Competition over visibility and popularity on Online Social Networks

Alexandre Reiffers-Masson

ACADÉMIE D'AIX-MARSEILLE
UNIVERSITÉ D'AVIGNON ET DES PAYS DE VAUCLUSE

THÈSE

présentée à l'Université d'Avignon et des Pays de Vaucluse
pour obtenir le diplôme de DOCTORAT

SPÉCIALITÉ : Informatique

École Doctorale 536 « Sciences et Agronomie »
Laboratoire d'Informatique (EA 4128)

Compétition sur la visibilite et la popularité dans les réseaux sociaux en ligne

par

Alexandre Reiffers-Masson

Soutenue publiquement le 12 janvier 2016 devant un jury composé de :

M.	Vivek Borkar	Professeur, Indian Institute of Technology Bombay	Rapporteur
M.	Georges Zaccour	Professeur, HEC Montréal	Rapporteur
Mme.	Nidhi Hegde	Chercheuse Senior, Bell Labs	Examinateur
Mme.	Clémence Magnien	Directrice de Recherche, CNRS	Examinateur
M.	Tijani Chahed	Professeur, Telecom SudParis	Examinateur
M.	Pierre-Henri Morand	Professeur, Université d'Avignon	Examinateur
M.	Eitan Altman	Directeur de Recherche, INRIA	Directeur de thèse
M.	Yezekael Hayel	Maître de Conférence (HDR), Université d'Avignon	Directeur de thèse

Laboratoire d'Informatique d'Avignon - CERI
Centre d'Enseignement et de Recherche en Informatique

Equipe MAESTRO
Institut national de recherche en informatique et en automatique

À ma femme et mes parents.

Remerciement

Tout d'abord je tiens à exprimer ma gratitude envers mes directeurs de thèse Eitan Altman et Yezekael Hayel pour leur participation et leur soutient constant. En m'accueillant dans leur laboratoire respectif, ils m'ont donné la chance de devenir un véritable chercheur.

Je remercie le laboratoire d'Avignon et l'équipe Maestro qui m'ont accueilli. J'y ai trouvé une atmosphère de travail sereine et créative qui m'a permis d'avancer.

Mes reconnaissances vont ensuite à Nahum Shimkin et Francesco De Pelligrini avec qui j'ai collaboré.

Je suis très reconnaissante envers Messieurs Vivek Borkar et Georges Zaccour qui ont accepté d'être les rapporteurs de ma thèse. Je remercie également Madame Nidhi Hegde, Madame Clémence Magnien et Messieurs Tijani Chahed, Pierre-Henri Morand de me faire l'honneur de faire partie de mon jury de thèse.

Je remercie également ma famille, mes amis et mes collègues qui m'ont permis de progresser dans la recherche.

Merci à mes collègues du laboratoire d'Avignon pour leur relecture enrichissante de ma thèse.

Enfin, je dédie cette thèse à ma femme qui m'a toujours soutenu durant mes études.

Résumé

Résumé

Cette thèse utilise la théorie des jeux pour comprendre le comportement des usagers dans les réseaux sociaux. Trois problématiques y sont abordées: "Comment maximiser la popularité des contenus postés dans les réseaux sociaux?";" Comment modéliser la répartition des messages par sujets?";"Comment minimiser la propagation d'une rumeur et maximiser la diversité des contenus postés?".
Après un état de l'art concernant ces questions développé dans le chapitre 1, ce travail traite, dans le chapitre 2, de la manière d'aborder l'environnement compétitif pour accroître la visibilité. Dans le chapitre 3, c'est le comportement des usagers qui est modélisé, en terme de nombre de messages postés, en utilisant la théorie des approximations stochastiques. Dans le chapitre 4, c'est une compétition pour être populaire qui est étudiée. Le chapitre 5 propose de formuler deux problèmes d'optimisation convexes dans le contexte des réseaux sociaux en ligne. Finalement, le chapitre 6 conclue ce manuscrit.

Mots-clés

théorie des jeux, réseaux sociaux en ligne, jeux différentiels, approximation stochastique, optimisation convexe, competition pour être populaire, competition pour être visible, propagation de rumeurs, maximisation de la popularité.

Abstract

Abstract

This Ph.D. is dedicated to the application of the game theory for the understanding of users behaviour in Online Social Networks. The three main questions of this Ph.D. are: " How to maximize contents popularity ? "; " How to model the distribution of messages across sources and topics in OSNs ? "; " How to minimize gossip propagation and how to maximize contents diversity? ".
After a survey concerning the research made about the previous problematics in chapter 1, we propose to study a competition over visibility in chapter 2. In chapter 3, we model and provide insight concerning the posting behaviour of publishers in OSNs by using the stochastic approximation framework. In chapter 4, it is a popularity competition which is described by using a differential game formulation. The chapter 5 is dedicated to the formulation of two convex optimization problems in the context of Online Social Networks. Finally conclusions and perspectives are given in chapter 6.

Keywords

game theory, online social network, differential game, stochastic approximation, convex optimization, popularity competition, visibility competition, posting behaviour, gossip minimization, popularity maximization.

Contents

Chapter 1

Introduction

Contents

Les réseaux sociaux en ligne font désormais partie de notre vie courante. Il y avait 968 millions d'utilisateurs quotidiens de Facebook, en moyenne, en Juin 2015. En d'autres termes un septième de la population mondiale utilise Facebook. Ainsi comprendre quel est l'impact que peut avoir l'utilisation d'un réseau social par ses abonnés constitue un défi important. Pour ce faire nous nous proposons de répondre aux trois problématiques suivantes:

- Comment maximiser la popularité des contenus postés dans les réseaux sociaux?
- Comment modéliser la répartition des messages par sujets?
- Comment minimiser la propagation d'une rumeur et maximiser la diversité des contenus postés?

Le chapitre 2 est basé sur deux articles . Le premier est "A Time and Space Routing Game Model applied to Visibility Competition on Online Social Networks" et le second s'intitule "Game theory approach for modeling competition over visibility on social networks". Premièrement, nous proposons dans ce chapitre une mesure de visibilité dans les réseaux sociaux en ligne. Ensuite, nous présentons notre jeu basé sur la maximisation de cette mesure de visibilité. Nous prouvons l'unicité de l'équilibre de Nash ainsi qu'une caractérisation par un problème d'optimisation concave. Ensuite, nous définissons un algorithme décentralisé où chaque utilisateur estime le débit total de messages dans chaque News Feed et utilise l'algorithme de gradient stochastique afin de mettre à jour ses propres flux.

Le chapitre 3 est basé sur l'article "Posting behaviour in Social Networks and Content Active Filtering". Dans ce chapitre, nous avons deux objectifs: d'abord, nous modélisons le nombre de messages postés par les usagers à l'aide de la théorie des approximations stochastiques. Le second objectif est de proposer une strategie de filtrage de contenu afin d'augmenter la diversité des contenus postés. Nous fournissons une expression mathématique du point d'équilibre de notre modèle qui est aussi équivalent à un équilibre de Nash d'un jeu non-coopératif. Enfin, nous illustrons nos résultats grâce à des simulations numériques et nous les validons avec des données réelles extraites des réseaux sociaux.

Le chapitre 4 est basé sur le papier "Differential Games of Competition in Online Content Diffusion ". Dans ce chapitre, nous utilisons la théorie des jeux différentiels pour modéliser la compétition se passant entre les créateurs de contenus qui paient le réseau social pour être populaires. Nous utilisons les équation d'Hamilton Jacobi pour calculer l'équilibre symétrique. Nous montrons qu'une politique bang-bang est un équilibre de Nash.

Le chapitre 5 est basé sur les papiers "Controlling the Katz-Bonacich Centrality in Social Network: Application to gossip in Online Social Networks" et "A Continuous Optimization Approach for Maximizing Content Popularity in Online Social Networks". Dans ce chapitre, nous considérons deux problèmes d'optimisation. Le premier est liée à l'optimisation de la centralité Katz-Bonacich à l'aide d'un contrôle topologique qui est appelé le problème d'optimisation Katz-Bonacich. Nous montrons d'abord que ce problème est équivalent à un problème d'optimisation linéaire. La deuxième partie de ce chapitre s'intéresse à l'optimisation de la popularité dans les réseaux sociaux. Nous modélisons l'évolution de la popularité du contenu en ligne sur les réseaux sociaus. Nous considérons le problème de maximisation de popularité lorsqu'un un usager contrôle les sujets associés à ses contenus et joue sur le nombre de contenus postés afin de maximiser sa popularité globale. Nous démontrons que ce problème peut être écrit comme un programme fractionnaire "généralisé" et nous fournissons un algorithme qui converge vers la solution optimale.

Dans le chapitre 6, nous décrivons les perspectives possibles de notre thèse.

1.1 OSN: an introduction

Online Social Networks (OSNs) are now part of our daily lives. There were 968 million of active daily Facebook users on average in June 2015, in other words one seventh of the world population is using Facebook. Thus understanding what is the impact of an OSN on its subscribers is an important challenge. To this end, the first step is to define what is an OSN. We adopt the definition of OSNs proposed in [1] with some small differences. From our perspective, OSNs are websites where users[1] can create a public profile, construct relationships with other users and get information on them in the most efficient way. The last point is, from our perspective, one of the keys to success of OSNs (as affirmed in TIME[2]). More precisely, the first novelty that OSNs bring is the fact that users can have a massive quantity of "friends" even if they do not know them in the real world. Indeed, OSNs allow users to see connections between their friends and proposed friends recommendations [1]. It is therefore easy for a user to find other people with common interests from all around the world. In an ideal world, OSNs would dissolve the boundaries of social classes. However, when researchers have observed friendship relations on real data it shows that the most common connections are between individuals who already know each other in the offline world [2]. The second novelty brought by OSNs is the fact that users can maintain a sustainable relationship with each one of his friends, which is not an easy task knowing that an average user has around 350 friends[3]. This is achievable because of technological progresses made by OSNs companies about the selection of relevant information. Indeed, the information that a

[1] OSN subscribers

[2] Luckerson, V. (July 9, 2015). Here's how facebook's news feed actually works. TIME.

[3] http://goo.gl/ii71G6

user receives about his friends are aggregated in a Feed[4]. Therefore, a user does not need to visit his friends profile to get a update on them. Depending on the OSN you are looking at, the name of this feed can change. For instance, on Facebook, it is called a news feed[5] and on Twitter a timeline[6]. Last year, the OSN Facebook updated its news feed algorithm seven times [7].

Because of the novelty of OSNs, a new need has arisen to understand them better. According to [3], the main issues of OSNs are:

How can a company use an OSN efficiently? Behind this wide issue, two major questions seem to be the key to solve the problem : How to become popular on OSNs? How to maintain a good reputation in OSNs? The first question will be the topic of a full section, later on in this thesis. For companies, OSNs are not just an advertisement tool. Indeed, they can be used to interact with their customers [4]. In France, one example of a company that succeeded in using OSNs to manage customer relations is Air France[8].

How to control spread of gossip in OSNs? Nowadays, everyone is able to have information about any event, no matter where. For instance, by using the classic information channels such as newspapers or, by using OSNs where your friends also provide some information about the event. However, the second source is not reliable. This raises the question of gossip detection [5] and how to contain it [6].

How to solve the privacy dilemma in OSNs? One aspect of the privacy dilemma can be explained as follows. If users want to interact with their friends in OSNs, they have to share personal information. However, this will generate privacy issues [7]. Thus the more a user interacts with his friends, the more there is a risk that another user posts undesired information on his news feed. Moreover, private information has an economic value and therefore there is a risk that OSNs collect user's personal information for commercial purposes.

