x_{c'}(n)), \\ 0 & \text{w.p} \quad 1 - p_c \dfrac{\lambda_c}{\sum_{c'} \lambda_{c'}} g_c(\theta_c)(1 - \sum_{c'} a_{cc'} x_{c'}(n)). \end{cases} \tag{10}$$

The rest point $\mathbf{x}^*$ of (9), according to theorem 1, is solution of the following system:

$$x_c^* = p_c \frac{\lambda_c}{\sum_{c'} \lambda_{c'}} g_c(\theta_c)(1 - \sum_{c'} a_{cc'} x_{c'}^*),\ \forall c \in \mathscr{C} \tag{11}$$

The Social Network defines a CAF $\mathbf{p}^*$ with two objectives in mind:

1) Determine $\mathbf{p}^* \in [0,1]^C$ such that if $\mathbf{x} \in [0,1]^C$ is the unique rest point of (9) then $\mathbf{x} \in [0,1]^C$ (rest point of the dynamics) satisfies (7) (the stationary regime of the average number of posts satisfies the diversity property).
2) Find $\mathbf{p}^* \in [0,1]^C$ such that the dynamical system given by equation (9) converges to $\mathbf{x}^*$.

The next theorem provides a characterization of a CAF $\mathbf{p}^*$ that satisfies the previous properties.

Theorem 2: Let assumption 1 hold and we define:

$$Y = \min_c \left(\frac{\lambda_c g_c(\theta_c)}{\lambda_c g_c(\theta_c)(\sum_{c' \in \mathscr{C}} a_{c,c'}) + \sum_{c'} \lambda_{c'}} \right), \tag{12}$$

$$c^* = \arg\min_c \left(\frac{\lambda_c g_c(\theta_c)}{\lambda_c g_c(\theta_c)(\sum_{c' \in \mathscr{C}} a_{c,c'}) + \sum_{c'} \lambda_{c'}} \right). \tag{13}$$

Then the CAF $\mathbf{p}^*$ defined by:

$$p_{c^*} = 1 \tag{14}$$

$$p_c = \frac{Y \sum_{c'} \lambda_{c'}}{\lambda_c g_c(\theta_c)(1 - Y \sum_{c' \in \mathscr{C}} a_{c,c'})}, \quad \forall c \neq c^*, \tag{15}$$

satifies the following properties:

1) the unique rest point $\mathbf{x}^* \in [0,1]^C$ of (9) satisfies the diversity property (7),
2) and the dynamical system given by equation (9) converges to $\mathbf{x}^*$.

A CAF is a simple control mechanism that can be used by a Social Network to control topics in the posts. CAF already exists, in various forms in Social Network, in order to limit the amount of received information by each user. For instance in Facebook, filtering occurs by limiting the amount of messages that are send to a user's news feed when a friend of him posts messages. CAF is also used to filter notifications that are send to members of groups on the group activity of other members. CAF can be criticize for being non democratic the sense that a Social Network may decide the contents that will be posted, without taking into account the preferences and opinions of users. However, in Facebook for example, every user could have access to two different News Feeds: a chronological one and other one created by the content active filtering (CAF) control[3]. Then, such type of content filtering exists in some Social networks, and can be designed in an efficient way given our analysis.

VI. NUMERICAL ILLUSTRATIONS

In this section, we illustrate trough simulations the theoretical results obtained in previous sections. First, we picture the impact of the negative externalities matrix A over the rest point $\mathbf{x}^*$. Second, we demonstrate that the diversity property can be obtained using the CAF defined in theorem 2. In both cases, simulations emphasize the convergence of the average number of posts for each topic depending on the number of users arriving in the social network (theorem 1).

