Posting behaviour Dynamics and Active Filtering for Content Diversity in Social Networks

Alexandre Reiffers Masson, Yezekael Hayel, Eitan Altman

Posting behaviour Dynamics and Active Filtering for Content Diversity in Social Networks

Alexandre Reiffers Masson, Yezekael Hayel, Eitan Altman

Abstract—In this paper, we have two objectives: First we model the posting behaviour of publishers in Social Networks which have externalities, and the second objective is to propose content active filtering in order to increase content diversity from different publishers. By externalities, we mean that when the quantity of posted contents from a specific publisher impacts the popularity of other posted contents. We introduce a dynamical model to describe the posting behaviour of publishers 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 and show that it can be obtained as the unique equilibrium of a non-cooperative game. 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 contents. As objective of the CAF we shall consider maximizing the diversity of posted contents. Finally, we illustrate our results through numerical simulations and we validate them with real data extracted from social networks.

Index Terms—Posting Behavior, Filtering Control, Stochastic Approximations, Game Theory, Online Social Networks.

I. INTRODUCTION

Nowadays, Online Social Networks (like Facebook or Twitter) have allowed a large population of users to post contents about different topics and to express different opinions. And with this amazing opportunity comes new problematics: Why do users of Social Network post? Why do they decide to post on a particular content? etc. Some works in Sociological Science try to find an answer to these questions [1] and [2]. The first goal of this paper is to model the posting behaviour of users also called publishers. In our model, users with some particular content, 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 he estimates to be potentially sufficiently popular[1] will be posted. To the best of our knowledge, mathematical models for posting behaviour do not involve popularity of contents [2], [3] and [1]. We also consider possible externalities between different contents, i.e. the number of contents from source c impacts the popularity of a content from any other sources including source c. Externalities may occur for example when content consumers have limited budget of attention [4]. In this case, negative externalities occur, meaning that the larger the number of contents the lower the popularity of other content. Other typical scenarios are described in the introduction of [5]. In previous works, the posting behaviour 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 by various publishers taking into account the externalities impact. These are modeled through a influence-correlation matrix. This matrix measures the co-influence between publisher's actions and is therefore different with the multi-variate Pearson correlation matrix. The relation between this matrix and other related concepts such as the influence can be found in [7]. 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 a non-cooperative game.

Our model has strong links to opinion diffusion models. Degroot et al., in [8], suggested the first model of opinion diffusion. He was assuming that the agents belong to social network and update their opinion by observing the one of the others and then by taking the average. The major extensions of Degroots model are summarized in [9]. In our model we assume that the opinion of a publisher concerning a topic has a linear relationship with the amount of messages posted, then our model can be viewed as an asynchronous Degroot model on a graph with positive and negative weights. Concerning the control of the posting behavior, recent papers [10], [11] derive, through the theory of optimization and control, efficient strategies for the control of opinion diffusion.Another example of such control is proposed in [12] where each individual learns his action based on signals from others. The authors determine under which conditions there is a convergence of the learning process to the right action.

The second goal of our paper is to propose content active filtering in order to increase content diversity at the rest point of the posting behaviour dynamics. 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 from various sources. 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 [13]. 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 publishers in a diverse way. We call this property content diversity. Such diversity is critical in Political News [14] and [15]. 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 [16] and [17]. 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, [18], [19] and [20].

To summarize, the main contributions of the paper are briefly described below.

- We first describe of a theoretical dynamic model of publisher's posting behavior into a system considering externalities between several publishers.
- We show that our dynamic model is similar to a game theoretic model, and therefore some results can be given.
- We optimize a measure of diversity into the contents posted by publishers and we propose a content active filtering mechanism that controls the incoming rates of posts.

Organisation of the paper. After a short state-of-the-art section, we motivate the externalities effect by studying data extracted from an online Social Network in section III. Then we introduce in section IV the posting behaviour model which include the externalities. We formulate the limit regime for appropriate scaling obtained through stochastic approximations. In the section IV-B, 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. Numerical simulations of the dynamics of the convergence of the posting behaviour dynamics are proposed in section VI. We also illustrate the usefulness of the CAF mechanism in a realistic scenario and we demonstrate how real data can be used to estimate the correlatioin matrix considering a reverse engineering viewpoint. 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 [21] and [22]. In [21], empirical analysis are proposed to the study connectivity properties and users' activities on different Social Networks, 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 [22], the authors consider a temporal context-aware mixture model to describe the behaviour of users. They assume that contents 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 [23] 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 [24], the authors propose a model in which several publishers maximize their visibility in a News Feed. This work assumes that the more a message stays on the top first messages of a News Feed,the 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 [25] and [26]. The model proposed in [25] 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 [26], 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 [27] or to improve civil discourses [28]. In [27] the authors propose to use a widget[2] that shows readers their reading diversity. A search engine algorithm is proposed in [27] to increase content diversity.

III. POPULARITY ANALYSIS BASED ON A CASE STUDY

In this section, we present an empirical study of the posting behaviour of 3 French news companies over an Online Social Network. In particular, we study the correlation between the proportion of messages from different publishers (news companies) and the average popularity of the messages. Therefore we first demonstrate the relationship between the popularity of messages from one publisher with respect to the posting rates of the other publishers. Secondly, we observe that this relationship implies negative or positive externalities.

Our dataset is composed of messages posted on three Facebook pages (Le monde.fr, Liberation, Le Figaro). We extract in total 52376 messages posted using the Netvizz app [29]. The data was collected from January 01 2010 to June 06 2015. Each message is associated with the publisher , the timestamp and the number of "likes".

The popularity of each message is measured as a function of the number of "likes" associated with it. On Figure 1, each point represents the average popularity of all messages published ruling a specific day from one news company, with respect to the visibility of other publisher's messages.

---

[2]http://balancestudy.org/

(a) Le Monde vs. Liberation
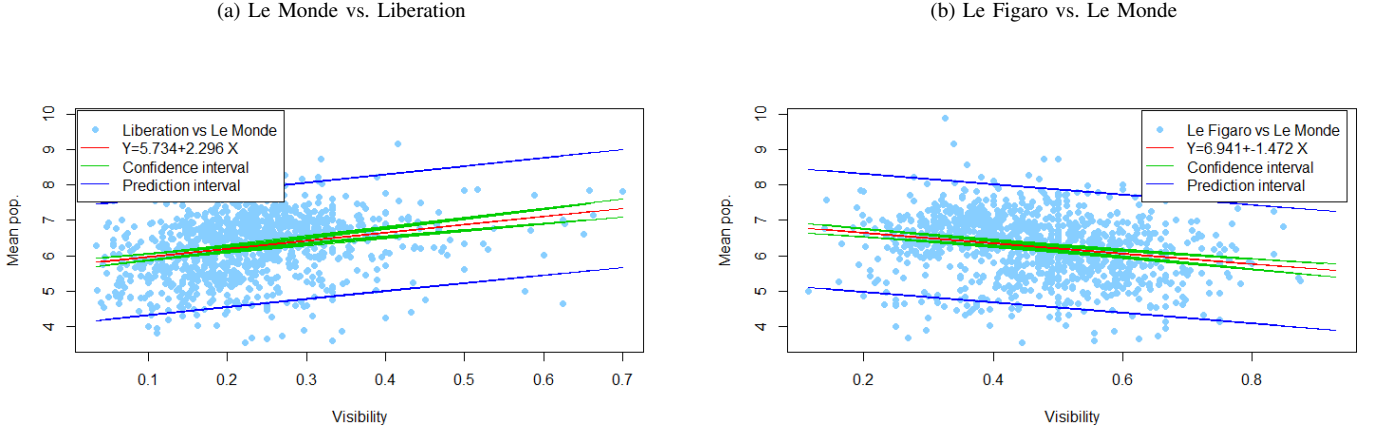
(b) Le Figaro vs. Le Monde

Fig. 1: Popularity impact over the visibility of other publisher's messages.