Also OSNs provide data that can be used to find answers to research issues such as understanding the diffusion of information. OSNs increase the diffusion of information and provide important datasets which were not available before [8]. There is an impressive amount of papers about this phenomenon, such as described in the survey [9]. The first and well known topic in it, is the propagation of a unique information (cascade model [10], SIR model [11]). Nonetheless, in the case of multiple information propagation, it is not clear how to model negative and positive interactions between the different propagation processes. In our opinion, the most advanced work around this subject is [12], which models positive and negative interactions, and where the parameters of the

[4] Putting together news that come from various friends of a subscriber is called *aggregation* of posts in a news feed.

[5] www.facebook.com/help/210346402339221

[6] support.twitter.com/articles/227251

[7] http://goo.gl/uYQEzq

[8] http://goo.gl/6JgMIo

model are estimated. Yet, the question of control is not addressed in it.

OSNs are hard and complex entities for the following reasons: First of all, the description of the content creators' strategies is puzzling. For instance, content creators (sources) can use different channels to express their opinions, where groups[9] and pages[10] are the majors. In [13], the authors show the importance of posting in the right community on the OSN Reddit[11]. These messages (or contents) need to be adapted for each channel, in what information and communication science call "transmedia". It can be defined as " [...] stories told across multiple media. At the present time, the most significant stories tend to flow across multiple media platforms " [14]. When the content is posted through different medias but it does not adapt its message, it is called "cross-media". Finally, complex issues can be found for content consumers (readers), which deal daily with a huge amount of information. The large number of choices that they are faced makes it difficult to find the information of their interest. This can be illustrated by the huge number of filtering algorithm proposed in research [15; 16; 17]. Fig. 1.1 attempt to capture the complexity of OSN users' behaviour.

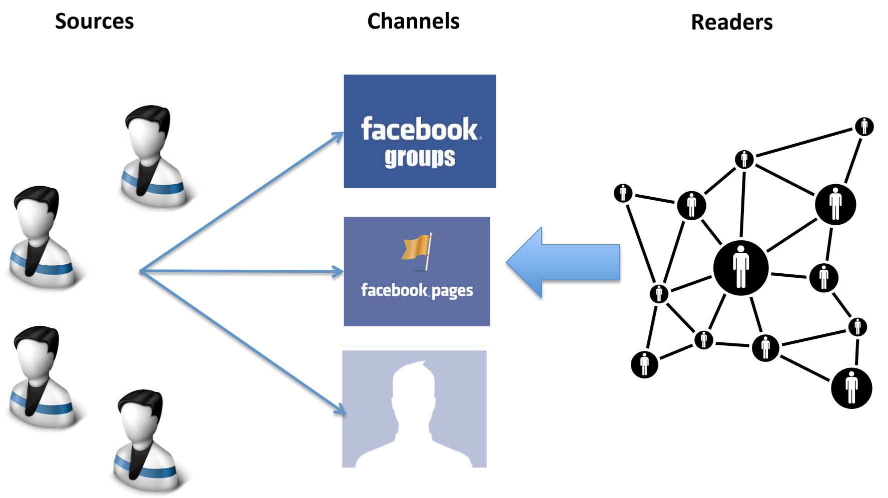


Figure 1.1: *OSN entity*

In this introduction we present three main issues, all of them belonging to the general challenges proposed above, and which we will study in this Ph.D. thesis : " How to maximize contents popularity ? "; " How to model the distribution of messages across sources and topics in OSNs ? "; " How to minimize gossip propagation and how to maximize contents diversity? ".

[9] http://mashable.com/2011/05/01/group-tumblr//
[10] https://www.facebook.com/pages/create/?ref_type=sitefooter
[11] reddit.com

In section 1.2, we first give an overview of the popularity optimization problem in which there is a unique source seeking to optimize the popularity of its contents. The intrinsic variables that increase popularity will be given. Then, we will describe our approach to tackle the popularity optimization problem.

Section 1.3 will be dedicated to the investigation of the posting behaviour of sources. We first explain why we assume that the main interest of sources is their contents popularity. Then we will introduce our game theoretical framework and explain the reason why it is the best adapted to model the posting behaviour of sources and the nature of the reached equilibrium.

Finally, section 1.4 will address the issue of gossip propagation and poor content diversity and then we will describe what kind of control which is possible to apply and its effect on the equilibrium.

1.2 Maximization of contents popularity

We propose understanding what drives the popularity of contents in OSNs and how to maximize it[12]. For companies it is an important issue to solve because they have already observed that investment in OSNs increases their sales:

"More than half of marketers who've been using social media for at least 2 years report it helped them improve sales.[13]"

However, even if the usefulness of OSNs was noticed, the number-one challenge marketers are interested in is:

"What social tactics are most effective?"[14]

This is the reason why understanding how to maximize popularity is of major concern. Firstly we will describe our methodology to tackle the following issue. Secondly an overview of previous works will be provided.

1.2.1 Definition of popularity

The popularity of a content on OSNs is defined as the number of likes, shares, views, comments, reactions to the content as suggested in [18]. In fig. 1.2, the popularity

[12] The associated optimization problem will be called the Popularity maximization problem.

[13] www.socialmediaexaminer.com/SocialMediaMarketingIndustryReport2015.pdfp18

[14] www.socialmediaexaminer.com/SocialMediaMarketingIndustryReport2015.pdfp6

evolution of a message posted on Facebook is plotted. This figure suggests two possible approaches to model the popularity. The first one is to restrict the study of a static problem which is obtained by computing the expected stationary regime of the popularity. If we do so, it facilitates the introduction of several contents. The second approach consists of trying to find a mathematical expression which captures the popularity evolution of a content. However, in this case, the question of how to model the interaction between contents is not fully understood.

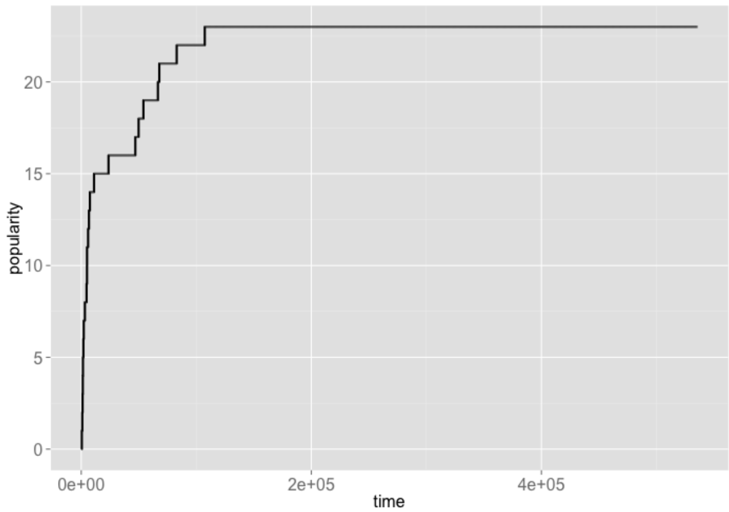


Figure 1.2: *Example of the popularity evolution of a message posted in Anne Hidalgo's Facebook page.*

First we study the popularity of contents associated with a particular source by introducing the visibility of this source. The visibility of a source j in an OSN m is defined as the proportion of content posted by j in m. We conduct a simple experiment to observe, whether or not the visibility of a source is correlated to its expected stationary regime popularity. We extract posted messages by a Facebook page associated with newspaper companies (Le Monde[15], Liberation[16], Le Figaro[17]) by using the Netvizz application [19]. Our dataset is composed of 63437 messages. The data was collected from January 01 2010 to June 06 2015. For different visibility of the Le Monde, a smooth density estimate of message popularity (according to a logarithmic scale) is depicted in fig. 1.3. We observe in this graph that the more the visibility increases the more the number of popular messages increases.

The visibility approach is the one proposed in chapter 2.

In the second approach proposed in chapter 3, we assume that the expected station-

[15] https://fr-fr.facebook.com/lemonde.fr
[16] https://www.facebook.com/Liberation?fref=ts
[17] https://www.facebook.com/lefigaro?fref=ts

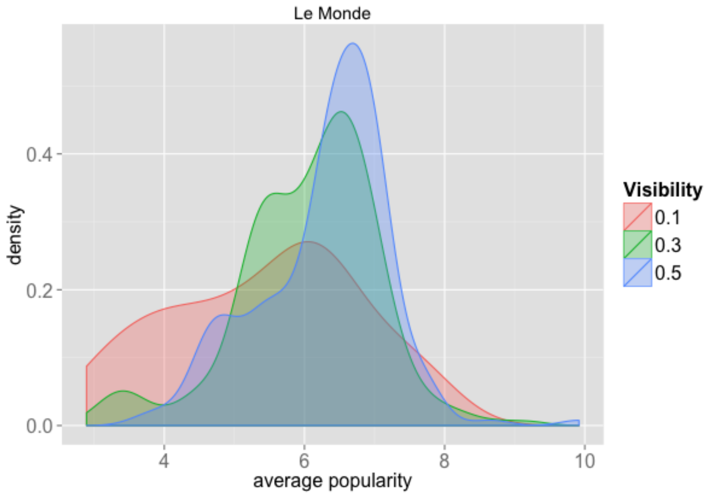


Figure 1.3: *Visibility and Popularity.*

ary regime popularity of a particular content could be expressed as a linear combination of other sources visibility. Therefore we assume that the popularity of a content associated with a particular source is related to what the other sources are doing. For instance, if two sources post contents about the same topic but with different information, it is interesting for a user to read both contents. In this case the visibility of one source has a positive influence over the popularity of the other source.

In chapter 4, we restrict the study of popularity on a unique content per source. Thus, in this case a differential equation is used to model the evolution of the popularity.

In chapter 5, we introduce a model that combines the previous approaches. As suggested in fig. 1.4, we model the popularity evolution of a source by a dynamical system which receives an impulse when a content is posted.

1.2.2 Control of popularity: an overview

In OSNs literature seven key points to control the improvement of content popularity have been identified:

- **Position in the news feed.** The simplest way of increasing popularity for a source is to improve the visibility of its contents. Visibility can be sustained by a source if its messages are always at the top of news feed. Indeed, in this case, to be in the top position allows a source's message to be seen by a user as soon as he logs onto the OSN[18]. There are three major ways of achieving high visibility. First, by paying OSNs [20], which can increase the probability of being in the

[18] http://www.nngroup.com/articles/scrolling-and-attention/

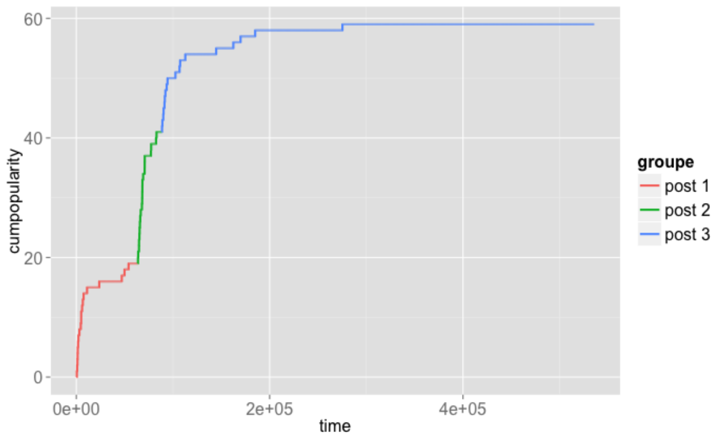


***Figure 1.4:** Evolution of the cumulative popularity when several contents are posted.*

top position[19]. We use this strategy in chapter 4. Secondly, by the quantity of messages posted [21] and thirdly by the design of OSNs, when the users can rank the post (as in Reddit.com).