A. Negative externalities

The contents posted in the social network are restricted to $\mathscr{C} = 2$ topics (*Culture* and *Sport* for instance), for the sake of simplicity. We propose to look at a diagonal negative externalities matrix A:

$$A = \begin{pmatrix} a & 0 \\ 0 & a \end{pmatrix}.$$

The contents have only interaction with themselves. We use the inverse function to model the popularity functions, i.e

[3]https://www.facebook.com/notes/facebook/facebook-tips-whats-the-difference-between-top-news-and-most-recent/414305122130

(a) $a = 0.1$ (b) $a = 0.9$

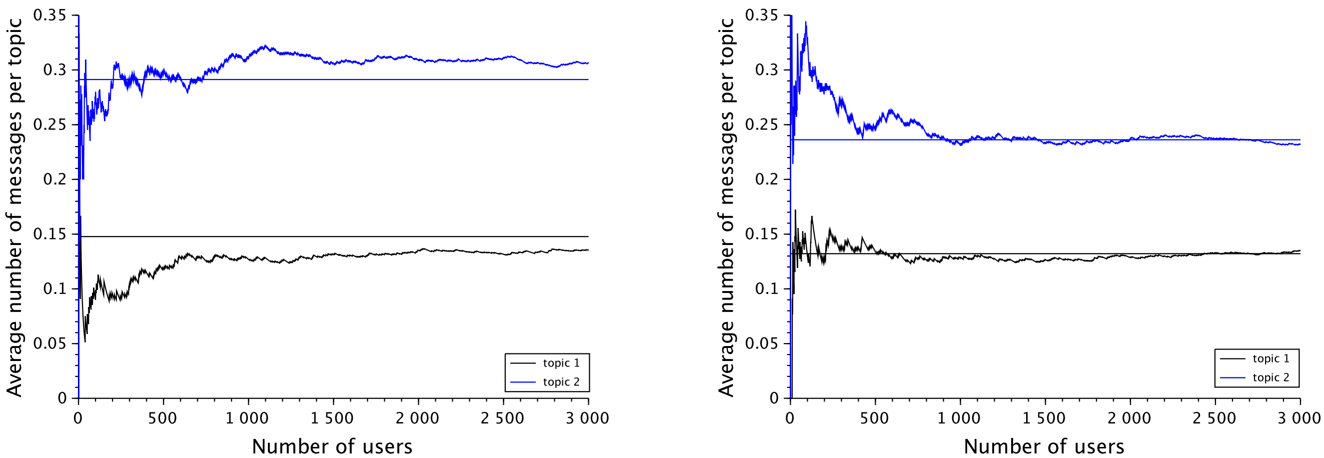

Fig. 2: Simulation of the arrival of users and messages associated to 2 topics.

$g_c(\theta_c) = \frac{1}{\theta_c}$ for each c. We assume that popularity thresholds are $\theta_1 = 3$, $\theta_2 = 2$. The intensity of the poisson process associated to users arrival is assumed to be the same for each topic c and equal to $\lambda_c = 10$.
In the Fig. 2a, we assume that $a = 0.1$ and Fig. 2b, $a = 0.9$. We can observe that, as expected, when a increases, the vector $\mathbf{x}^*$ decreases. Indeed, for $a = 0.1$, $x_1^* \approx 0.15$ and $x_2^* \approx 0.29$ and for $a = 0.9$, $x_1^* \approx 0.13$ and $x_2^* \approx 0.24$.

B. CAF simulation

We then consider a social network with $\mathscr{C} = 5$ topics.We generate, randomly, a matrix A:

$$A = \begin{pmatrix} 0.613 & 0.107 & 0.077 & 0.127 & 0.076 \\ 0.098 & 0.583 & 0.109 & 0.094 & 0.114 \\ 0.0820 & 0.094 & 0.6194 & 0.108 & 0.100 \\ 0.107 & 0.127 & 0.078 & 0.595 & 0.092 \\ 0.104 & 0.063 & 0.103 & 0.112 & 0.615 \end{pmatrix}.$$

We assume that popularity thresholds are $\theta_1 = 316$, $\theta_2 = 156$, $\theta_3 = 1136$, $\theta_4 = 446$, and $\theta_5 = 649$.
In the Fig. 3a, we observe the convergence of the dynamics of the average number of posts for these 5 topics considering the system with and without CAF. In our scenario, the CAF policy is given by the vector $\mathbf{p}^* = (0.278, 0.137, 1, 0.393, 0.571)$. This control policy implies the convergence of the dynamics to a distribution x^* among topics which satisfies the diversity property.