As we observe in the scatter plots, there is a correlation which can be positive or negative between the average popularity and the posting rates/visibilities. For instance, in fig. 1(a), the posting rate of Le Monde impacts positively the average popularity of messages posted by Liberation. However, not only positive effects are observed as it can be noticed in fig. 1(b). In this case, there is a negative relationship between the posting rate of Le Figaro and the average popularity of Le Monde's messages. The influence-correlation matrix between the 3 news companies, also called publisher s, is provided below:

$$\begin{bmatrix} 0.2564758 & -0.08708539 & -0.167834 \\ 0.2359287 & -0.2455585 & -0.01392799 \\ 0.3441988 & -0.09750635 & -0.2416787 \end{bmatrix}, \quad (1)$$

where $c \in \{$Le Monde, Libération, Le Figaro$\}$. We then demonstrate that both positive and negative correlations are possible between the popularity of messages and the posting rates of publishers. In the next section, we develop the evolution of the posting behaviour of a publisher, taking into account the correlation between the popularity of messages and the posting rates. This externality effect is captured through a influence-correlation matrix.

IV. EVOLUTION OF THE POSTING BEHAVIOUR

Based on the observations made in the previous section we propose a model of the evolution of the posted messages fraction associated to each publisher when they are sensitive to popularity. We try to find answers to the following questions:

- Can we expect a convergence of the posting behaviour over time?
- If the convergence occurs, can we have a closed form of the limit?

We provide positive answers to these questions.

A. Model

We consider that sources generate new content and then publishers decide to post it or not depending on the potential popularity of the content. More precisely, we denote by

$t_n \in \mathbb{R}_+$ the arrival instant of the nth content. For example, in Newspapers or the area of business of News, several sources like Associated Press or reporters generate new contents and send these to the publishers, which could be interested in it and therefore may want to publish it. A publisher may also be any user of a Social Network. Each content has a potential level of popularity $Z_c(n) \in \mathbb{R}_+$. For any publisher c, the arrival rate of contents 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 publisher associated to the nth content. The probability that the nth opportunity is concerning the publisher c is given by [30]:

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

For any publisher c, let $y_c(n)$ denote the total number of posts of publisher c in the Social Network during $[0, t_n]$. For each publisher c, let $x_c(n) := \frac{y_c(n)}{n}$ be the average number of posts from c. Denote that by definition we have for all publisher c and arrival instant t_n, $x_c(n) \in [0, 1]$. We denote the vector $\mathbf{x}(n) = (x_1(n), \ldots, x_C(n))$. 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. We assume that a publisher will decide to post a new content 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 publisher has enough experience to predict the popularity of his content. The random variables $Z_c(n)$ are independent. The probability that the publisher c will post his content is given by, for each n:

$$P(Z_c(n) \geq \theta_c \mid \xi(n) = c, \mathbf{x}(n)) := \quad (\frac{1}{2} \sum_{c' \in \mathscr{C}} a_{cc'}^-(1 - x_{c'}(n))$$
$$+ \frac{1}{2} \sum_{c' \in \mathscr{C}} a_{cc'}^+ x_{c'}(n)), \quad (2)$$

where the matrix $A^+ := \{a_{cc'}^+\}_{cc'}$ and $A^- := \{a_{cc'}^-\}_{cc'}$ describe, respectively, the positive (resp. the negative) impacts or externalities between publishers. These two matrices are assumed to be substochastic. The difference of $a_{cc}^+ - a_{cc}^-$ gives the global

influence of c' over c. For example if $a_{cc}^+ - a_{cc}^-$ is positive then the publisher c' has a positive influence over the publisher c.

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

$$P(Z_c(n) \geq \theta_c, \xi(n) = c \mid \mathbf{x}(n))$$

$$= P(\xi(n) = c) \times P(Z_c(n) \geq \theta_c \mid \xi(n) = c, \mathbf{x}(n))$$

$$= \frac{\lambda_c}{\sum_{c'} \lambda_{c'}} \left(\frac{1}{2} \sum_{c' \in \mathscr{C}} a_{cc'}^- (1 - x_{c'}(n)) + \frac{1}{2} \sum_{c' \in \mathscr{C}} a_{cc'}^+ x_{c'}(n) \right).$$

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 content and the previous contents he received. The advantage of using linear function is that we can have matrix representation of the externality between contents. Moreover, this probability function can be fit by using the maximum likelihood as in [5].

The posting probability function $P(Z_c(n) \geq \theta_c, \xi(n) = c \mid \mathbf{x}(n))$ considered here, can be related to demand functions in Cournot competition [31] or delay function in routing games [32]. In fact, in an economic framework, the inverse linear demand is a classical assumption [31] and [33]. It is also the case in network economic models [34].

Example from case study. Fig. 2 depicts information flows in a News Business taken from the case study described in section III. Different sources generate contents to publishers, some of them are exclusive to publishers and some of them are common sources like AP and AFP. Publishers, which are famous French News companies (Le Monde, Liberation and Le Figaro) decide to post or not their incoming content on Social Networks like Facebook, Tweeter, etc.

B. Rest point analysis

We first propose to study the asymptotic behaviour of $\mathbf{x}(n)$. We provide an assumption such that a unique $\mathbf{x}^*$ exists where $\lim_{n \to \infty} \mathbf{x}(n) = \mathbf{x}^*$. For each publisher 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 content c posts $\zeta_c(n)$, is given by:

$$\zeta_c(n) := \begin{cases} 1 & \text{w.p} \quad P(Z_c(n) \geq \theta_c, \xi(n) = c \mid \mathbf{x}(n)), \\ 0 & \text{w.p} \quad 1 - P(Z_c(n) \geq \theta_c, \xi(n) = c \mid \mathbf{x}(n)). \end{cases}$$
$$(3)$$

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

$$x_c(n+1) = x_c(n) + \frac{1}{n+1} (\zeta_c(n) - x_c(n)). \tag{4}$$

The posting rates vector $\mathbf{x}^*$ is called the rest point associated to (4). Intuitively, equation(4) can be seen as a finite difference Euler scheme of the following system of differential equations:

$$\dot{x}_c(t) = P(Z_c(t) \geq \theta_c, \xi(t) = c \mid \mathbf{x}(t)) - x_c(t), \, x_c(0) = x_c^0. \tag{5}$$

However, equation (4) is a stochastic difference equation, which is not the case for the classical Euler scheme. The theory of stochastic approximations [6] links the asymptotic behaviour of $\mathbf{x}(n)$ and $\mathbf{x}(t)$. We study these dynamics by introducing the following matrix $B = (B_{cc'})$ and vector $\mathbf{D}$ as:

$$\forall c, c', \quad B_{cc'} = \frac{\lambda_c}{2 \sum_{c' \in \mathscr{C}} \lambda_{c'}} (a_{cc'}^+ - a_{cc'}^-) - 1_{c=c'}, \tag{6}$$

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

$$\mathbf{D} = \left[\frac{\lambda_1}{2 \sum_{c' \in \mathscr{C}} \lambda_{c'}} \sum_{c'} a_{1c'}^-, \ldots, \frac{\lambda_C}{2 \sum_{c' \in \mathscr{C}} \lambda_{c'}} \sum_{c'} a_{Cc'}^- \right]. \tag{7}$$

To ensure the convergence of the stochastic process, we consider the following assumption.

Assumption 1: The matrix $\frac{1}{2}(B + B^T)$ is assumed to be definite negative.