- **The topic and the content of the post.** The choice of content and of topic of each message is crucial for the popularity of a post. In this paper [22] the authors predict the popularity of a message based on its topic. It is also interesting to understand the influence that different topics can have on each other. Is it a good strategy to post a content about "terrorism" followed by a content about "art"? In [23], the authors have illustrated and studied interactions between different "epidemics" in OSNs concerning different topics.

- **Network concern.** It is common knowledge that the popularity of a content depends on who is posting and its degree of influence in the OSN [24]. In this reference, the position in the network has been proved to be useful to determine the evolution of content popularity.

- **Payment.** A pure economic strategy can be adopted, where sources need to decide how much to pay to OSNs to be visible. There are different payment modalities [20], depending on the OSN and different Ad formats [21]. Most of the time auction mechanisms are used [25] to implement the competition between sources.

- **Headline.** The headline is known as an important feature to become popular. Indeed, the headline of content will be decisive for a reader to decide whether to read/watch a content or not. Authors of [13] create a model that predicts the best

[19] https://www.facebook.com/business/ads-guide/?tab0=Mobile%20News%20Feed
[20] https://support.google.com/adsense/answer/160525?hl=en
[21] https://support.google.com/youtube/answer/2467968?hl=en

headline for a message in the OSN reddit.com.

- **Timing.** The "when to post" question is an important and complex issue. It is not enough to give the right time as suggested in [26] because if every source posts at the same time, then it is unlikely that a source will be in the top position, it means that it will be less visible. Thus in [27], the authors study, in a game theory setting, what happens when sources have to decide when to post.

All these controls are to be taken with caution due to unexpectedness phenomenon. Indeed, because of the power law distribution of contents popularity (as illustrated in [28]), and the high value of the variance, it is hard to predict the virality of a content.

1.3 Prospection of posting behaviour in OSNs

Understanding the posting behaviour of sources is of main interest. For Internet Service providers, being able to understand traffic patterns will allow them to design efficient infrastructures. For the owner of the OSN service providers, it can help to enhance user experience by improving content filtering action. From companies perspective, understanding what drives topics tendency is useful for the understanding of the consumers' preferences. In this section, we explain the different approaches used to model the posting behaviour. Then we make a survey of the different methods developped to capture the sources' behaviour in OSNs. We will describe the approach used during the Ph.D. thesis to model the posting behaviour of sources.

Firstly we assume the following:

The actions in the OSN taken by the sources are driven by the search for popularity.

Of course if we restrict our framework to companies that use OSNs as a communication tool, it is reasonable that they try to become popular. But, we extend this assumption to other type of users. There is empirical evidence which shows that "classical users"[22] look for interaction with their community [29].

One of the main reasons why people will not post is because they think it will be boring for their friends, and because they will not receive enough attention from their community. There is also some empirical evidence that newspapers will cover stories that will be more popular [30]. We validate our hypothesis by a simple experiment, which measures the impact of popularity over posting behaviour. We can observe in fig. 1.5 the following phenomenon:

"The more I'm popular today, the more I will post tomorrow."

[22]that use OSNs without any commercial purpose

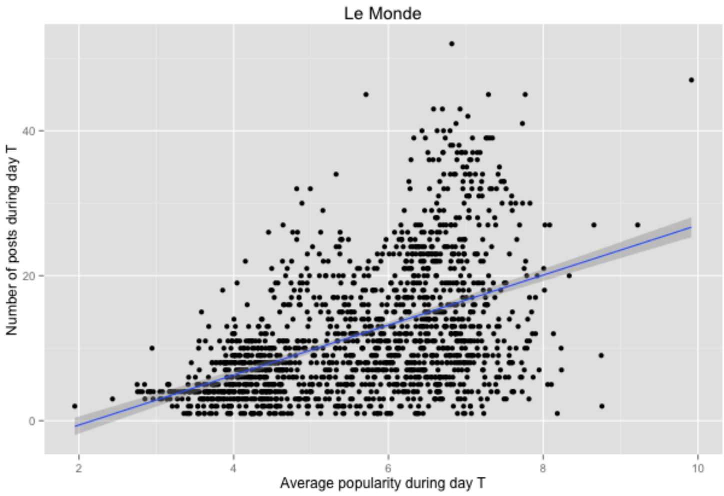


Figure 1.5: *Popularity vs. number of posts per day*

Secondly, we define a model where sources can directly send messages (according to a Poisson process) to users and users have a variable that measures the interest in content associated with each source. Moreover, each source can send messages to another source. In order to understand the equilibrium behaviour when several sources try to maximize their own popularity, the Nash equilibrium seems to be the right mathematical tool. Indeed, the Nash equilibrium concept [31], with all its extensions is often used in economics and telecommunications. A strategy vector (for instance a specific arrival rate of messages for each source) is a Nash equilibrium if no player can improve its utility by a unilateral deviation.

Finally, we propose a second model inspired from queuing theory. We assume that sources connect to OSN according to a Poisson process and will decide to post or not a message if they think that its messages will be popular enough. Thus this kind of model is inspired from queuing theory where users can decide to enter or not in a queue according to the observed size of it [32]. This model is a stochastic model because the arrival of users is random, but also because the popularity of content is a random variable. We use the framework of stochastic approximation in order to model this posting process.

We now provide an overview of previous works about posting behaviour in OSNs. The posting behaviour of users has been studied most of the time from a machine learning perspective. Several approaches can be used. The classical one is to identify the key features that determine, for instance, the number of messages posted each day. For instance, in [33] the authors show the importance of two features, the own intrinsic interest and the global temporal context. Yet, even if it seems an issue easy to address, a little reflection leads to the enumerations of different cases, which appear not to be linked together. For example, we examine the behaviour of Online Trolling. *"Online* *trolling is the practice of behaving in a deceptive, destructive, or disruptive manner in a social setting on the Internet with no apparent instrumental purpose."*[23]. In this case, it was proved that motivation behind trolling is fun and sadism [34]. Yet another example is about the users who post about themselves, who try to fight loneliness [35]. Finally, our last example is the one of companies, which use OSNs to make advertisements or to manage customer relations' [4]. This profusion of goals of sources shows how complex OSNs can be, and illustrates the fact that even the simplest questions can have hard answers in this context. However, we can notice that in each case, there is a search of reactions. Indeed, I will not do online trolling if users do not complain about it. I will not post information about myself in order to regulate my emotions if people do not like my posts [24]. Thus it seems that everyone is trying to be popular, to receive attention from their community. So popularity could be seen as an important driver to explain the posting behaviour of sources in OSNs.

1.4 Minimization of gossip propagation and maximization of contents diversity

The need to improve the posting behaviour of users is becoming a major issue. Indeed, the main dangers of OSNs is the propagation of false information.

This is the reason why sometimes, there is the need for a regulatory authority able to detect rumors [36] and to remove the source of the gossip [37]. For example, if we observe messages concerning the Boston Marathon (8 million relevant tweets concerning this topic[25]), it can perfectly illustrate the dissemination of misinformation. In the twenty most popular tweets, 30 % are false information. Another reason for controlling messages posted by users is to improve the quality of experience of each user. For instance, the regulatory authority can propose a safety net such that unwanted messages do not appear in the users' news feed [38]. The last illustration of the need of control is about helping the creation of newcomer contribution in OSNs. If you are a newcomer source it is more difficult to get popular, because other sources are already taking all the attention of readers. The authors, in [39], design a mechanism to increase the popularity of newcomer source, with the intervention of a regulatory authority. In this section we first describe how we will control the posting behaviour of users. Then we will present an overview of research about the improvement of the posting behaviour.

We propose a Content Active Filtering (CAF)[26]. We define two specific goals for

[23] Buckels, E. E., Trapnell, P. D., Paulhus, D. L. (2014). Trolls just want to have fun. Personality and individual Differences, 67, 97-102.

[24] http://goo.gl/qNBZAz

[25] http://www.smithsonianmag.com

[26] where the control is the probability to accept a message from a particular source

the CAF. We suggest using the CAF in order to increase the diversity of posted content by trying to provide an equal access to content for everyone. We are interested in the minimization of the propagation of a gossip by removing nodes. We assume that the regulatory authority can block some nodes in the Social Network. For instance, by sending a warning to the friends of a user, the control can disturb communication between users. The minimization of the gossip process is proved to be equivalent to a convex optimization problem.

CAF already exists in many OSNs. Indeed, in the case of Facebook, until 2011, EdgeRank was the name of the algorithm used by Facebook to determine what content should appear in a user's news feed. Usually CAF has a main objective which is the satisfaction of users. Papers [40], [38] propose a filter algorithm based on user profiles, without taking into account the overall popularity of the feed. However, none of these works were designed for OSN. In the paper [41], the authors introduce the popularity of contents in their proposal. To the best of our knowledge, this is the first time that a CAF is used to improve diversity and to decrease gossip.

1.5 Mathematical tools and mathematical contributions

This section is dedicated to the description of the mathematical tools used in this Ph.D. thesis.

1.5.1 Game theory

Firstly we applied and adapt, in chapter 2, the theory of *atomic congestion games with splittable stocks* [42] to visibility competition on OSNs. A congestion game is defined as follows: There is a finite number of players $\mathscr{J} \in \{1,\ldots,J\}$ and a finite number of links $\mathscr{C} \in \{1,\ldots,C\}$. Each player $j \in \mathscr{J}$ has a demand $\phi_j > 0$ to split between the links, where λ_{jc} is the proportion of ϕ_j send by player $j \in \mathscr{J}$ in link $c \in \mathscr{C}$. The total loss of a player j is defined as

$$\sum_{c=1}^{C} \lambda_{jc} D_c(\sum_{i=1}^{J} \lambda_{ic}),$$

where $D_c(\sum_{i=1}^{J} \lambda_{ic})$ is the loss associated with link $c \in \mathscr{C}$ which depends on the total flow $\sum_{i=1}^{J} \lambda_{ic}$. In chapter 1, we demonstrate that the total loss associated with a visibility competition is given by the following function:

$$\sum_{c=1}^{C} \lambda_{jc} \left[\gamma_{jc} - D_c (\sum_{d=1}^{C} \sum_{i=1}^{J} \lambda_{jd}) \right]. \tag{1.1}$$

It can be noticed that the loss in congestion game is slightly different than the loss associated with a visibility competition, and we cannot use some important properties as the concavity properties. However, in chapter 2, we managed to prove the uniqueness of the interior Nash equilibrium and we propose a learning algorithm that converges to the symmetric Nash equilibrium.

Secondly, in chapter 4, we use differential game theory [43], to model a competition between sources who wish to see their content become popular. The differential game theory comes from the optimal control theory [44].

1.5.2 Stochastic dynamical systems

Firstly we use the theory of stationary point process to model the number of messages posted the sources [45].