VII. Conclusion

In this work we first model the posting behavior in Social Networks in several topics which have negative externalities one over the other, and we propose a content active filtering in order to increase content diversity. We use dynamical approach (based on stochastic approximation theory) to model the posting behavior of users taking into account these externalities. The convergence of the posting behavior is proved. Then we propose a content active filtering control, with an explicit form, in order to improve content diversity in the social network.

Acknowledgment

This work was supported by the European project CONGAS.

Appendix

Proof of Theorem 1

The proof is made in fours steps. In the first step we highlight the link between $\mathbf{x}(n)$ and the system of differential equations (3). Then the rest point of the associated system of differential equations is proved to be unique. In the third step, by using Monotone Operator Theory [31] we prove that $\lim_{t\to\infty} \mathbf{x}(t) = \mathbf{x}^*$. Finally by the using the theory of stochastic approximation [6] we derive the convergence of $\mathbf{x}(n)$ to $\mathbf{x}^*$.
Step 1: For each c, the evolution of $x_c(n)$ is described by the following stochastic approximation:

$$x_c(n+1) = x_c(n) + \frac{1}{n+1}\left(Z_c(n) - x_c(n)\right)$$

$$= x_c(n) + \frac{1}{n+1}\left(\frac{\lambda_c}{\sum_{c'\in\mathscr{C}} \lambda_{c'}} g_c(\theta_c)(1 - \sum_{c'\in\mathscr{C}} a_{c,c'} x_{c'}(n)) - x_c(n)\right)$$

$$+ \frac{1}{n+1}\left(Z_c(n) - \frac{\lambda_c}{\sum_{c'\in\mathscr{C}} \lambda_{c'}} g_c(\theta_c)(1 - \sum_{c'\in\mathscr{C}} a_{c,c'} x_{c'}(n))\right).$$

It can be notice that, for each user n and topic c:

$$M_c(n) :=$$

$$E\left[Z_c(n) - \frac{\lambda_c}{\sum_{c'\in\mathscr{C}} \lambda_{c'}} g_c(\theta_c)(1 - \sum_{c'\in\mathscr{C}} a_{c,c'} x_{c'}(n)) \mid Z_c(m),\, m \le n\right]$$