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

Proposition 1: Let assumption 1 hold. The sequence $\{\mathbf{x}(n)\}$ converges almost surely to $\mathbf{x}^*$ with

$$\forall c, \quad x_c^* = -\frac{1}{2 \sum_{c''} \lambda_{c''}} \sum_{c'} [B^{-1}]_{cc'} \lambda_{c'} \sum_{c'} a_{cc'}^-.$$

The vector $\mathbf{x}^*$ is the unique positive solution of:

$$B\mathbf{x}^* + \mathbf{D} = 0. \tag{8}$$

In the next section, we illustrate our framework by showing that the rest point $\mathbf{x}^*$ is equivalent to a stable strategic situation (particularly a Nash Equilibrium) of a non-cooperative game between the subscribers in which each one tries to maximize his popularity by posting the contents.

C. A game theoretic perspective

In this section we demonstrate that the steady state of the posting behaviour dynamics studied in previous sections can be obtained by assuming a posting competition between the publishers. In particular, we show that the rest point $\mathbf{x}^*$ obtained in proposition 1 is a Nash equilibrium of a non-cooperative game in which each publisher strategically determines his posting rate in order to maximize the popularity of his own messages. We also propose a decentralized algorithm that computes the Nash equilibrium. Game theory yields a natural framework to study such competitive setting and the Nash equilibrium is an ideal situation in the sense that no publisher has an interest to deviate from the equilibrium. Then, in other words, the rest point of the dynamical system described in 4 is strategically stable.

Each publisher c controls his posting rate x_c, which denotes the amount of contents posted by the publisher c per unit of time (we also call it the posting rate) in the Social Network. We assume that the average popularity of a content posted by publisher c is a function of the posting rates vector $\mathbf{x}$ of all publishers. Publishers are interested in maximizing the cumulative popularity of their contents minus a cost for sending contents. In fact, sending contents is not free, as content has to be generated or obtained through economic trades. We assume a standard quadratic cost function that
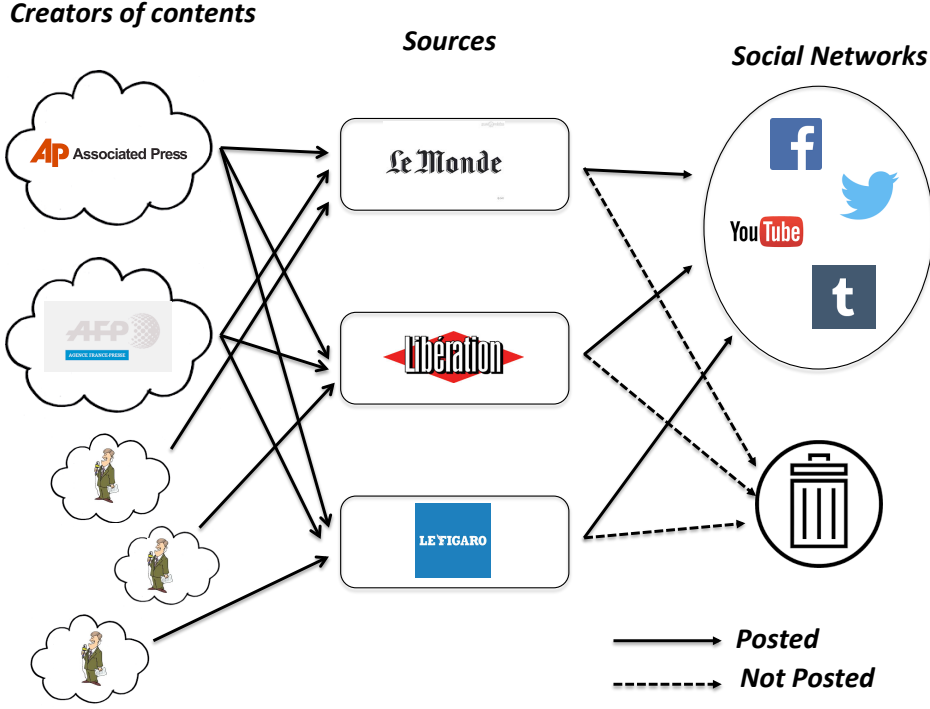
Fig. 2: Information flows in News business

depends on the posting rate x_c. The cumulative popularity is approximated by the posting rate x_c times the average posting probability $P(Z_c \geq \theta_c, \xi = c \mid \mathbf{x})$. In fact, we recall that contents are published if and only if their expected popularity is above a given threshold θ_c. Moreover, the cumulative popularity is strongly related to the number of published contents and therefore, can be approximated by the product $x_c P(Z_c \geq \theta_c, \xi = c \mid \mathbf{x})$. Finally, the average posting probability is:

$$P(Z_c \geq \theta_c, \xi = c \mid \mathbf{x}) = \frac{\lambda_c}{\sum_{c'} \lambda_{c'}} \left(\frac{1}{2} \sum_{c' \in \mathscr{C}} a_{cc'}^-(1 - x_{c'}(n)) + \frac{1}{2} \sum_{c' \in \mathscr{C}} a_{cc'}^+ x_{c'}(n) \right).$$

Then the utility function of publisher c is considered as the following:

$$U_c(x_c, \mathbf{x}_{-c}) = x_c P(Z_c \geq \theta_c, \xi = c \mid \mathbf{x}) - \frac{1}{2}(x_c)^2,$$

$$= x_c \frac{\lambda_c}{\sum_{c'} \lambda_{c'}} \left(\frac{1}{2} \sum_{c' \neq c} a_{cc'}^+ x_{c'} + \frac{a_{cc}^+}{4} x_c \right) \tag{9}$$

$$+ \frac{1}{2} \sum_{c' \neq c} a_{cc'}^-(1 - x_{c'}) + \frac{a_{cc}^-}{4}(1 - x_c) \Big) - \frac{1}{2}(x_c)^2, \tag{10}$$

where $\mathbf{x}_{-c} = [x_1, \ldots, x_{c-1}, x_{c+1}, \ldots, x_C]$. The Nash Equilibrium (NE) $\mathbf{x}^{NE}$ is a specific vector of posting rates defined as follows.

Definition 1: A vector $\mathbf{x}^{NE} \in \mathbb{R}^+$ is a Nash Equilibrium if for all publisher c,

$$x_c^{NE} \in \arg\max_{x_c \geq 0} U_c(x_c, \mathbf{x}_{-c}^{NE}). \tag{11}$$

In the following analysis of the Nash Equilibrium of the non-cooperative game, we assume first the existence of an interior Nash Equilibrium. This assumption is well known in similar types of games which are routing games [32].

Asumption 2: A vector of posting rates $\mathbf{x}^{NE}$ exists that satisfies equation (11) and such that for each publisher c, $x_c^{NE} > 0$.

Definition 2: For each c, the noisy best response of publisher c againt $\mathbf{x}_{-c}$ is given by:

$$x_c^{NE} \in \arg\max_{x_c \geq 0} U_c(x_c, \mathbf{x}_{-c}^{NE} + \varepsilon_{-c}). \tag{12}$$

where each component of $\varepsilon_{-c} \in \mathbb{R}^c$ is a zero mean gaussian noise with finite variance.