Secondly, in chapter 3, we use stochastic approximation theory [46] to model the posting behavior of users. A *stochastic approximation scheme* is given by:

$$\mathbf{x}(n+1) = \mathbf{x}(n) + a(n)[h(\mathbf{x}(n)) + \mathbf{M}(n+1)],\ n \geq 0, \tag{1.2}$$

with for each $n \geq 0$, $\mathbf{x}(n) \in \mathbb{R}^I$, $\mathbf{M}(n+1) \in \mathbb{R}^I$ is a martingale difference sequence with respect to the increasing family of σ-fields $\mathscr{F}(n) := \sigma(x_m, \mathbf{M}(n+1), m \leq n)$. This scheme can be seen as a noisy approximation of the following system of differential equation:

$$\frac{d\mathbf{x}(t)}{dt} = h(\mathbf{x}(t)),\ t \geq 0.$$

Moreover, the model in chapter 3 can be seen as a stochastic extension of the Degroot model [47; 48] with a signed matrix.

1.6 Organization of the document

Chapter 2 is based on two papers. The first one is "A Time and Space Routing Game Model applied to Visibility Competition on Online Social Networks" and the second one is "Game theory approach for modeling competition over visibility on social networks". First, we propose a visibility measure on a news feed. Next, we present our game based on visibility measures, which is similar to the Weighted Allocation Game. Following this consideration, we provide the uniqueness of the Nash Equilibrium, a characterization via concave programming, and its closed form. We propose a two time scale decentralized algorithm where each user estimates the total flow of messages in each News Feed and uses stochastic gradient algorithm to update their own flows. Finally, we propose an extension of our model where a sharing strategy is added.

Chapter 3 is based on the paper "Posting behaviour in Social Networks and Content Active Filtering". In this chapter, we have two objectives: First we model the posting behaviour of publishers in OSNs 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.

Chapter 4 is based on the paper "Differential Games of Competition in Online Content Diffusion". In this chapter, we use the theory of differential game to model the competition between a source who pay the OSN to get their content visible. The cumulative popularity of a content is modeled by using a differential equation and we use Hamilton Jacobi Equations to compute the symmetric equilibrium. We prove that a bang-bang policy is a Nash equilibrium.

Chapter 5 is based on the paper "Controlling the Katz-Bonacich Centrality in Social Network: Application to gossip in Online Social Networks" and on "A Continuous Optimization Approach for Maximizing Content Popularity in Online Social Networks". In this chapter, we consider two optimization problems. The first one is related to the optimization of the Katz-Bonacich centrality using a topological control which is called the Katz-Bonacich optimization problem. We first prove that this problem is equivalent to a linear optimization problem. Thus, in the context of large graphs, we can use algorithms such as Simplex algorithm. We provide a specific applications of the Katz-Bonacich centrality minimization problem based on the minimization of gossip propagation and make experiments on real networks. The second section of the chapter is concerned with popularity optimization in OSN. We model the evolution of content popularity on Online Social Networks considering the following property: most of the popularity (influence, likes, number of views) of a content is obtained while the content is in the first position in the news feed of an OSN. This property is checked on a real dataset and therefore a mathematical optimization model of the cumulative popularity is proposed. We consider the popularity maximization problem in which a source of contents controls the topics of its contents and the number of posts in order to maximize its overall popularity. We demonstrate that the popularity maximization problem can be written as a "generalized" fractional program and we provide an algorithm that converges to the optimal solution.

In chapter 6, we give the conclusion of our Ph.D. thesis and we describe possible perspectives.

Chapter 2

Visibility Competition on Online Social Networks

Contents

This chapter is based on two papers *A Time and Space Routing Game Model applied to Visibility Competition on Online Social Networks* and *Game theory approach for modeling competition over visibility on social networks*.

2.1 Introduction

Lately, new questions concerning OSNs and the visibility of its contents are attracting the attention of researchers. For each user or company that wants to become popular and visible in OSNs, answers to the following questions seem to be essential: Where and when should messages be posted? On what topic? How many messages should a company post? In this chapter we consider that these questions have to be treated following a competitive scenario, since there are several users/companies aiming to become popular on a common OSNs. Routing game theory [49] seems to be an interesting tool to answer these questions. So far, to our knowledge, only one paper has adopted the routing game theory to solve problems concerning OSNs [15].

We propose a model where the sources control the flow of messages sent in a news feed and try to maximize the visibility of their posted messages. We assume that content visibility in a news feed is positively correlated with content popularity. In the case of Facebook, a *news feed* is a Feed where the user's friends are published. These posts are displayed in chronological order from the newest to the oldest. It is possible to find in the introduction of [21] more definitions of a news feed.

The reason why we use Routing game theory is because, for a source, the maximization of message visibility is equivalent to the minimization of a delay function. Routing game theory supposes a game where players have to decide where they send their flows of data. In this sense, the player's flow has the *splittable property*. In OSNs sources can choose to post messages in different news feeds. Another property of routing games is the *congestion property*. Mathematically speaking, this property can be explained by the fact that the utility of a player on a link decreases in the total amount of flow sent on this link. This congestion property is also found in OSNs, when the visibility of a message decreases on the total flow of messages on this OSN.

In section 2.2, we present the latest related works concerning game theory applied to competition over visibility in OSNs. Then, in the case of our model, we present the mathematical models that are similar to ours.

In this chapter, we consider that sources can post in several news feeds, and concerning different topics. In section 2.3, we propose a visibility measure and we also develop the routing game associated.

There is a difference between the routing games defined in this chapter and classical routing games [49]. This is why we adapt the mathematical framework of [49] to the

Routing game defined in this chapter. We develop it in section 2.4. First, we demonstrate the existence of a Nash equilibrium. Then, we prove that at the Nash equilibrium, under some hypothesis, a source only sends messages about a unique topic per news feed. Moreover, we give a characterization of the Nash equilibrium based on Concave Programming and we prove its uniqueness.

Section 2.5 deals with more practical issues. Indeed, the routing game theory developed in [50] provides existence and uniqueness results about the Nash equilibrium. However, it does not propose realistic decentralized learning algorithm that converges to it. However, based on the results of 2.4 , we are able to provide the convergence of a decentralized mechanism to the Nash equilibrium.

Finally, in section 2.6, we extend the model to study rerouting of posts, which corresponds to sharing content.

2.2 Related works

2.2.1 Related models

Studies concerning users/companies that want to become visible on OSN news feeds have been recently treated in [50] and [21]. This is not the usual way of modeling competition over visibility on OSNs. Usually there are two approaches. The first one is the advertising allocation. In this case, sources can choose to pay OSNs in order to get their content promoted. Papers about this subject are often linked to marketing differential games [51; 52; 53]. The second one is to consider an epidemic process on an OSN. In this case, the question would concern which node a source would need to use in order to maximize the spread of its contents [54; 55; 56].

In [21] the authors use a Poisson process to model the flow of messages on a unique news feed. However, it can be expected that the Poisson process assumption does not fit with a realistic scenario. That's the reason why we model the flow of messages of each source by a Stationary Point Process, as it is proposed in [50]. This model would allow sources to post on different news feeds and concerning different topics.

2.2.2 Related mathematical results

Under some hypothesis, and from the fact that at equilibrium player's messages concern only one topic per news feed, the game that we study in this chapter is similar to the game defined in [57]. This game is called *Weighted Allocation game* and could be defined as follows: Let $\{K_1,\ldots,K_L\}$ be a set of resources that is shared among J strategic users. Each user proposes a bid $\lambda_{jl} \in [0,\overline{\lambda}]$ for each resource l. Each resource is

Table 2.1: Main notations used throughout the chapter 2 until section 2.6

Symbol	*Meaning*
J	number of sources
C	number of topics
N_{jc}	stationary point process that counts the number of messages posted by source j about topic c
λ_{jc}	intensity of the stationary point process N_{jc}; $\lambda_j := (\lambda_{j1}, \ldots, \lambda_{jC})$ $\lambda_{jc} \in [0, \overline{\lambda}]$, with $\overline{\lambda} < \infty$
γ_{jc}	preference of user j over visibility of topic c; $\gamma_j^* = \max_c \{\gamma_{jc}\}$
γ_0	linear cost associated with the creation of messages; $\Gamma := \sum_{i=1}^{J} \frac{\gamma_0}{\gamma_i^*}$

assumed to be shared proportionally to the bids, so that the amount of resource l that player j receives is given by:

$$K_l \frac{\lambda_{jl}}{\sum_i \lambda_{il}},$$

and each player j has a linear cost γ_0 for his bid. Finally, his objective function is given by:

$$U(\lambda_j, \lambda_{-j}) = \sum_l \gamma_l \frac{\lambda_{jl}}{\sum_i \lambda_{il}} - \gamma_0 \sum_l \lambda_{jl},$$

where $\lambda_j := (\lambda_{j1}, \ldots, \lambda_{jL})$ and $\lambda_{-j} := (\lambda_1, \ldots, \lambda_{j-1}, \lambda_{j+1}, \ldots, \lambda_J)$. The authors, in [57], prove the existence of the Nash equilibrium. Also, they study efficiency measurements and they provide learning algorithms concerning the Nash equilibrium. The utility function in this chapter does not have the same mathematical form as the one of the Weighted Allocation game, indeed, in the game we consider, the utility is:

$$U(\lambda_j, \lambda_{-j}) = \frac{\sum_l \gamma_{jl} \lambda_{jl}}{\sum_{i,l} \lambda_{il}} - \gamma_0 \sum_l \lambda_{jl}.$$

In our case, we prove the uniqueness and the characterization of the equilibrium based on the routing game model. Concerning the learning algorithm, considerable works have been done to provide learning algorithm in Allocation games [57; 58; 59; 60; 61] and in routing games [49; 62]. We propose a new algorithm based on a decentralized scheme proposed in [63].

2.3 Model

The main symbols used in the chapter, until section 2.6 are reported in Tab. 2.1.

2.3.1 Creation of a visibility measure on a news feed

In this section, we define a measure of visibility of a source j, $j \in \{1,\dots,J\}$ in a news feed. A news feed in an OSN is a continuous transmission of data in the user's personal account. The content of this data comes from other users and pages that the user choses to be subscribed to. We consider that the messages in a news feed of users are ordered from the newest to the oldest. This type of news feed configuration can be found in the most popular OSNs, like Twitter, Facebook and Tumblr. In order to model the visibility of a content in a news feed, we uses the theory of Point Processes [45]. As a first step, each source j is associated with a stationary point process N_j which represents the messages sent by source j in the news feed. With the help of this model, a first visibility measure of the content of j is proposed. Then we distinguish messages by topic and our final measure of visibility of a source j is given.

Since most web users give more attention to the information above the page fold[1], we focus our visibility measure on the first message of the news feed. The arrival time of any message in the news feed of users is modeled by a stationary point process (N,θ_t,P), where θ_t is the associated flow, P the associated probability measure and $\lambda := E[N(0,1]]$ the intensity of the process with $0<\lambda<\infty$. T_n denotes the arrival time of the n^{th} message. We define the sequence of mark $\{Z_n\}_n$ associated with (N,θ_t,P) with $Z_n \in \{1,\dots,J\}$ for all n. Z_n tell us the origin of the n^{th} message. We define the continuous process $Z(t)$ where $Z(t)=Z_n$ if $t\in(T_n,T_{n+1}]$. We define a new process N_j:

$$N_j((a,b]) := \sum_{n\in\mathbb{Z}} 1_{\{Z_n=j\}} 1_{\{T_n\in(a,b]\}}. \tag{2.1}$$

This process counts the number of messages from j during the interval $(a,b]$. We assume that the intensity of this new process is λ_j, with $0<\lambda_j<\infty$. We consider that the visibility measure of a source j at time t is given by:

$$\lim_{h\to 0} P(Z(T_+(t)) = j \mid N((t,t+h]) \geq 1) \tag{2.2}$$

where $T_+(t) = t + inf\{h \mid N((t,t+h]) = 1\}$.