(a) without CAF (b) with CAF

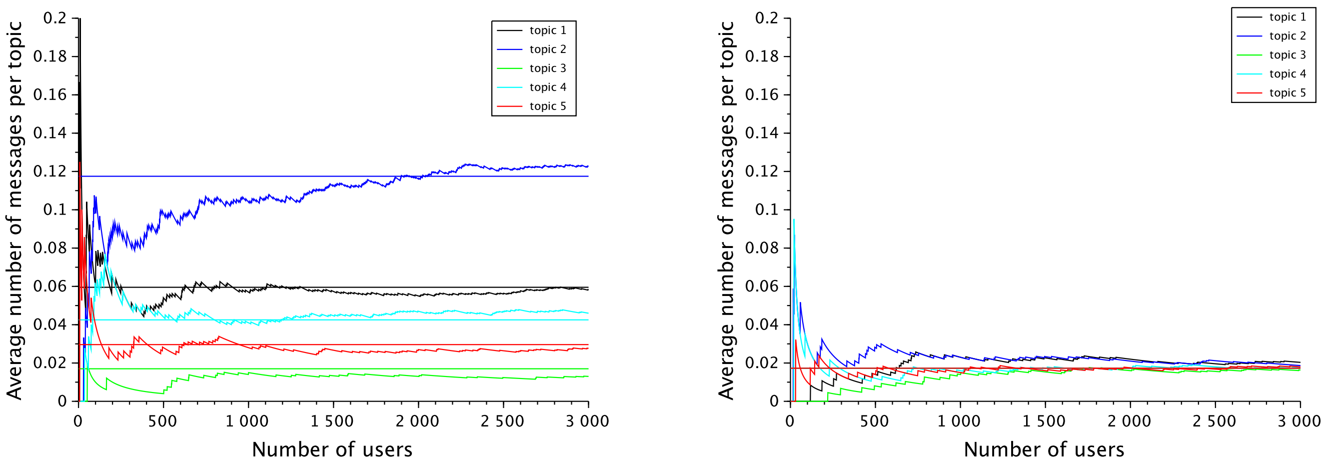

Fig. 3: Simulation of the arrival of users and messages associated to 5 topics.

$$= E[Z_c(n) \mid Z_c(m),\, m \le n] - \frac{\lambda_c}{\sum_{c' \in \mathscr{C}} \lambda_{c'}} g_c(\theta_c)(1 - \sum_{c' \in \mathscr{C}} a_{c,c'} x_{c'}(n))$$

$$= \frac{\lambda_c}{\sum_{c' \in \mathscr{C}} \lambda_{c'}} g_c(\theta_c)(1 - \sum_{c' \in \mathscr{C}} a_{c,c'} x_{c'}(n)) \times 1$$

$$+ (1 - (\frac{\lambda_c}{\sum_{c' \in \mathscr{C}} \lambda_{c'}} g_c(\theta_c)(1 - \sum_{c' \in \mathscr{C}} a_{c,c'} x_{c'}(n)))) \times 0$$

$$- \frac{\lambda_c}{\sum_{c' \in \mathscr{C}} \lambda_{c'}} g_c(\theta_c)(1 - \sum_{c' \in \mathscr{C}} a_{c,c'} x_{c'}(n)) = 0.$$

Thus $M_c(n)$ is a martingale difference sequence of zero mean (see appendix of [6] for a definition of martingale difference sequence). Following the previous argument, the sequence $\mathbf{x}(n)$ can be thought as a noisy discretization of the following system of differential equations:

$$\dot{x}_c(t) = \frac{\lambda_c}{\sum_{c' \in \mathscr{C}} \lambda_{c'}} g_c(\theta_c)(1 - \sum_{c' \in \mathscr{C}} a_{c,c'} x_{c'}(t)) - x_c(t),\ \forall c. \tag{16}$$

Step 2: Let us first define a new matrix $B \in [0,1]^{C^2}$. The cc'^{th} entry of the matrix B is given by:

$$b_{cc'} = \begin{cases} -\frac{\lambda_c}{\sum_{c' \in \mathscr{C}} \lambda_{c'}} g_c(\theta_c) a_{cc'} & if \quad c \neq c' \\ -\frac{\lambda_c}{\sum_{c' \in \mathscr{C}} \lambda_{c'}} g_c(\theta_c) a_{cc} - 1 & if \quad c = c' \end{cases} \tag{17}$$

It can be easily deduce that the rest point $\mathbf{x}^*$ of (3) is solution of:

$$B\mathbf{x}^* + \mathbf{D} = 0 \tag{18}$$

where $\mathbf{D}$ is given by:

$$\mathbf{D} = \left[\frac{\lambda_1}{\sum_{c' \in \mathscr{C}} \lambda_{c'}} g_1(\theta_1), \ldots, \frac{\lambda_C}{\sum_{c' \in \mathscr{C}} \lambda_{c'}} g_C(\theta_C) \right].$$