In particular, it is possible to compute a closed form expression of the best response function as:

$$0 = \frac{\partial U_c}{\partial x_c}(x_c, x_{-c} + \varepsilon_{-c})$$

$$= \frac{\lambda_c}{\sum_{c'} \lambda_{c'}} \left(\sum_{c'} a_{cc'}^+ x_{c'} + \sum_{c'} a_{cc'}^-(1 - x_{c'}) \right)$$

$$- x_c + \frac{\lambda_c}{\sum_{c'} \lambda_{c'}} \sum_{c'} (a_{cc'}^+ - a_{cc'}^-) \varepsilon_{c'}$$

$$\Leftrightarrow \left(\frac{\sum_{c'} \lambda_{c'} + \lambda_c(a_{cc}^- - a_{cc}^+)}{\sum_{c'} \lambda_{c'}} \right) x_c$$

$$= \frac{\lambda_c}{\sum_{c'} \lambda_{c'}} \left(\sum_{c' \neq c} a_{cc'}^+ x_{c'} + \sum_{c' \neq c} a_{cc'}^-(1 - x_{c'}) + a_{cc}^- \right)$$

$$+ \frac{\lambda_c}{\sum_{c'} \lambda_{c'}} \sum_{c'} (a_{cc'}^+ - a_{cc'}^-) \varepsilon_{c'}$$

$$\Leftrightarrow x_c = \frac{\lambda_c}{(\sum_{c'} \lambda_{c'} + (a_{cc}^+ - a_{cc}^-)\lambda_c)} \left(\sum_{c' \neq c} a_{cc'}^+ x_{c'} \right.$$

$$\left. + \sum_{c' \neq c} a_{cc'}^-(1 - x_{c'}) + a_{cc}^- \right)$$

$$+ \frac{\lambda_c}{(\sum_{c'} \lambda_{c'} + (a_{cc}^+ - a_{cc}^-)\lambda_c)} \sum_{c'} (a_{cc'}^+ - a_{cc'}^-) \varepsilon_{c'}. \tag{13}$$

The previous relationship between x_c and $\mathbf{x}_{-c}$ provides an algorithm way to compute the Nash equilibrium. Indeed, a decentralized algorithm that computes iteratively the best response of each publisher against the actions of the other players, converges to the Nash equilibrium. We next describe this algorithm called the Best-response algorithm, where $\mathbf{x}(t)$ is the posting rates vector of the publishers at time slot t:

Best-response algorithm [35]
Initialization: $\mathbf{x}(1)$ is randomly chosen in $[0,1]^C$.
For each round $t = 1, 2, \ldots$

1) Each publisher observes a noisy version of $\mathbf{x}_{-c}(t)$, i.e $\widetilde{\mathbf{x}}_{-c}(t) = \mathbf{x}_{-c}(t) + \varepsilon(t)$ where for each c, $\varepsilon_c(t)$ is a zero mean gaussian noise.

2) Then each publisher updates his decision by computing his best response:

$$x_c(t+1) = \mathrm{argmax}_{x_c \in [0,1]} U_c(x_c, \widetilde{\mathbf{x}}_{-c}(t))$$
$$= \frac{\lambda_c}{(\sum_{c'} \lambda_{c'} + (a_{cc}^+ - a_{cc}^-)\lambda_c)} \left(\sum_{c' \neq c} a_{cc'}^+ x_{c'}(t) \right.$$
$$\left. + \sum_{c' \neq c} a_{cc'}^-(1 - x_{c'}(t)) + a_{cc}^- \right)$$
$$+ \frac{\lambda_c}{(\sum_{c'} \lambda_{c'} + (a_{cc}^+ - a_{cc}^-)\lambda_c)} \sum_{c'} (a_{cc'}^+ - a_{cc'}^-) \varepsilon_{c'}(t). \quad (14)$$

3) Stop when $\max_c | x_c(t+1) - x_c(t) | < \varepsilon$ for $\varepsilon \ll 1$ otherwise go to step 1.

We illustrate the convergence of the best response algorithm in the numerical section VI. Finally, in the next proposition, we prove the equivalence between $\mathbf{x}^*$, the rest point of (5), and a vector $\mathbf{x}^{NE}$ which is an interior Nash equilibrium of the non-cooperative popularity game between publishers.

Proposition 2: Let Assumption 1 and Assumption 2 hold. $\mathbf{x}^* \in [0,1]^C$ is a rest point of (5) iff $\mathbf{x}^{NE} \in [0,1]^C$ is Nash Equilibrium of the previous game.

Based on this result, in the next section we describe a control mechanism, which aims to reach a diverse content flow in this dynamical system at the stationary regime, equivalent to the Nash Equilibrium point.

V. CONTENT DIVERSITY

The second goal of our paper is to increase the diversity of contents in a Social Network. We are looking for contents that are posted by different publishers. Note that even if the same content, provided by a common source, is posted by two different publishers, their posted messages will not necessarily be exactly identical and then perceived differently by the social network consumers. In this section, we first define the diversity of contents in a Social Network in terms of a relationship between the average numbers of posts of all the publishers. 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 contents published in a Social Network is highly related to equal proportion such that subscribers have

access to all contents from different publishers in an equal way. We therefore define the diversity property of publisher rate vector $\mathbf{x} \in \mathbb{R}^C$.

Definition 3: Let $\mathbf{x} \in \mathbb{R}^C$ denote a vector of the average number of posts associated to each publisher in a Social Network. We say that $\mathbf{x}$ satisfies the *diversity property* if for all $(c, c') \in \mathscr{C}$, we have:

$$x_c = x_{c'}. \quad (15)$$

Note that our definition of content diversity is related to the posting rates of the publishers and not by the content itself. In fact, messages published by different publishers may talk about the same content, but these contents are coming from different sources or are exposed differently by the publishers. We therefore talk about content diversity in order to specify that the messages posted on the Social Network come from different publishers. The next proposition provides a necessary and sufficient condition for the rest point $\mathbf{x}^*$ to satisfy the *diversity property* given by equation (15).

Proposition 3: The rest point $\mathbf{x}^* \in [0,1]^C$ satisfies the *diversity property* if and only if: $\forall (c, c'') \in \mathscr{C}^2$

$$\frac{\lambda_c \sum_{c' \in \mathscr{C}} a_{cc'}^-}{\lambda_c \sum_{c' \in \mathscr{C}} (a_{cc'}^- - a_{cc'}^+) + 2\sum_{c'} \lambda_{c'}}$$
$$= \frac{\lambda_{c''} \sum_{c' \in \mathscr{C}} a_{c''c'}^-}{\lambda_{c''} \sum_{c' \in \mathscr{C}} (a_{c''c'}^- - a_{c''c'}^+) + 2\sum_{c'} \lambda_{c'}}.$$

Diverse measures of diversity are already proposed in different scientific fields. But the most well-known are the Gini index defined in [36] and the Shannon entropy defined in [37]. Usually, a diversity measure is defined as a function $D : \mathbb{R}^C \to \mathbb{R}$ satisfying the following property:

$$\mathbf{x}^E = \mathrm{argmax}_x D(\mathbf{x}) \quad \text{iff} \quad x_c^E = x_{c'}^E, \forall (c, c') \in \mathscr{C}. \quad (16)$$

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 the next section how to control the posting behaviour 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 news 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 publisher type control and we denote by $\mathbf{p} = [p_1, \ldots, p_C]$ the control vector. The element $p_c \in [0,1]$ represents the probability that the content from publisher c is accepted, and therefore posted on the Social Network. We assume that each publisher knows the Social Network CAF policy $\mathbf{p}$. Considering a given CAF $\mathbf{p}$, for the publisher c the probability that a new arrival message is posted is then given by:

$$P(Z_c(n) \geq \theta_c \mid \xi(n) = c, p_c, \mathbf{x}(n)) =$$
$$p_c \left(\frac{1}{2} \sum_{c'} a_{cc'}^+ x_{c'}(n) + \frac{1}{2} \sum_{c'} a_{cc'}^-(1 - x_{c'}(n)) \right). \quad (17)$$

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

$$x_c(n+1) = x_c(n) + \frac{1}{n+1}\left(\zeta_c^1(n) - x_c(n)\right), \qquad (18)$$

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

$$\zeta_c^1(n) := \begin{cases} 1 & \text{w.p } \frac{\lambda_c}{\sum_{c'}\lambda_{c'}}P(Z_c(n) \geq \theta_c \mid \xi(n) = c, p_c, \mathbf{x}(n)), \\ 0 & \text{w.p } 1 - \frac{\lambda_c}{\sum_{c'}\lambda_{c'}}P(Z_c(n) \geq \theta_c \mid \xi(n) = c, p_c, \mathbf{x}(n)). \end{cases} \tag{19}$$

The rest point $\mathbf{x}^*$ of (18), according to proposition 1, is solution of the following system:

$$x_c^* = p_c\frac{\lambda_c}{\sum_{c'}\lambda_{c'}}\left(\frac{1}{2}\sum_{c'}a_{cc'}^+x_{c'}^* + \frac{1}{2}\sum_{c'}a_{cc'}^-x_{c'}^*\right), \; \forall c \in \mathscr{C}. \tag{20}$$

We observe that the stationary posting rates vector $\mathbf{x}^*$ is solution of a system which is linear with respect to the CAF vector $\mathbf{p}$. The Social Network determines 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 (18) then $\mathbf{x} \in [0,1]^C$ (rest point of the dynamics) satisfies (15) (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 (18) converges to $\mathbf{x}^*$.