This measure is the probability that, when a message arrives in the news feed, it comes from source j. The next proposition gives an explicit formula for this probability based on palm probability theory.

Proposition 2.1. *For all t and for all j,*

$$\lim_{h\to 0} P(Z(T_+(t)) = j \mid N((t,t+h]) \geq 1) = \frac{\lambda_j}{\sum_i \lambda_i}. \tag{2.3}$$

[1] http://www.nngroup.com/articles/scrolling-and-attention

The fact that we restrict our measure to arrival time allows us to have an explicit form for any stationary point process. Moreover, the visibility measure of the source j gives the proportion of messages that comes from j in all the news feed content.

We next consider the topics treated in messages. We propose that a message can give information about a topic c, $c \in \{1,\dots,C\}$. We must therefore define for each c a point process (N_c, θ_t, P_c) of intensity λ_c. We assume that each (N_c, θ_t, P_c) and $(N_{c'}, \theta_t, P_c)$ are independent. This is why by defining Z_c, N_{jc} and λ_{jc} as previously, we can prove that:

$$\lim_{h\to 0} P(Z_c(T_+(t)) = j \mid N((t,t+h]) \geq 1) = \frac{\lambda_{jc}}{\sum_i \sum_{c'} \lambda_{ic'}}. \tag{2.4}$$

We consider that $V_{jc} := \dfrac{\lambda_{jc}}{\sum_i \sum_{c'} \lambda_{ic'}}$ is a measure of the visibility of contents posted by j about topic c.

2.3.2 Game model applied to OSN

Due to the presence of many sources wishing to be visible in OSNs news feeds, we define a non-cooperative game between them. Firstly, the strategies of sources is described. Secondly, the utility of each player is given. We propose the utility of a player as the weighted sum of measures of visibility.

Source's Strategies: The strategy vector of the source j is $\lambda_j = (\lambda_{j1},\dots,\lambda_{jC})$. We assume that a source can only send an expected limited number of messages per time unit. Thus, for each j and each c, $\lambda_{jc} \in [0, \overline{\lambda}]$.

Source's Utility: In order to define the utility of a source, we recall the visibility measure vector of j is given by:

$$\mathbf{V}_j = \Big(\frac{\lambda_{j1}}{\sum_i \sum_{c'} \lambda_{ic'}}, \dots, \frac{\lambda_{jC}}{\sum_i \sum_{c'} \lambda_{ic'}}\Big). \tag{2.5}$$

Each source j has a linear cost given by $\lambda_{cost} := \gamma_0 \sum_{c'} \lambda_{jc'}$ associated with the creation of messages. The utility of source j is a weighted sum of the visibilities measures of j minus its cost:

$$U_j(\lambda_j, \lambda_{-j}) = \left[\frac{\sum_c \gamma_{jc}\lambda_{jc}}{\sum_i \sum_{c'} \lambda_{ic'}}\right] - \gamma_0 \sum_{c'} \lambda_{jc'}, \tag{2.6}$$

where $\lambda_{-j} := (\lambda_1, \dots, \lambda_{j-1}, \lambda_{j+1}, \dots, \lambda_J)$. In the next section, our interest turns to study the game that we defined previously.

2.4 Game study

In this section, we study the properties of Nash equilibrium. To start with, we provide some characterizations of it. Next, we prove that at any Nash equilibrium, each player sends messages about a unique topic. At this point we characterize Nash equilibrium via Concave Programming. Finally, the uniqueness of the Nash equilibrium is proved, a closed form is provided and the price of anarchy is studied.

The concept of Nash equilibrium (NE) is defined as follows:

Definition 2.2. $(\lambda_1^{NE}, \ldots, \lambda_J^{NE})$ *is a NE iff for all j:*

$$U_j(\lambda_j^{NE}, \lambda_{-j}^{NE}) = \max_{\lambda_j \in [0, \overline{\lambda}]} U_j(\lambda_j, \lambda_{-j}^{NE}). \tag{2.7}$$

In other words, a source has no interest in changing his intensity unilaterally at the Nash equilibrium. The existence of at least one NE is not clear. The utility function $U_j(\cdot, \lambda_{-j})$ is not concave in λ_j. However, this game has several important properties. Firstly, at the NE at least one source sends messages in the news feed as it is stated in the next proposition.

Proposition 2.3. *The decision vector* $(0, \ldots, 0)$ *is not a Nash equilibrium.*

Secondly, we prove that if a source for a specific OSN has no equal preferences on two different topics then it only sends messages that concern the unique topic that it prefers the most. We first impose the following assumption:

(H1) There exists a NE such that for all j and c, $\lambda_{jc}^{NE} < \overline{\lambda}$.

The assumption **(H1)** characterizes situations where an equilibrium exists such that any source cannot saturate its upper constraints. The next proposition states the expected result:

Proposition 2.4. *Consider a source j, two topics* (c, c') *and* $\gamma_{jc} > \gamma_{jc'}$. *Then for any* λ_{jc}^{NE} *and* $\lambda_{jc'}^{NE}$ *that verified **(H1)**,*

$$\lambda_{jc}^{NE} = 0, \; \lambda_{jc'}^{NE} = 0 \tag{2.8}$$

or

$$\lambda_{jc}^{NE} > 0, \; \lambda_{jc'}^{NE} = 0. \tag{2.9}$$

We define a new hypothesis:

(H2) For each source j and topics c and c', we have $\gamma_{jc} \neq \gamma_{jc'}$.

According to the previous proposition, when a source maximize its visibility it is optimal for him to send messages about a unique topic, which is harmful in terms of content diversity in OSNs. More precisely, with the help of **(H2)** we have the following corollary.

Corollary 2.5. *Consider that **(H1)** and **(H2)** are satisfied. Then for each j there exists a unique $c_j \in \{1,\dots,C\}$ such that $\gamma_j^* = \max_{c'}\{\gamma_{jc'}\}$. Therefore $c_j = \text{argmax}_{c'}\{\gamma_{jc'}\}$. Then for j and l associated with c, $\lambda_{jc_j}^{NE} \geq 0$ and for all $c' \neq c_j$, $\lambda_{jc'}^{NE} = 0$. In this case the payoff of a player j can be rewritten in the following manner:*

$$U_j(\lambda_j,\lambda_{-j}) = \left[\frac{\gamma_j^*\lambda_j}{\sum_i \lambda_i}\right] - \gamma_0\lambda_j. \tag{2.10}$$

The previous corollary provides assumptions such that the game defined in this chapter is equivalent to the generalized Kelly mechanism [64]. With the help of corollary 2.5, $U(\cdot,\lambda_{-j})$ is concave in $\lambda_{jc_j} \in [0,\overline{\lambda}]$ and so the existence theorem from [65] can be applied. Moreover, $\lambda^{NE} > 0$ is a Nash equilibrium if and only if for each j, $\frac{\partial U_j}{\partial \lambda_j}(\lambda_j^{NE},\lambda_{-j}^{NE}) = 0$. This game has been well-studied, however, we propose deriving new results about uniqueness by using tools from routing game theory. The next theorem provides a new characterization of Nash equilibrium. We prove that, at a Nash equilibrium, the sum of messages flow, in other words $\sum_i \lambda_{ic_i}^{NE}$, is the unique solution of a concave optimization problem.

Theorem 2.6. *Consider that **(H1)** and **(H2)** are satisfied. If $\lambda^{NE} = (\lambda_1^{NE},\dots,\lambda_J^{NE}) > 0$ is a NE then $\sum_i \lambda_{ic_i}^{NE}$ is the unique solution of the following concave optimization problem:*

$$\max_{\Lambda\in[0,J\overline{\lambda}]} (J-1)\log(\Lambda) - \Gamma\Lambda, \tag{2.11}$$

with $\Gamma = \sum_i \frac{\gamma_0}{\gamma_i^*}$.

This new characterization is fundamental to prove the main result of this chapter. With it we prove the uniqueness of the interior Nash equilibrium because $\lambda_{jc_j}^{NE}$ is uniquely defined by $\lambda_l^{NE} := \sum_i \lambda_{ic_i}^{NE}$.

Theorem 2.7. *Consider that **(H1)** and **(H2)** are satisfied. Let $\lambda_1^{NE} > 0$ and $\lambda_2^{NE} > 0$ be two Nash equilibria. Under these assumptions, for all j*

$$\lambda_1^{NE} = \lambda_2^{NE}.$$

In the next proposition, a closed form of the equilibrium is given.

Proposition 2.8. *Consider that **(H1)** and **(H2)** are satisfied. If there exists a Nash equilibrium such that $\lambda^{NE} > 0$, and if for all j,*

$$\gamma_j^* > \frac{(J-1)}{(\sum_i \frac{1}{\gamma_i^*})}, \tag{2.12}$$

then

$$\lambda_{jc_j}^{NE} = \frac{(J-1)}{\gamma_0(\sum_i \frac{1}{\gamma_i^*})}\left[1 - \frac{1}{\gamma_j^*}\frac{(J-1)}{(\sum_i \frac{1}{\gamma_i^*})}\right]. \tag{2.13}$$

Moreover

$$U_j(\lambda_j^{NE}, \lambda_{-j}{}^{NE}) = (1 - \frac{\gamma_0}{\gamma_j^*}\frac{(J-1)}{(\sum_i \frac{1}{\gamma_i^*})})(\gamma_j^* - \frac{(J-1)}{\gamma_0(\sum_i \frac{1}{\gamma_i^*})}).$$

In the previous theorem, there is a condition (2.12) over the cost γ_j^* all j such that $\lambda_{jc_j} > 0$ for all j. If this condition is not satisfied, intensities can be equal to 0 at equilibrium for some sources. This result is described in the next theorem. Given the explicit expression of the Nash equilibrium of the game, we are able to understand better the behaviour of the equilibrium when the parameters of the game change. Thus from proposition 2.8, it can be noticed that for all source j, the intensity λ_{jc_j} is strictly increasing in γ_j^* and quadratic concave in J.

Thus we can now compute a general form of the equilibrium and not just restrict ourselves to an equilibrium where $\gamma_j^* > \frac{(J-1)}{(\sum_i \frac{1}{\gamma_i^*})}$, in other words when all λ_j^{NE} are positive. We rearrange the sources from the smallest $\frac{\gamma_0}{\gamma_j^*}$ to the largest, such that $\frac{\gamma_0}{\gamma_1^*} > \frac{\gamma_0}{\gamma_2^*} > \ldots > \frac{\gamma_0}{\gamma_J^*}$.

Theorem 2.9. *Assume that j' is such that $j' = max\{j \mid \gamma_j^* < \frac{(J-1)}{(\sum_{i=j}^J \frac{1}{\gamma_i^*})}\}$ i.e j' is the source with the largest $\frac{\gamma_0}{\gamma_j^*}$ who does not send any messages in the news feed. In this case the Nash equilibrium is uniquely defined by:*

$$\begin{cases} \lambda_{jc_j}^{NE} = \dfrac{(J'-1)}{\gamma_0(\sum_{i=j'+1}^J \frac{1}{\gamma_i^*})}\left[1 - \dfrac{1}{\gamma_j^*}\dfrac{(J'-1)}{(\sum_{i=j'+1}^J \frac{1}{\gamma_i^*})}\right], \\ \lambda_{jc_j}^{NE} = 0, \qquad \forall j \le j', \end{cases} \tag{2.14}$$

where $J' = J - \mid \{1, \ldots, j'\} \mid$.