It can be notice that $B \in [0,1]^{C^2}$ is a strictly diagonally dominant matrix [32]. Indeed, for all c,

$$\begin{aligned} \mid b_{cc} \mid - \sum_{c' \neq c} \mid b_{cc'} \mid &= 1 + \frac{\lambda_c}{\sum_{c' \in \mathscr{C}} \lambda_{c'}} g_c(\theta_c) a_{cc} \\ &\quad - \frac{\lambda_c}{\sum_{c' \in \mathscr{C}} \lambda_{c'}} g_c(\theta_c) \sum_{c' \neq c} a_{cc'} \\ &\ge 0 \end{aligned}$$

Thus because B is strictly diagonally dominant, B is invertible and this implies that (3) has a unique rest point given by:

$$\forall c, \quad x_c^* = -\frac{1}{\sum_{c''} \lambda_{c''}} \sum_{c'} [B^{-1}]_{cc'} \lambda_{c'} g_{c'}(\theta_{c'}).$$

Step 3: We now study the stability of the unique rest point of (16). We propose to study the function:

$$V(\mathbf{x}(t)) := \sum_c (x_c(t) - x_c^*)^2, \tag{19}$$

where $\mathbf{x}^*$ is solution of (18). Then, for all $\mathbf{x}(t)$ we get

$$\begin{aligned} &\frac{d}{dt} V(\mathbf{x}(t)) \\ &= 2 < \mathbf{x}(t) - \mathbf{x}^*, B\mathbf{x}(t) + \mathbf{D} > \\ &\overset{0=B\mathbf{x}^*+\mathbf{D}}{=} 2 < \mathbf{x}(t) - \mathbf{x}^*, B\mathbf{x}(t) \\ &\qquad + \mathbf{D} - B\mathbf{x}^* - \mathbf{D} > \\ &= 2 < \mathbf{x}(t) - \mathbf{x}^*, B\mathbf{x}(t) - B\mathbf{x}^* > \\ &< 0 \end{aligned}$$

where the last inequality is coming from the fact that $-B\mathbf{x} - \mathbf{D}$ is a monotone operator. Indeed the fact that B is strictly diagonally dominant and has negative diagonal elements implies that each eigenvalue of B has a negative real part [32]. And if B is a negative definite matrix then it is also the case for the matrix $2(B+B^T)$. The following implies the monotony of $-B\mathbf{x}-\mathbf{D}$. Finally V is a Lyapunov function and using the LaSalle's Invariance Principle [6], we can deduce the following:

$$\lim_{t\to\infty}\mathbf{x}(t)=\mathbf{x}^*.$$

Step 4: Finally, we apply theorem 2 p.15 of [6] to show that $\mathbf{x}(n)$ (whose components are given by (2)) converges almost surely to the compact invariant set of the o.d.e (3) given by the singleton $\mathbf{x}^*$.

Proof of Proposition 1

We prove this proposition by looking at the necessary first, and sufficient condition second.

- Let us prove the first necessary condition, i.e. if the rest point $\mathbf{x}^* \in [0,1]^C$ satisfies (7) then $\forall(c,c'') \in \mathscr{C}^2$, we have:

$$\frac{\lambda_c g_c(\theta_c)}{\lambda_c g_c(\theta_c)(\sum_{c'\in\mathscr{C}} a_{c,c'})+\sum_{c'}\lambda_{c'}} = \frac{\lambda_{c''} g_{c''}(\theta_{c''})}{\lambda_{c''} g_{c''}(\theta_{c''})(\sum_{c'\in\mathscr{C}} a_{c'',c'})+\sum_{c'}\lambda_{c'}}.$$

  According to Definition 1 and Theorem 1, if $\mathbf{x}^* := [x,\dots,x]$ satisfies the diversity property then, for all topic $c\in\mathscr{C}$, it is the unique solution of

$$\frac{\lambda_c}{\sum_{c'}\lambda_{c'}} g_c(\theta_c)(1-\textstyle\sum_{c'\in\mathscr{C}} a_{c,c'}x^*)=x^*, \tag{20}$$

$$\Leftrightarrow x^* = \frac{\lambda_c g_c(\theta_c)}{\lambda_c g_c(\theta_c)(\sum_{c'\in\mathscr{C}} a_{c,c'})+\sum_{c'}\lambda_{c'}}. \tag{21}$$