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

Proposition 4: Let assumption 1 hold and we define:

$$Y = \min_c\left(\frac{\lambda_c\sum_{c'\in\mathscr{C}}a_{cc'}^-}{\lambda_c\sum_{c'\in\mathscr{C}}(a_{cc'}^- - a_{cc'}^+) + 2\sum_{c'}\lambda_{c'}}\right), \tag{21}$$

$$c^* = \arg\min_c\left(\frac{\lambda_c\sum_{c'\in\mathscr{C}}a_{cc'}^-}{\lambda_c\sum_{c'\in\mathscr{C}}(a_{cc'}^- - a_{cc'}^+) + 2\sum_{c'}\lambda_{c'}}\right). \tag{22}$$

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

$$p_{c^*} = 1 \tag{23}$$

$$p_c = \frac{2Y\sum_{c'}\lambda_{c'}}{\lambda_c(\sum_{c'}a_{cc'}^- - Y(\sum_{c'}a_{cc'}^- - a_{cc'}^+))}, \quad \forall c \neq c^*, \tag{24}$$

satisfies the following properties:

1) the unique rest point $\mathbf{x}^* \in [0,1]^C$ of (18) satisfies the diversity property (15),
2) and the dynamical system given by equation (18) converges to $\mathbf{x}^*$.

A CAF is a simple control mechanism that can be used by a Social Network to control the contents posted by the publishers. 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 sent to a user's news feed when a friend of his posts messages. CAF is also used to filter notifications that are sent to members of groups on the group activity of other members. CAF can be criticized for being non democratic in 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 another 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 in our analysis.

VI. NUMERICAL ILLUSTRATIONS AND REVERSE ENGINEERING

A. Stochastic Approximation and CAF

We first illustrate through simulations the theoretical results obtained in previous sections related to the posting behaviour dynamics and the CAF. First, we plot the convergence of the stochastic approximation to the rest point $\mathbf{x}^*$. Second, we demonstrate that the diversity property can be obtained using the CAF defined in proposition 4.

The contents posted in the Social Network are restricted to $\mathscr{C} = 3$ publishers, like the News business example with French News companies described in section III. Estimating real correlation matrices is out of the scope of the paper. By the way, correlation matrices estimators can be used as input in our framework. We consider for the sake of example, the following correlation matrices:

$$A^- = \begin{pmatrix} 0.6532 & 0.1888 & 0.1577 \\ 0.1838 & 0.6652 & 0.1507 \\ 0.1204 & 0.2129 & 0.6664 \end{pmatrix},$$

$$A^+ = \begin{pmatrix} 0.658 & 0.2159 & 0.1258 \\ 0.1418 & 0.7092 & 0.1487 \\ 0.1718 & 0.1328 & 0.6951 \end{pmatrix}.$$

In this example, the publishers have stronger influence with themselves than on another publisher. The intensity of the Poisson point process associated to contents arrival is assumed to be the same for each publisher c and equal to $\lambda_c = 10$. In the Figures 3a and 3b, we observe the convergence of the dynamics of the average number of posts for the publishers considering the system with and without CAF. In our scenario, the CAF policy is given by the vector $\mathbf{p}^* = (1, 0.5, 0.16)$. This control policy implies the convergence of the dynamics to a distribution x^* among contents which satisfies the diversity property, i.e. $x_1^* = x_2^* = x_3^*$.

B. Estimation of influence through reverse engineering

Reverse engineering is a system engineering principle based on treated information usually through disassembling a system in order to reproduce it. In our context, this idea is applied considering the data set as an information that guides us to determine the correlation matrices. Practically, we assume that the publishers are competing for maximizing the popularity of their content as proposed in section IV-C. We assume that each day, denoted by t, publishers are competing and they update their decision with the help of the best-response algorithm. Let

---

[3]https://www.facebook.com/notes/facebook/facebook-tips-whats-the-difference-between-top-news-and-most-recent/414305122130
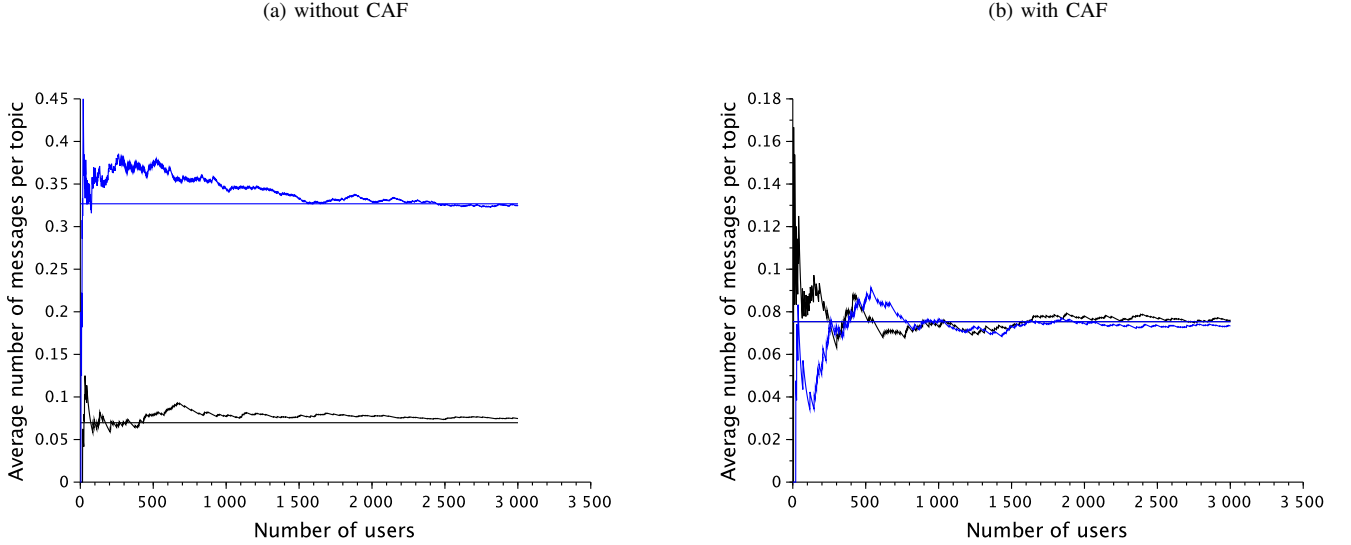
(a) without CAF