After having proved the uniqueness of the Nash equilibrium and characterizing the NE, the next section is dedicated to a learning mechanism that converges to the symmetric Nash equilibrium.

2.5 Learning algorithm for symmetric sources

In this section we study a decentralized algorithm that converges to the NE given previously.The game does not admit a potential function in the sense of [66], and does not verify the submodularity property [67] for $J > 2$ and $L > 2$. Thus, we cannot use conventional tools of game theory to find out decentralized learning algorithm. Finding a decentralized algorithm is a well-known challenging issue in the context of routing game [42]. First, with the help of the theorem 2.6 and the proposition 2.8, we are able to design a singularly perturbated algorithm where:
- Each source imitates the average behaviour of all sources on each news feed,
- The source uses a gradient scheme to compute its optimal strategy.
This decentralized scheme has already been used in [63] in the case of distributed deterministic and stochastic gradient optimization algorithms. We are able to prove the convergence of this decentralized algorithm to the interior equilibrium which is, in the symmetric case the Nash equilibrium.

We assume throughout this section that **(H1)** and **(H2)** are satisfied. From **(H2)** and corollary 2.5, we restrict the learning algorithm to only one topic per news feed. Moreover we assume that for each j and c, $\lambda_{jc} \in [\underline{\lambda}, \overline{\lambda}]$. Let $t \in \mathbb{R}$ be the time of update of the decentralized learning algorithm. Let $\lambda_{jc_j}(t)$ be the flow of messages sent by player j at time t. Let $\varepsilon \in \mathbb{R}_+$ be the perturbation parameter.

$$\frac{d\lambda_{jc_j}(t)}{dt} = \left[\underbrace{\frac{1}{J}\sum_i \lambda_{ic_i}(t) - \lambda_{jc_j}(t)}_{Imitation} + \varepsilon \underbrace{\frac{\partial U_j}{\partial \lambda_j}(\lambda_{jc_j}(t), \lambda_{-j}(t))}_{gradient} \right]_{\underline{\lambda}}^{\overline{\lambda}} \tag{2.15}$$

$$= \left[\frac{1}{J}\sum_i \lambda_{ic_i}(t) - \lambda_{jc_j}(t) + \varepsilon \left(\frac{\gamma_j^* \sum_{i \neq j} \lambda_{ic_i}(t)}{(\sum_i \lambda_{ic_i}(t))^2} - \gamma_0 \right) \right]_{\underline{\lambda}}^{\overline{\lambda}} \tag{2.16}$$

(2.17)

Theorem 2.10. *Consider that* ***(H1)*** *and* ***(H2)*** *are satisfied. Then* $\forall \varepsilon << 1$

$$\lambda_{jc_j}(t) \underset{t \to +\infty}{\to} \lambda_{jc_j}^* \ \forall j \tag{2.18}$$

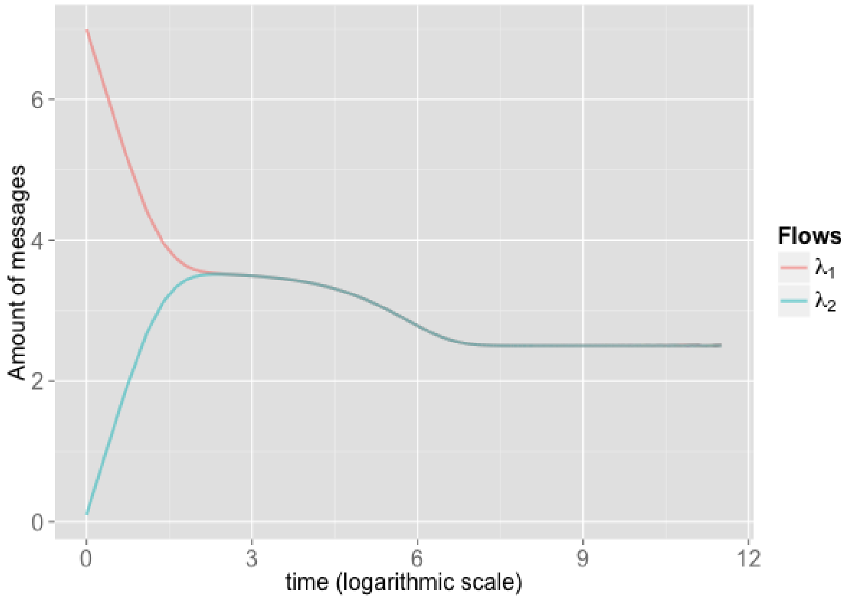


Figure 2.1: *Evolution of $\lambda_{1c_1}(t)$ and $\lambda_{2c_2}(t)$.*

Table 2.2: *Main notations used throughout section 2.6*

Symbol	*Meaning*
J	number of sources
L	number of news feeds
N_{jl}	stationary point process that counts the number of messages posted by j in news feed l
λ_{jl}	intensity of the stationary point process N_j^l; $\lambda_j := (\lambda_j^1, \dots, \lambda_j^L)$ $\lambda_{jl} \in [0, \overline{\lambda}]$, with $\overline{\lambda} < \infty$
P	sharing matrix where the ll'-entry is denoted by $p_{ll'}$
γ_0	linear cost

where $\lambda^*_{jc_j} = \frac{1}{J}\sum_i \lambda^*_{ic_i}$ *and* $\sum_i \lambda^*_{ic_i} = \frac{\Gamma}{(J-1)}$ *where* $\Gamma := \gamma_0 \sum_i \frac{1}{\gamma_i^*}$. *Moreover if for each j and i, $\gamma_j^* = \gamma_i^*$, then $\lambda^*_{jc_j} = \lambda^{NE}_{jc_j}$.*

In fig. 2.1 we illustrate the convergence of (2.15) with two symmetric players and two News Feeds. We impose $\varepsilon = 0.1$, $\gamma_1^* = \gamma_2^* = 1$ and $\gamma_0 = 0.1$. As expected, we can observe that the trajectories of λ_{1c_1} and λ_{2c_2} follow quickly the same behaviour and then we can see the convergence of the whole system.

2.6 Network extension

The main symbols used in this section are reported in Tab. 2.2. We consider a generalization of our model to take into account the propagation effect between several news feeds. In fact, some messages posted on one news feed can be relayed (propagated) to another news feed. We assume that posting a message on a news feed has a cost for the source, whereas if a message is relayed, it is free for the source. In this section, we study this topological effect on the non-cooperative game between the sources. The described scenario is illustrated in fig. 2.2.

We assume that there are L news feed and $l \in \{1,\dots,L\}$ denotes a news feed index. Each source $j \in \{1,\dots,J\}$ can send messages to each news feed l. It means that each source j controls L point processes, which represent flow of messages of source j in each news feed, with intensity $\lambda_{jl} = E[N_{jl}((0,1])]$. We assume that $\lambda_{jl} \in [0,\overline{\lambda}]$ for each j and l. For each news feed l, any new message is copied and posted, with a probability $p_{ll'}$, to the other news feed $l' \neq l$. This process defines, for each source j, a new point process in each news feed l with intensity X_{jl} which is the solution of the following linear system:

$$X_{jl} = \lambda_{jl} + \sum_{l'=1,\, l'\neq l}^{L} p_{l'l} X_{jl'},\ \forall j,\ l. \tag{2.19}$$

We denote $X_j = (X_{j1},\dots,X_{jL})$ for each j. For each l, we assume that $\sum_{l'=1,\, l'\neq l}^{L} p_{ll'} < 1$. By applying similar analysis as in section 2.3, we obtain the following closed-form expression of the probability that when a message arrives in news feed l, it is coming from source j:

$$\lim_{h\to 0} P(Z_l(T_+(t)) = j \mid N(t,t+h] \geq 1) = \frac{X_{jl}}{\sum_i X_{il}}. \tag{2.20}$$

Let γ_0 be the cost associated with source j for sending one message to news feed l. We denote the flow vector of source j by $\lambda_j = \{\lambda_{j1},\dots,\lambda_{jL}\}$. The objective function of each source j is:

$$U_j(\lambda_j,\lambda_{-j}) = \sum_l \frac{X_{jl}}{\sum_i X_{il}} - \gamma_0 \sum_l \lambda_{jl}, \tag{2.21}$$

where for each j and l, X_{jl} is solution of:

$$X_{jl} = \lambda_{jl} + \sum_{l'=1,\, l'\neq l}^{L} p_{l'l} X_{jl'},\ \forall j,\ l. \tag{2.22}$$

This game is a symmetric game and we define $X_l^{NE} := X_{jl}^{NE}$ the symmetric equilibrium for each source j and each news feed l.

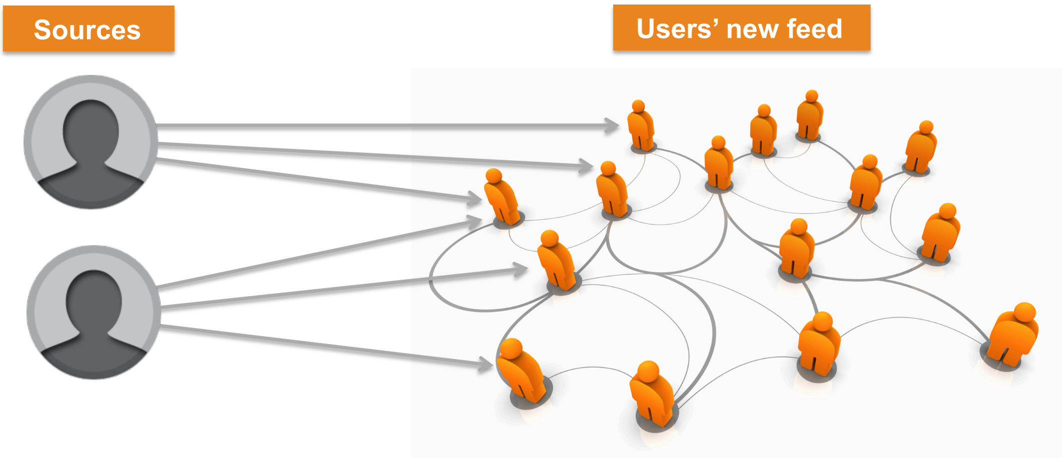


Figure 2.2: *Sources sending messages to different news feeds.*

Proposition 2.11. *Consider that there exists a Nash equilibrium such that for each j and l,*

$$\frac{J-1}{J^2}\left[\frac{1}{\gamma_0(1-\sum_{l'\neq l} p_{ll'})} - \sum_{l'} \frac{p_{l'l}}{\gamma_0(1-\sum_{l''\neq l'} p_{l'l''})}\right] \in (0,\overline{\lambda}),$$

then

$$\lambda_{jl}^{NE} = \frac{J-1}{J^2}\left[\frac{1}{\gamma_0(1-\sum_{l'\neq l} p_{ll'})} - \sum_{l'} \frac{p_{l'l}}{\gamma_0(1-\sum_{l''\neq l'} p_{l'l''})}\right]. \tag{2.23}$$

Moreover

$$U_j(\lambda_j^{NE}, \lambda_{-j}^{NE}) = \frac{L}{J^2}. \tag{2.24}$$

2.7 Conclusion

In this chapter we have defined a routing game where sources have to decide how many messages they have to post and the topic of their messages. Nash equilibrium uniqueness, an explicit form and a characterization via concave programming is provided. Also a decentralized algorithm that converges to the NE was proposed. Finally, we extend our work to multiple news feeds, where the sharing phenomenon is modeled.