  Then this implies that for all $\forall(c,c'')\in\mathscr{C}^2$, we have:

$$\frac{\lambda_c g_c(\theta_c)}{\lambda_c g_c(\theta_c)(\sum_{c'\in\mathscr{C}} a_{c,c'})+\sum_{c'}\lambda_{c'}} = \frac{\lambda_{c''} g_{c''}(\theta_{c''})}{\lambda_{c''} g_{c''}(\theta_{c''})(\sum_{c'\in\mathscr{C}} a_{c'',c'})+\sum_{c'}\lambda_{c'}}.$$

- Let us prove now the sufficient condition, i.e. if $\forall(c,c'')\in\mathscr{C}^2$ we have:

$$\frac{\lambda_c g_c(\theta_c)}{\lambda_c g_c(\theta_c)(\sum_{c'\in\mathscr{C}} a_{c,c'})+\sum_{c'}\lambda_{c'}} = \frac{\lambda_{c''} g_{c''}(\theta_{c''})}{\lambda_{c''} g_{c''}(\theta_{c''})(\sum_{c'\in\mathscr{C}} a_{c'',c'})+\sum_{c'}\lambda_{c'}},$$

  then the rest point $\mathbf{x}^* \in [0,1]^C$ satisfies the diversity condition given by equation (7). We notice that for each topic $c\in\mathscr{C}$, we have that

$$y := \frac{\lambda_c g_c(\theta_c)}{\lambda_c g_c(\theta_c)(\sum_{c'\in\mathscr{C}} a_{c,c'})+\sum_{c'}\lambda_{c'}}$$

  is the unique solution (because of the non singularity of matrix B used in proof of theorem 1) of the following equation:

$$\frac{\lambda_c}{\sum_{c'}\lambda_{c'}} g_c(\theta_c)(1-\sum_{c'} a_{cc'}y)=y. \tag{22}$$

  An this last equality completes the sufficient condition and hence the proof.

Proof of Theorem 2

According to proposition 1, having the equal proportion is equivalent to find $\mathbf{p}\in[0,1]^C$, $\forall c\in\mathscr{C}-\{C\}$, such that:

$$\frac{p_c\lambda_c g_c(\theta_c)}{p_c\lambda_c g_c(\theta_c)(\sum_{c'\in\mathscr{C}} a_{c,c'})+\sum_{c'}\lambda_{c'}} = \frac{p_{c''}\lambda_{c''} g_{c''}(\theta_{c''})}{p_{c''}\lambda_{c''} g_{c''}(\theta_{c''})(\sum_{c'\in\mathscr{C}} a_{c'',c'})+\sum_{c'}\lambda_{c'}}, \tag{23}$$

We can notice that we are in presence of non-convex non-concave function. Indeed we need to find $\mathbf{p}$ such that:

$$\frac{a_c p_c}{b_c p_c+d}-\frac{a_{c'} p_{c'}}{b_{c'} p_{c'}+d}=0,$$

where all the parameters are positive. Thus

$$F(p_c,p_{c'})=\frac{a_c p_c}{b_c p_c+d}-\frac{a_{c'} p_{c'}}{b_{c'} p_{c'}+d},$$

is concave in p_c and convex $p_{c'}$. Because of this property, finding $\mathbf{p}$ such that (23) is satisfied for a generic $F(p_c,p_{c'})$ remains an open question. However in this particular case we can find a way to solve this issue. Indeed if we fix a particular p_{c^*}, which provide us

$$Y := \frac{p_{c^*}\lambda_{c^*} g_{c^*}(\theta_{c^*})}{p_{c^*}\lambda_{c^*} g_{c^*}(\theta_{c^*})(\sum_{c'\in\mathscr{C}} a_{c^*,c'})+\sum_{c'}\lambda_{c'}},$$