(b) with CAF

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

$x_c^{NE}(t)$ posting rate of publisher c during day t. Thus according to (14), for each c,

$$
\begin{aligned}
x_c(t+1) &= \mathrm{argmax}_{x_c \in [0,1]} U_c(x_c, \widetilde{\mathbf{x}}_{-c}(t)) \\
&= \frac{\lambda_c}{(\sum_{c'} \lambda_{c'} + (a_{cc'}^+ - a_{cc'}^-)\lambda_c)} \Big(\sum_{c' \neq c} a_{cc'}^+ x_{c'}(t) \\
&\quad + \sum_{c' \neq c} a_{cc'}^- (1 - x_{c'}(t)) + a_{cc}^- \Big) \\
&\quad + \frac{\lambda_c}{(\sum_{c'} \lambda_{c'} + (a_{cc'}^+ - a_{cc'}^-)\lambda_c)} \sum_{c'} (a_{cc'}^+ - a_{cc'}^-) \varepsilon_{c'}(t).
\end{aligned}
$$

We propose to use the least square method in order to estimate for each $(c, c') \in \{\text{Le Monde}, \text{Liberation}, \text{Le Figaro}\}^2$ and each c', the following coefficients: $A_{cc'}^* := \frac{a_{cc'}^+ - a_{cc'}^-}{a_{cc}^+ - a_{cc}^-}$ (the associated matrix is A^*) and $D_c^* := \frac{\sum_{c'} a_{cc'}^-}{a_{cc}^+ - a_{cc}^-}$ (where the associated vector is D^*). In the least square method, we solve the following optimisation problem:

$$
\min_{A^*, D^*} \| Id_C \mathbf{x}(t+1) - A^* \mathbf{x}(t) - D^* \|^2,
$$

where Id_C is the diagonal matrix of size $C \times C$. Without loss of generality, we assume that $\lambda_c = 1$ for each c. Based on our data set presented in section III, we obtain the following matrices:

$$
A^* = \begin{pmatrix} 0 & 0.49938 & 0.20559 \\ 0.30016 & 0 & -0.01863 \\ 0.34687 & 0.09159 & 0 \end{pmatrix},
$$

and

$$
D^* = \begin{pmatrix} 4.69251 \\ 6.75663 \\ 13.66454 \end{pmatrix}.
$$

Note that this method aims to estimate the total influence of each publisher on the others. In fact, we can assume that each publisher c knows his own influence, i.e. a_{cc}^+ and a_{cc}^-. Given this local knowledge and estimating the coefficient $A_{cc'}^*$, each publisher c has an estimator of the total influence of each

publisher c' on him, which is exactly the term $a_{cc'}^+ - a_{cc'}^-$. We first observe that there is not only positive or negative influences in the matrix A^*, which can be confirmed by the fact that the influence-correlation matrix (1) has also positive and negative coefficients. Moreover, it can be noticed that the publisher which has the most influence on the others is Le Monde which makes sense because it is one of the most famous newspapers compared to the others[4].

C. Best response algorithm

We describe in this section the convergence of the best-response algorithm described in section IV-C. Publishers decide simultaneously and only once their posting rate for each time slot, based on noisy observations of the rates of the others. The fact that publishers do not observe the real posting rate of the others comes from real word considerations. In order to show the impact of the uncertainty on the observations of the actions of the other players, we also plot on fig. 4 the best response algorithm assuming perfect knowledge. The dotted lines represents the case where the variance of the noise is equal to 0.1. The straight line is for the case where publishers have full knowledge of $\mathbf{x}_{-c}(t)$. We notice that the best response algorithm converges to the Nash Equilibrium, which is equal, by computation to $(0.35, 0.15, 0.87)$, in the case of perfect knowledge. In the noisy scenario, the best response gives decisions that are close to the optimal ones without noise.

VII. CONCLUSION

In this work we first model the posting behaviour of publishers in Social Networks in several contents which have externalities impact one over the other. Secondly, we propose a content active filtering in order to increase content diversity. We use dynamical approach (based on stochastic

---

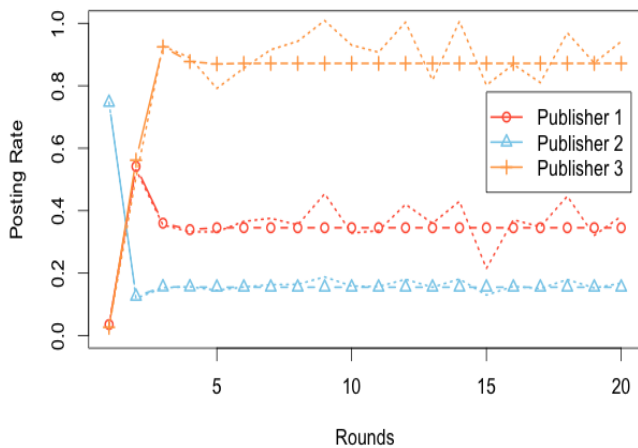[4]https://fr.wikipedia.org/wiki/Presse_en_Francecite_note-audipresseOne2012-6

Fig. 4: Dynamic of the best response algorithm. The dash lines correspond to convergence of the learning algorithm with a noisy observations of the rates of the other publishers.

approximation theory) to model the posting behaviour of publishers taking into account these externalities. The convergence of the posting behaviour is proved and an equivalent game theoretical is proposed. Then we define a content active filtering control, with an explicit form, in order to improve content diversity in the Social Network. Finally, all the theoretical results are illustrated through simulations and a data set extracted from a real Social Network is used to assess our results.

REFERENCES

[1] M. Sleeper, R. Balebako, S. Das, A. L. McConahy, J. Wiese, and L. F. Cranor, "The post that wasn't: exploring self-censorship on facebook," in *Proceedings of the 2013 conference on Computer supported cooperative work*. ACM, 2013, pp. 793–802.

[2] B. Ferwerda, M. Bruce, and M. Tkalcic, "To post or not to post: The effects of persuasive cues and group targeting mechanisms on posting behavior," *Academy of Science and Engineering (ASE), USA, ASE 2014*, 2014.

[3] Z. Xu, Y. Zhang, Y. Wu, and Q. Yang, "Modeling user posting behavior on social media," in *Proceedings of the 35th international ACM SI-GIR conference on Research and development in information retrieval*. ACM, 2012, pp. 545–554.

[4] B. Jiang, N. Hegde, L. Massoulié, and D. Towsley, "How to optimally allocate your budget of attention in social networks," in *INFOCOM, 2013 Proceedings IEEE*. IEEE, 2013, pp. 2373–2381.

[5] S. A. Myers and J. Leskovec, "Clash of the contagions: Cooperation and competition in information diffusion." in *ICDM*, vol. 12. Citeseer, 2012, pp. 539–548.

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

[7] E. L. F. Chua, H. Lauw, "Generative models for item adoptions using social correlation," *IEEE TRANSACTIONS ON KNOWLEDGE AND DATA ENGINEERING*, vol. 25, no. 9, pp. 2036–2049, 2013.

[8] M. H. DeGroot, "Reaching a consensus," *Journal of the American Statistical Association*, vol. 69, no. 345, pp. 118–121, 1974.

[9] M. O. Jackson, *Social and economic networks*. Princeton University Press Princeton, 2008, vol. 3.

[10] V. S. Borkar, J. Nair, and N. Sanketh, "Manufacturing consent," in *Communication, Control, and Computing (Allerton), 2010 48th Annual Allerton Conference on*. IEEE, 2010, pp. 1550–1555.

[11] K. Bimpikis, A. Ozdaglar, and E. Yildiz, "Competing over networks," 2014. [Online]. Available: https://asu.mit.edu/sites/default/files/documents/publications/network-marketingOR.pdf

[12] I. L. A. O. D. Acemoglu, M. Dahleh, "Bayesian learning in social networks," *The Review of Economic Studies*, vol. 78, no. 4, pp. 1201–1236, 2011.

[13] M. Eslami, A. Aleyasen, K. Karahalios, K. Hamilton, and C. 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*. ACM, 2015, pp. 65–68.