APPENDIX

Proof of proposition 2.1
Consider $j \in \{1,\ldots,J\}$ and $t \in \mathbb{R}$ then,

$$P_N^0(Z(0)=j) = \lim_{h\to 0} P(Z(T_+(t)) = j \mid N(t,t+h] \geq 1)$$

where $P_N^0(Z(0)=j) = \dfrac{1}{\lambda(b-a)} E\left[\sum_{n\in\mathbb{Z}} 1_{Z_n=j} 1_{T_n\in(a,b]}\right]$. P_N^0 is called the Palm probability of the marks Z_n. The first equality comes from the local interpretation of Palm probability (p.40 [45]). From the definition of the intensity of N_i, it is possible to write:

$$\lambda_j(b-a) = E\left[\sum_{n\in\mathbb{Z}} 1_{Z_n=j} 1_{T_n\in(a,b]}\right] = \lambda(b-a)P_N^0(Z(0)=j).$$

The first equality comes from the definition of the intensity of N_j and the second equality comes from the definition of $P_N^0(Z(0)=j)$. From the definition of N_j we have $\lambda = \sum_i \lambda_i$. Then,

$$P_N^0(Z(0)=j) = \frac{\lambda_j}{\sum_i \lambda_i}.$$

Proof of proposition 2.3
If all sources $\{1,\cdots,J\} - \{j\}$ play 0, j can always find a very small λ_{jc}^{NE} such that $\lambda_j^{NE} < \dfrac{\gamma_{jc}}{\gamma_0}$. Then, we have that $U_j(\lambda_j^{NE}, 0) > 0$ and the decision vector $(0,\ldots,0)$ is not a Nash equilibrium.

Proof of proposition 2.4
Let $j \in \{1,\ldots,J\}$ be a fixed source. We assume the existence of $(c,c') \in \{1,\ldots,C\}^2$ such that $\lambda_{jc'}^{NE} > 0$, $\lambda_{jc}^{NE} > 0$. Then $\lambda_{jc'}^{NE}$ and λ_{jc}^{NE} are solution of the first order necessary condition:

$$\begin{cases} \dfrac{U_j}{\partial \lambda_{jc'}}(\lambda_j^{NE}, \lambda_{-j}^{NE}) & = & 0 \\ \dfrac{U_j}{\partial \lambda_{jc}}(\lambda_j^{NE}, \lambda_{-j}^{NE}) & = & 0 \end{cases} \tag{2.25}$$

$$\Leftrightarrow \begin{cases} \dfrac{\gamma_{jc'}}{\sum_i \sum_{c''} \lambda_{ic''}^{NE}} - \dfrac{\sum_{c''}\gamma_{jc''}\sum_{c''}\lambda_{jc''}^{NE}}{(\sum_i \sum_{c''} \lambda_{ic''}^{NE})^2} - \gamma_0 & = & 0 \\ \dfrac{\gamma_{jc}}{\sum_i \sum_{c''} \lambda_{ic''}^{NE}} - \dfrac{\sum_{c''}\gamma_{jc''}\sum_{c''}\lambda_{jc''}^{NE}}{(\sum_i \sum_{c''} \lambda_{ic''}^{NE})^2} - \gamma_0 & = & 0 \end{cases} \tag{2.26}$$

$$\Leftrightarrow \frac{\gamma_{jc'}}{\sum_i \sum_{c''} \lambda_{ic''}^{NE}} - \frac{\sum_{c''}\gamma_{jc''}\sum_{c''}\lambda_{jc''}^{NE}}{(\sum_i \sum_{c''} \lambda_{ic''}^{NE})^2} - \gamma_0 = \frac{\gamma_{jc}}{\sum_i \sum_{c''} \lambda_{ic''}^{NE}} - \frac{\sum_{c''}\gamma_{jc''}\sum_{c''}\lambda_{jc''}^{NE}}{(\sum_i \sum_{c''} \lambda_{ic''}^{NE})^2} - \gamma_0, \tag{2.27}$$

implies $\gamma_{jc} = \gamma_{jc'}$ which contradicts $\gamma_{jc} > \gamma_{jc'}$. So by contradiction the proposition follows.

Proof of theorem 2.6
From the sufficient optimality condition and from the fact that for all j, $\overline{\lambda} > \lambda_{jc_j}^{NE} > 0$, $\lambda_{jc_j}^{NE}$ is solution of

$$\left[\frac{1}{\sum_i \lambda_{ic_i}^{NE}} - \frac{\lambda_{jc_j}^{NE}}{(\sum_i \lambda_{ic_i}^{NE})^2}\right] - \frac{\gamma_0}{\gamma_j^*} = 0, \ \forall j. \tag{2.28}$$

Then by taking the sum over j in (2.28) and by defining $\Lambda = \sum_i \lambda_{ic_i}$ and $\Gamma = \sum_i \frac{\gamma_0}{\gamma_i^*}$ we get:

$$\frac{J-1}{\Lambda^{NE}} - \Gamma = 0, \tag{2.29}$$

which are the KKT condition of (2.11).

Proof of theorem 2.7
From the first order optimality condition and from the fact that for all j, $\lambda_{jc_j}^{NE} > 0$, we can deduce that for all j

$$\lambda_{jc_j}^{NE} = \sum_i \lambda_{ic_i}^{NE} - \frac{\gamma_0}{\gamma_j^*}(\sum_i \lambda_{ic_i}^{NE})^2. \tag{2.30}$$

Thus, we observe that $\lambda_{jc_j}^{NE}$ is uniquely defined by $\sum_i \lambda_{ic_i}^{NE}$ for each j. Moreover, because of the theorem 2.6, $\sum_i \lambda_{ic_i}^{NE}$ is unique and so is $\lambda_{jc_j}^{NE}$.

Proof of proposition 2.8
From the first order optimality condition and from the fact that for all j, $\lambda_{jc_j}^{NE} > 0$, we deduce, for all j

$$\lambda_{jc_j}^{NE} = \sum_i \lambda_{ic_i}^{NE} - \frac{\gamma_0}{\gamma_j^*}(\sum_i \lambda_{ic_i}^{NE})^2. \tag{2.31}$$

By taking the sum over j in (2.31) we get:

$$\begin{aligned}
\sum_i \lambda_{ic_i}^{NE} &= J\sum_i \lambda_{ic_i}^{NE} - (\sum_i \frac{\gamma_0}{\gamma_i^*})(\sum_i \lambda_{ic_i}^{NE})^2 && (2.32)\\
1 &= J - (\sum_i \frac{\gamma_0}{\gamma_i^*})(\sum_i \lambda_{ic_i}^{NE}) && (2.33)\\
\sum_i \lambda_{ic_i}^{NE} &= (J-1)\frac{1}{(\sum_i \frac{\gamma_0}{\gamma_i^*})}. && (2.34)
\end{aligned}$$

The rest of the proof follows from equation (2.31).

Proof of theorem 2.10

First when $\varepsilon \to 0$, we can consider for all j, $\lambda_{jc_j}(t)$ is solution of

$$\dot{\lambda}_{jc_j}(t) = \left[\frac{1}{J}\sum_i \lambda_{ic_i}(t) - \lambda_{jc_j}(t)\right]_{\underline{\lambda}}^{\overline{\lambda}}. \tag{2.35}$$

We are in the case of cooperative ode [46] and then for all j the unique globally attractive rest point of this system is:

$$0 = [\frac{1}{J}\sum_i \lambda_{ic_i} - \lambda_{jc_j}]_{\underline{\lambda}}^{\overline{\lambda}} = \frac{1}{J}\sum_i \lambda_{ic_i} - \lambda_{jc_j}, \, \forall i, \, \lambda_{ic_i} \in [\underline{\lambda}, \overline{\lambda}]. \tag{2.36}$$

The previous equation implies for all j and l,

$$\begin{aligned} \dot{\lambda}_{jc_j}(t) &= \left[\frac{1}{\sum_i \lambda_{ic_i}(t)} - \frac{\sum_i \lambda_{ic_i}(t)}{J(\sum_i \lambda_{ic_i}(t))^2} - \frac{\gamma_0}{\gamma_j^*}\right]_{\underline{\lambda}}^{\overline{\lambda}} && (2.37) \\ &= \left[\frac{J-1}{J\sum_i \lambda_{ic_i}(t)} - \frac{\gamma_0}{\gamma_j^*}\right]_{\underline{\lambda}}^{\overline{\lambda}} && (2.38) \end{aligned}$$

It can be notice that (2.37) only depends of $\sum_i \lambda_{ic_i}(t)$. By taking the sum over i in (2.37) we obtain that $\sum_i \lambda_{ic_i}(t)$ is solution of

$$\sum_i \dot{\lambda}_{ic_i}(t) = \left[\frac{J-1}{\sum_i \lambda_{ic_i}(t)} - \sum_i \frac{\gamma_0}{\gamma_i^*}\right]_{\underline{\lambda}}^{\overline{\lambda}}. \tag{2.39}$$

(2.39) converges to the $\sum_i \lambda_{ic_i}^{NE}$. Indeed, (2.11) can be taken as a Lyapunov function for (2.39) and by using the Lasalle principle the convergence can be proved (p.118, [46]). From (2.36) we can conclude that:

$$\lim_{t\to\infty} \lambda_{jc_j}(t) = \lim_{t\to\infty} \frac{1}{J}\sum_i \lambda_{ic_i}(t). \tag{2.40}$$

Proof of proposition 2.11
The objective of a source j is given by;

$$\max_{\lambda_j \in [0,\overline{\lambda}]^L} \sum_{l=1}^{L} \frac{X_{jl}}{\sum_i X_{il}} - \gamma_0 \sum_{l=1}^{L} \lambda_{jl}, \tag{2.41}$$

where for all l, X_{jl} is solution of the following linear system:

$$X_{jl} = \lambda_{jl} + \sum_{l' \neq l} p_{l'l} X_{jl'}. \tag{2.42}$$

The following linear system as a unique solution due to the fact that the matrix P is substochastic. Therefore, we can deduce that this game is equivalent to the game where the objective of a source j is given by:

$$\max_{\lambda_j\in[0,\overline{\lambda}]^L,\, X_j} U(\lambda_j,\lambda_{-j},X_j,X_{-j}) := \sum_{l=1}^{L}\frac{X_{jl}}{\sum_i X_{il}} - \gamma_0\sum_{l=1}^{L}\lambda_{jl}, \tag{2.43}$$

subject to:

$$X_{jl} = \lambda_{jl} + \sum_{l'\neq l} p_{l'l}X_{jl'},\ \forall l, \tag{2.44}$$

where $X_j := (X_{j1},\ldots,X_{jL})$ and $X_{-j} := (X_1,\ldots,X_{j-1},X_{j+1},\ldots,X_J)$. We define the lagrangian function:

$$L_j(\lambda_j,X_j,\alpha_j) = \sum_{l=1}^{L}\frac{X_{jl}}{\sum_i X_{il}} - \gamma_0\sum_{l=1}^{L}\lambda_{jl} + \sum_{l=1}^{L}\alpha_{jl}(X_{jl}-\lambda_{jl}-\sum_{l'\neq l}p_{l'l}X_{jl}), \tag{2.45}$$

with $\alpha_j := (\alpha_{j1},\ldots,\alpha_{jL})$. By assuming that $\lambda_j^{NE} > 0$ and $X_j^{NE} > 0$ for all j, the first order optimality condition implies that:

$$\left\{\begin{array}{lcl} \dfrac{\partial L_j}{\partial \lambda_{jl}}(\lambda_j^{NE},X_j^{NE},\alpha_j^{NE}) & = & 0 \\ \dfrac{\partial L_j}{\partial X_{jl}}(\lambda_j^{NE},X_j^{NE},\alpha_j^{NE}) & = & 0 \end{array}\right. \tag{2.46}$$

$$\Leftrightarrow \left\{\begin{array}{lcl} -\gamma_0 & = & \alpha_{jl} \\ \dfrac{1}{\sum_i X_{il}^{NE}} - \dfrac{X_{jl}}{(\sum_i X_{il}^{NE})^2} + \alpha_{jl} - \sum_{l'\neq l}\alpha_{jl'}p_{ll'} & = & 0 \end{array}\right. \tag{2.47}$$

$$\Leftrightarrow \left\{\begin{array}{lcl} -\gamma_0 & = & \alpha_{jl} \\ \dfrac{1}{\sum_i X_{il}^{NE}} - \dfrac{X_{jl}}{(\sum_i X_{il}^{NE})^2} & = & \gamma_0(1-\sum_{l'\neq l}p_{ll'}) \end{array}\right. \tag{2.48}$$

$$\overset{\text{symmetric Nash}}{\Leftrightarrow} \left\{\begin{array}{lcl} -\gamma_0 & = & \alpha_{jl} \\ \dfrac{1}{JX_l^{NE}} - \dfrac{X_l^{NE}}{(JX_l^{NE})^2} & = & \gamma_0(1-\sum_{l'\neq l}p_{ll'}) \end{array}\right. \tag{2.49}$$

$$\Leftrightarrow \left\{\begin{array}{lcl} -\gamma_0 & = & \alpha_{jl} \\ X_l^{NE} & = & \dfrac{J-1}{J^2\gamma_0(1-\sum_{l'\neq l}p_{ll'})}. \end{array}\right. \tag{2.50}$$