then the problem becomes to find $\mathbf{p}_{-c^*}\in[0,1]^{C-1}$ such that

$$Y = \frac{p_{c''}\lambda_{c''} g_{c''}(\theta_{c''})}{p_{c''}\lambda_{c''} g_{c''}(\theta_{c''})(\sum_{c'\in\mathscr{C}} a_{c'',c'})+\sum_{c'}\lambda_{c'}}, \tag{24}$$

where $\mathbf{p}_{-c^*}=(p_1,\dots,p_{c^*-1},p_{c^*+1},\dots,p_C)$. We propose to use

$$c^* = \text{ argmin }_c\{\frac{\lambda_c g_c(\theta_c)}{\lambda_c g_c(\theta_c)(\sum_{c'\in\mathscr{C}} a_{c,c'})+\sum_{c'}\lambda_{c'}}\},$$

and $p_{c^*}=1$. Finally we just need to need to solve (24), and we get the solution. Moreover the control $\mathbf{p}$ does not impact the strict diagonally of matrix B. This is the reason why (9) still converge to $\mathbf{x}^*$.

References

[1] Manya Sleeper, Rebecca Balebako, Sauvik Das, Amber Lynn McConahy, Jason Wiese, and Lorrie Faith Cranor. The post that wasn't: exploring self-censorship on facebook. In *Proceedings of the 2013 conference on Computer supported cooperative work*, pages 793–802. ACM, 2013.

[2] Bruce Ferwerda, Markus Bruce, and Marko Tkalcic. To post or not to post: The effects of persuasive cues and group targeting mechanisms on posting behavior. 2014.

[3] Zhiheng Xu, Yang Zhang, Yao Wu, and Qing Yang. Modeling user posting behavior on social media. In *Proceedings of the 35th international ACM SIGIR conference on Research and development in information retrieval*, pages 545–554. ACM, 2012.

[4] Bo Jiang, Nidhi Hegde, Laurent Massoulié, and Don Towsley. How to optimally allocate your budget of attention in social networks. In *INFOCOM, 2013 Proceedings IEEE*, pages 2373–2381. IEEE, 2013.

[5] Seth A Myers and Jure Leskovec. Clash of the contagions: Cooperation and competition in information diffusion. In *ICDM*, volume 12, pages 539–548. Citeseer, 2012.

[6] Vivek S Borkar et al. Stochastic approximation. *Cambridge Books*, 2008.

[7] Motahhare Eslami, Amirhossein Aleyasen, Karrie Karahalios, Kevin Hamilton, and Christian Sandvig. Feedvis: A path for exploring news feed curation algorithms. In *Proceedings of the 18th ACM Conference Companion on Computer Supported Cooperative Work & Social Computing*, pages 65–68. ACM, 2015.

[8] Jisun An, Daniele Quercia, and Jon Crowcroft. Why individuals seek diverse opinions (or why they don't). In *Proceedings of the 5th Annual ACM Web Science Conference*, pages 15–18. ACM, 2013.

[9] Sean A Munson and Paul Resnick. Presenting diverse political opinions: how and how much. In *Proceedings of the SIGCHI conference on human factors in computing systems*, pages 1457–1466. ACM, 2010.

[10] Gretchen K Berland, Marc N Elliott, Leo S Morales, Jeffrey I Algazy, Richard L Kravitz, Michael S Broder, David E Kanouse, Jorge A Muñoz, Juan-Antonio Puyol, Marielena Lara, et al. Health information on the internet: accessibility, quality, and readability in english and spanish. *Jama*, 285(20):2612–2621, 2001.

[11] KT Goodall, LA Newman, and PR Ward. Improving access to health information for older migrants by using grounded theory and social network analysis to understand their information behaviour and digital technology use. *European journal of cancer care*, 23(6):728–738, 2014.