[14] J. An, D. Quercia, and J. Crowcroft, "Why individuals seek diverse opinions (or why they don't)," in *Proceedings of the 5th Annual ACM Web Science Conference*. ACM, 2013, pp. 15–18.

[15] S. A. Munson and P. Resnick, "Presenting diverse political opinions: how and how much," in *Proceedings of the SIGCHI conference on human factors in computing systems*. ACM, 2010, pp. 1457–1466.

[16] G. K. Berland, M. N. Elliott, L. S. Morales, J. I. Algazy, R. L. Kravitz, M. S. Broder, D. E. Kanouse, J. A. Muñoz, J.-A. Puyol, M. Lara *et al.*, "Health information on the internet: accessibility, quality, and readability in english and spanish," *Jama*, vol. 285, no. 20, pp. 2612–2621, 2001.

[17] K. Goodall, L. Newman, and P. 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*, vol. 23, no. 6, pp. 728–738, 2014.

[18] C. Castillo, M. Mendoza, and B. Poblete, "Information credibility on twitter," in *Proceedings of the 20th international conference on World wide web*. ACM, 2011, pp. 675–684.

[19] N. P. Nguyen, G. Yan, M. T. Thai, and S. Eidenbenz, "Containment of misinformation spread in online social networks," in *Proceedings of the 4th Annual ACM Web Science Conference*. ACM, 2012, pp. 213–222.

[20] C. Budak, D. Agrawal, and A. El Abbadi, "Limiting the spread of misinformation in social networks," in *Proceedings of the 20th international conference on World wide web*. ACM, 2011, pp. 665–674.

[21] R. Motamedi, R. Gonzalez, R. Farahbakhsh, R. Rejaie, A. Cuevas, and R. Cuevas, "Characterizing group-level user behavior in major online social networks," Technical report available at: http://mirage. cs. uoregon. edu/pub/CIS-TR-2013-09. pdf, Tech. Rep., 2014.

[22] H. Yin, B. Cui, L. Chen, Z. Hu, and X. Zhou, "Dynamic user modeling in social media systems," *ACM Transaction on Information Systems*, 2015.

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

[24] 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.

[25] K. Bimpikis, A. Ozdaglar, and E. Yildiz, "Competing over networks," *working paper*, 2013.

[26] S. Goyal, H. Heidari, and M. Kearns, "Competitive contagion in networks," *Games and Economic Behavior*, 2014.

[27] S. A. Munson, S. Y. Lee, and P. Resnick, "Encouraging reading of diverse political viewpoints with a browser widget." in *ICWSM*, 2013.

[28] E. Yom-Tov, S. Dumais, and Q. Guo, "Promoting civil discourse through search engine diversity," *Social Science Computer Review*, 2013.

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

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

[31] K. Bimpikis, S. Ehsani, and R. Ilkilic, "Cournot competition in networked markets," in *Proceedings of the 15th ACM conference of economics and computation*, 2014.

[32] A. Orda, R. Rom, and N. Shimkin, "Competitive routing in multiuser communication networks," *IEEE/ACM Transactions on Networking (ToN)*, vol. 1, no. 5, pp. 510–521, 1993.

[33] X. Vives, "Strategic supply function competition with private information," *Econometrica*, vol. 79, no. 6, pp. 1919–1966, 2011.

[34] E. Altman, T. Basar, T. Jimenez, and N. Shimkin, "Competitive routing in networks with polynomial costs," *Automatic Control, IEEE Transactions on*, vol. 47, no. 1, pp. 92–96, 2002.

[35] D. Fudenberg and D. K. Levine, *The theory of learning in games*. MIT press, 1998, vol. 2.

[36] C. Gini, "Measurement of inequality of incomes," *The Economic Journal*, pp. 124–126, 1921.

[37] C. E. Shannon, "A mathematical theory of communication," *ACM SIGMOBILE Mobile Computing and Communications Review*, vol. 5, no. 1, pp. 3–55, 2001.

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

[39] R. A. Horn and C. R. Johnson, *Matrix analysis*. Cambridge university press, 2012.

APPENDIX

Proof of Proposition 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 (5). 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 [38] 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:

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

$$= x_c(n) + \frac{1}{n+1}(P(Z_c(n) \geq \theta_c, \xi(n) = c \mid \mathbf{x}(n)) - x_c(n))$$

$$+ \frac{1}{n+1}(Z_c(n) - P(Z_c(n) \geq \theta_c, \xi(n) = c \mid \mathbf{x}(n))).$$

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

$$M_c(n) :=$$

$$E[Z_c(n) - P(Z_c(n) \geq \theta_c, \xi(n) = c \mid \mathbf{x}(n)) \mid Z_c(m), \, m \leq n]$$

$$= E[Z_c(n) \mid Z_c(m), \, m \leq n] - P(Z_c(n) \geq \theta_c, \xi(n) = c \mid \mathbf{x}(n))$$

$$= P(Z_c(n) \geq \theta_c, \xi(n) = c \mid \mathbf{x}(n)) \times 1$$

$$+ P(Z_c(n) \geq \theta_c, \xi(n) = c \mid \mathbf{x}(n)) \times 0$$

$$- P(Z_c(n) \geq \theta_c, \xi(n) = c \mid \mathbf{x}(n)) = 0.$$

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:

$$\dot{x}_c(t) = P(Z_c(t) \geq \theta_c, \xi(t) = c \mid \mathbf{x}(t)) - x_c(t), \, \forall c. \quad (25)$$

Step 2: It can be easily deduce that the rest point $\mathbf{x}^*$ of (5) is solution of:

$$B\mathbf{x}^* + \mathbf{D} = 0. \quad (26)$$

It can be notice that $B \in [0,1]^{C^2}$ is a strictly diagonally dominant matrix [39]. Indeed, for all c,

$$|b_{cc}| - \sum_{c' \neq c}|b_{cc'}| = 1 + \frac{\lambda_c}{2\sum_{c' \in \mathscr{C}}\lambda_{c'}}(a_{cc}^+ - a_{cc}^-)$$

$$- \frac{\lambda_c}{\sum_{c' \in \mathscr{C}}\lambda_{c'}}\sum_{c' \neq c}(a_{cc'}^+ - a_{cc'}^-)$$

$$\geq 0$$

Thus because B is strictly diagonally dominant, B is invertible and this implies that (5) has a unique rest point given by:

$$\forall c, \quad x_c^* = -\frac{1}{2\sum_{c''}\lambda_{c''}}\sum_{c'}[B^{-1}]_{cc'}\lambda_{c'}(\theta_{c'})\sum_{c'}a_{cc'}^-.$$

Step 3: We now study the stability of the unique rest point of (25). We propose to study the function:

$$V(\mathbf{x}(t)) := \sum_c(x_c(t) - x_c^*)^2, \quad (27)$$

where $\mathbf{x}^*$ is solution of (26). Then, for all $\mathbf{x}(t)$ we get

$$\frac{d}{dt}V(\mathbf{x}(t))$$

$$= 2 < \mathbf{x}(t) - \mathbf{x}^*, B\mathbf{x}(t) + \mathbf{D} >$$

$$\overset{0 = B\mathbf{x}^* + \mathbf{D}}{=} 2 < \mathbf{x}(t) - \mathbf{x}^*, B\mathbf{x}(t)$$

$$+ \mathbf{D} - B\mathbf{x}^* - \mathbf{D} >$$

$$= 2 < \mathbf{x}(t) - \mathbf{x}^*, B\mathbf{x}(t) - B\mathbf{x}^* >$$

$$< 0$$

where the last inequality is coming from the fact that $\frac{1}{2}(B + B^T)$ is a negative definite matrix. 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 (4)) converges almost surely to the compact invariant set of the o.d.e (5) given by the singleton $\mathbf{x}^*$.

Proof of Proposition 2

The proof is made in two steps. In the first step, we provide a characterisation of the Nash Equilibrium. In the second step we prove the equivalence. According to definition 11, a vector

$x^{NE} \in \mathbb{R}^C$ is a Nash Equilibrium if for each publisher $c \in \{1,\dots,C\}$,

$$x_c^{NE} \in \arg\max_{x_c \geq 0} \left(x_c \frac{\lambda_c}{2\sum_{c'} \lambda_{c'}} (Y_c + \frac{a_{cc}^+}{2} x_c + \frac{a_{cc}^-}{2}(1-x_c)) - \frac{1}{2}(x_c)^2 \right). \tag{28}$$

where $Y_c := \sum_{c' \neq c} a_{cc'}^- (1 - x_{c'}^{NE}) + \sum_{c' \neq c} a_{cc'}^+ x_{c'}^{NE}$. We are dealing with a concave optimisation problem and the first order optimality conditions, for each publisher c, gives us the following characterisation. A posting rates vector $\mathbf{x}^{NE}$ is a Nash Equilibrium if and only if for each publisher c,

$$\frac{\partial U_c}{\partial x_c}(x_c^{NE}, \mathbf{x}_{-c}^{NE}) = 0, \text{ with } x_c^{NE} > 0, \tag{29}$$

$$\frac{\partial U_c}{\partial x_c}(x_c^{NE}, \mathbf{x}_{-c}^{NE}) \leq 0, \text{ with } x_c^{NE} \geq 0, \tag{30}$$

which are equivalent to the following:

$$\frac{\lambda_c}{2\sum_{c' \in \mathscr{C}} \lambda_{c'}} \left(\sum_{c' \in \mathscr{C}} a_{c,c'}^-(1 - x_{c'}^{NE}) + \sum_{c' \in \mathscr{C}} a_{c,c'}^+ x_{c'}^{NE} \right) = x_c^{NE},$$
$$\text{if } x_c^{NE} > 0, \tag{31}$$

$$\frac{\lambda_c}{2\sum_{c' \in \mathscr{C}} \lambda_{c'}} \left(\sum_{c' \in \mathscr{C}} a_{c,c'}^-(1 - x_{c'}^{NE}) + \sum_{c' \in \mathscr{C}} a_{c,c'}^+ x_{c'}^{NE} \right) \leq x_c^{NE},$$
$$\text{if } x_c^{NE} \geq 0. \tag{32}$$

Assuming the existence of an interior Nash equilibrium (Assumption 2), we then obtain that the first order conditions given by equation (31) are equivalent to the rest point solution obtained in Proposition 1. Thus, an interior Nash equilibrium of the non-cooperative posting game is equivalent to the rest point of the posting behaviour dynamics proposed in equations (26).

Proof of Proposition 3

We prove this proposition by looking at the necessary condition 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 (15) then $\forall (c,c'') \in \mathscr{C}^2$, we have:

$$\frac{\lambda_c \sum_{c' \in \mathscr{C}} a_{cc'}^-}{\lambda_c \sum_{c' \in \mathscr{C}} (a_{cc'}^- - a_{cc'}^+) + 2\sum_{c'} \lambda_{c'}}$$
$$= \frac{\lambda_{c''} \sum_{c' \in \mathscr{C}} a_{c''c'}^-}{\lambda_{c''} \sum_{c' \in \mathscr{C}} (a_{c''c'}^- - a_{c''c'}^+) + 2\sum_{c'} \lambda_{c'}}.$$

According to Definition 3 and proposition 1, if $\mathbf{x}^* := [x,\dots,x]$ satisfies the diversity property then, for all content $c \in \mathscr{C}$, it is the unique solution of

$$\frac{\lambda_c}{2\sum_{c'} \lambda_{c'}} (\sum_{c' \in \mathscr{C}} a_{c,c'}^-(1 - x^*) + \sum_{c' \in \mathscr{C}} a_{c,c'}^+ x^*) = x^*, \tag{33}$$

$$\Leftrightarrow x^* = \frac{\lambda_c \sum_{c' \in \mathscr{C}} a_{cc'}^-}{\lambda_c \sum_{c' \in \mathscr{C}} (a_{cc'}^- - a_{cc'}^+) + 2\sum_{c'} \lambda_{c'}}. \tag{34}$$

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

$$\frac{\lambda_c \sum_{c' \in \mathscr{C}} a_{cc'}^-}{\lambda_c \sum_{c' \in \mathscr{C}} (a_{cc'}^- - a_{cc'}^+) + 2\sum_{c'} \lambda_{c'}}$$
$$= \frac{\lambda_{c''} \sum_{c' \in \mathscr{C}} a_{c''c'}^-}{\lambda_{c''} \sum_{c' \in \mathscr{C}} (a_{c''c'}^- - a_{c''c'}^+) + 2\sum_{c'} \lambda_{c'}}.$$

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

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

then the rest point $\mathbf{x}^* \in [0,1]^C$ satisfies the diversity condition given by equation (15). We notice that for each content $c \in \mathscr{C}$, we have that

$$y := \frac{\lambda_c \sum_{c' \in \mathscr{C}} a_{cc'}^-}{\lambda_c \sum_{c' \in \mathscr{C}} (a_{cc'}^- - a_{cc'}^+) + 2\sum_{c'} \lambda_{c'}}$$

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

$$\frac{\lambda_c}{2\sum_{c'} \lambda_{c'}} (\sum_{c'} a_{cc'}^+ y + \sum_{c'} a_{cc'}^-(1 - y)) = y. \tag{35}$$

This last equality completes the sufficient condition and hence the proof.

Proof of Proposition 4

According to proposition 3, having the equal proportion is equivalent to find $\mathbf{p} \in [0,1]^C$, $\forall c \in \mathscr{C} - \{C\}$, such that:

$$\frac{p_c \lambda_c \sum_{c' \in \mathscr{C}} a_{cc'}^-}{p_c \lambda_c \sum_{c' \in \mathscr{C}} (a_{cc'}^- - a_{cc'}^+) + 2\sum_{c'} \lambda_{c'}}$$
$$= \frac{p_{c''} \lambda_{c''} \sum_{c' \in \mathscr{C}} a_{c''c'}^-}{p_{c''} \lambda_{c''} \sum_{c' \in \mathscr{C}} (a_{c''c'}^- - a_{c''c'}^+) + 2\sum_{c'} \lambda_{c'}}.$$

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 (15) 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

$$Y := \frac{p_{c*} \lambda_{c*} \sum_{c' \in \mathscr{C}} a_{c*c'}^-}{p_{c*} \lambda_{c*} \sum_{c' \in \mathscr{C}} (a_{c*c'}^- - a_{c*c'}^+) + 2\sum_{c'} \lambda_{c'}}$$

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

$$Y = \frac{p_c \lambda_c \sum_{c' \in \mathscr{C}} a_{cc'}^-}{p_c \lambda_c \sum_{c' \in \mathscr{C}} (a_{cc'}^- - a_{cc'}^+) + 2\sum_{c'} \lambda_{c'}} \tag{36}$$

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

$$c^* = \arg\min_c \left\{ \frac{\lambda_c \sum_{c' \in \mathscr{C}} a_{cc'}^-}{\lambda_c \sum_{c' \in \mathscr{C}} (a_{cc'}^- - a_{cc'}^+) + 2\sum_{c'} \lambda_{c'}} \right\},$$

and $p_{c^*} = 1$. Finally we just need to need to solve (36), and we get the solution. Moreover the control $\mathbf{p}$ does not impact the strict diagonally of matrix B. This is the reason why (18) still converge to $\mathbf{x}^*$.