[12] Carlos Castillo, Marcelo Mendoza, and Barbara Poblete. Information credibility on twitter. In *Proceedings of the 20th international conference on World wide web*, pages 675–684. ACM, 2011.

[13] Nam P Nguyen, Guanhua Yan, My T Thai, and Stephan Eidenbenz. Containment of misinformation spread in online social networks. In *Proceedings of the 4th Annual ACM Web Science Conference*, pages 213–222. ACM, 2012.

[14] Ceren Budak, Divyakant Agrawal, and Amr El Abbadi. Limiting the spread of misinformation in social networks. In *Proceedings of the 20th international conference on World wide web*, pages 665–674. ACM, 2011.

[15] Reza Motamedi, Roberto Gonzalez, Reza Farahbakhsh, Reza Rejaie, Angel Cuevas, and Ruben Cuevas. Characterizing group-level user behavior in major online social networks. Technical report, Technical report available at: http://mirage. cs. uoregon. edu/pub/CIS-TR-2013-09. pdf, 2014.

[16] Hongzhi Yin, Bin Cui, Ling Chen, Zhiting Hu, and Xiaofang Zhou. Dynamic user modeling in social media systems. *ACM Transaction on Information Systems*, 2015.

[17] Francesco De Pellegrini, Alexandre Reiffers, and Eitan Altman. Differential games of competition in online content diffusion. In *Networking Conference, 2014 IFIP*, pages 1–9. IEEE, 2014.

[18] Alexandre Reiffers-Masson, Eitan Altman, and Yezekael 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.

[19] Kostas Bimpikis, Asuman Ozdaglar, and Ercan Yildiz. Competing over networks. *submitted for publication*, 2013.

[20] Sanjeev Goyal, Hoda Heidari, and Michael Kearns. Competitive contagion in networks. *Games and Economic Behavior*, 2014.

[21] Sean A Munson, Stephanie Y Lee, and Paul Resnick. Encouraging reading of diverse political viewpoints with a browser widget. In *ICWSM*, 2013.

[22] Elad Yom-Tov, Susan Dumais, and Qi Guo. Promoting civil discourse through search engine diversity. *Social Science Computer Review*, page 0894439313506838, 2013.

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

[24] Yang Wang, Gregory Norcie, Saranga Komanduri, Alessandro Acquisti, Pedro Giovanni Leon, and Lorrie Faith Cranor. I regretted the minute i pressed share: A qualitative study of regrets on facebook. In *Proceedings of the Seventh Symposium on Usable Privacy and Security*, page 10. ACM, 2011.

[25] Kostas Bimpikis, Shayan Ehsani, and Rahmi Ilkilic. Cournot competition in networked markets. In *Proceedings of the 15th acm conference of economics and coputation*, 2014.

[26] Ariel Orda, Raphael Rom, and Nahum Shimkin. Competitive routing in multiuser communication networks. *IEEE/ACM Transactions on Networking (ToN)*, 1(5):510–521, 1993.

[27] Xavier Vives. Strategic supply function competition with private information. *Econometrica*, 79(6):1919–1966, 2011.

[28] Eitan Altman, Tamer Basar, Tania Jimenez, and Nahum Shimkin. Competitive routing in networks with polynomial costs. *Automatic Control, IEEE Transactions on*, 47(1):92–96, 2002.

[29] Corrado Gini. Measurement of inequality of incomes. *The Economic Journal*, pages 124–126, 1921.

[30] Claude Elwood Shannon. A mathematical theory of communication. *ACM SIGMOBILE Mobile Computing and Communications Review*, 5(1):3–55, 2001.

[31] Heinz H Bauschke and Patrick L Combettes. *Convex analysis and monotone operator theory in Hilbert spaces*. Springer Science & Business Media, 2011.

[32] Roger A Horn and Charles R Johnson. *Matrix analysis*. Cambridge university press, 2012.