According to (2.19), we can deduce that for each l and l':

$$\begin{aligned} \lambda_{jl}^{NE} &= X_l^{NE} - \sum_{l'} p_{l'l}X_{l'}^{NE} && (2.51) \\ &= \frac{J-1}{J^2}\left[\frac{1}{\gamma_0(1-\sum_{l'\neq l}p_{ll'})} - \sum_{l'}p_{l'l}\frac{1}{\gamma_0(1-\sum_{l''\neq l'}p_{l'l''})}\right]. && (2.52) \end{aligned}$$

Moreover,

$$
\begin{aligned}
U_j(\lambda_j^{NE}, \lambda_{-j}^{NE}) &= \sum_{l=1}^{L} \frac{X_{jl}^{NE}}{\sum_i X_{il}^{NE}} - \gamma_0 \sum_{l=1}^{L} \lambda_{jl}^{NE} && (2.53)\\
&= \sum_{l=1}^{L} \frac{X_l^{NE}}{JX_l^{NE}} - \gamma_0 \sum_{l=1}^{L} (X_l^{NE} - \sum_{l' \neq l} p_{l'l} X_{l'}^{NE}) && (2.54)\\
&= \frac{L}{J} - \gamma_0 \sum_{l=1}^{L} X_l^{NE} (1 - \sum_{l' \neq l} p_{ll'}) && (2.55)\\
&= \frac{L}{J} - \gamma_0 \sum_{l=1}^{L} \frac{J-1}{J^2 \gamma_0 (1 - \sum_{l' \neq l} p_{ll'})} (1 - \sum_{l' \neq l} p_{ll'}) && (2.56)\\
&= \frac{L}{J} - \frac{L(J-1)}{J^2} && (2.57)\\
&= \frac{L}{J^2}. && (2.58)
\end{aligned}
$$

Chapter 3

Posting Behavior Dynamics in OSN

Contents

This chapter is based on the paper *Posting behavior in Online Social Networks and Content Active Filtering*.

3.1 Introduction

Nowadays, OSNs (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 problematic: When do users post in OSNs? Why do they decide to post on a particular content? etc. Some works in Sociological Science try to find an answer to these questions [68] and [69].
The first goal of this chapter is to model the posting behaviour of users, also called sources. In our model, users with some particular content, connect to the OSN according to a Poisson Process. Following [68], we assume that there is some self-censorship exercised by the user: only contents that he estimates to be potentially popular[1] will be posted. To the best of our knowledge, mathematical models for posting behaviour do not involve popularity of contents [69], [70; 68]. 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 a limited budget of attention [15]. In this case, negative externalities occur, meaning that the larger the number of contents is, lower is the popularity of other content. Other typical scenarios are described in the introduction of [12]. In previous works, the posting behaviour of a source in an OSN was modelled using three main factors: trends, interest of the source and of its neighbours [70]. We introduce a model based on stochastic approximations [46] in order to study the evolution of the amount of posted contents by various sources taking into account the externalities impact. 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.

The second goal of this chapter 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 OSN 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 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 [71]. As objective of the CAF we shall consider maximizing the diversity of posted contents for the following reasons. We find it desirable that OSNs provide an access to content of different sources in a diverse way. We call this property *content diversity*. Such diversity is critical in Political News [72] and [73]. Websites Initiatives (like `politifact.com` or `FactCheck.org`)

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

that try to introduce the same visibility to different opinions. Another context where content diversity is an important issue is Health Information [74; 75]. 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 contents. A similar issue arises when we are interested in decreasing misinformation propagation and gossip in OSNs, [76], [6] and [36].

Organization of the chapter. After a short state-of-the-art section, we motivate the externalities effect by studying data extracted from an OSN in section 3.3. Then, in section 3.4, we introduce the posting behaviour model which includes the externalities. We formulate the limit regime for appropriate scaling obtained through stochastic approximations. In section 3.4.2, 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 3.5, we define content diversity and we propose a Content Active Filtering (CAF) with a closed-form. Numerical simulations of posting behaviour convergence are proposed in section 3.6. 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 influence matrix considering a reverse engineering viewpoint. Finally, in section 3.7 we give some conclusions and perspectives of the chapter.

3.2 Related works

Many recent works study empirically the behaviour of users in OSNs, for example [29] and [33]. In [29], empirical analysises are proposed to study the connectivity properties and user activity on different OSNs, in particular on Facebook, Twitter and Google+. The authors notice that on Facebook and Google+, there are more creation of messages than reshares (to relay a received post), and that the opposite scenario occurs on Twitter. In [33], 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 an OSN dataset, that their model is more realistic than other models from the literature.

The question of visibility and popularity of contents in OSNs have recently been subjects of interest from a theoretical point of view. For example, in [53] the authors consider a dynamical model to describe the popularity of various contents in Youtube. Each source 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 sources. In [77], the authors propose a model in which several sources 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 its popularity increases. The goal of this work is to understand how to model the dynamics of message visibility and to understand the consequences of a competition over visibility between different sources.

Negative externalities are well-known phenomena in diffusion process [78] and [79]. The model proposed in [78] describes a situation in which several users of an OSN 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 [79], the authors discuss 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 content diversity improvement in OSNs. In social sciences, there exist some studies that try to design tools in order to increase diversity in political opinions [80] or to improve civil discourses [81]. In [80] the authors use a widget[2] that shows to users their reading diversity. A search engine algorithm is proposed in [80] to increase content diversity.

3.3 Popularity Analysis based on a Case Study

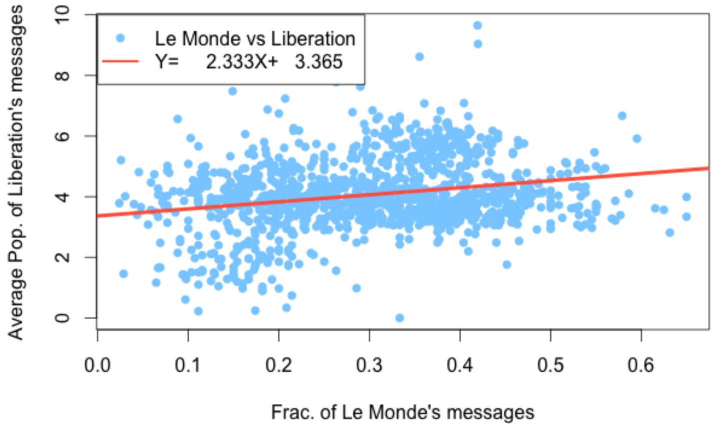
In this section, we present an empirical study of the posting behaviour of 3 French news companies over an OSN. Particularly, we study the correlation between the proportion of messages from different source (news company) and the average popularity of the messages. Therefore we first demonstrate the relationship between the popularity of messages from one source with respect to the posting rates of the other sources. Second, we observe that this relationship implies negative or positive externalities.

Our dataset is composed of messages posted on three Facebook pages (Le monde, Liberation, Le Figaro). We extract in total 52376 messages posted using the Netvizz app [19]. The data was collected from January 01 2010 to June 06 2015. Each message is associated with the source , the timestamp and the number of "like".

The popularity of each message is measured as a function of the number of "like" associated with it. On figure 3.1, each point represents the average popularity of all messages published ruling a specific day from one news company, with respect to the posting rate of another news company.

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. 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

[2] http://balancestudy.org/



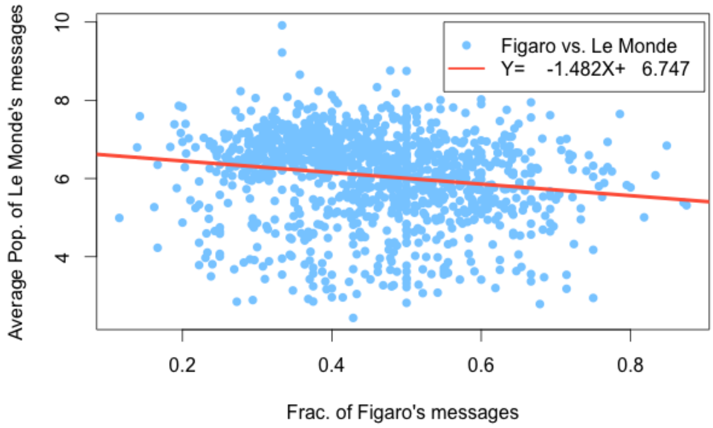


Figure 3.1: *Popularity impact over the posting strategy*

Le Figaro and the average popularity of Le Monde's messages. The correlation matrix between the 3 news companies, also called sources, is provided below:

$$\begin{bmatrix} 0.2564758 & -0.08708539 & -0.167834 \\ 0.2359287 & -0.2455585 & -0.01392799 \\ 0.3441988 & -0.09750635 & -0.2416787 \end{bmatrix}. \tag{3.1}$$

We then demonstrate that both positive and negative correlations are possible between the popularity of messages and the posting rates of sources. In the next section, we develop the evolution of the posting behaviour of a source, taking into account the correlation between the popularity of messages and the posting rates. This externality effect is captured through an influence matrix.

Table 3.1: Main notations used throughout the chapter 3

Symbol	*Meaning*
C	number of sources
t_n	arrival instant of the n^{th} content
$Z_c(n)$	potential level of popularity
λ_c	arrival rate of contents from source c
$\xi(n)$	source of content n
$y_c(n)$	total number of posts from c; $x_c(n) := \frac{y_c(n)}{n}$ $\mathbf{x}(n) = (x_1(n), \ldots, x_C(n))$, $\mathbf{x}(0) = \mathbf{x}_0$
A^+ (A^-)	the positive (resp. the negative) impacts or externalities matrix between sources
θ_c	popularity threshold associated with source c
p_c	probability that the content from source c is accepted in the OSN $\mathbf{p} := (p_1, \ldots, p_C)$

3.4 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 with each source when they are sensitive to popularity. We try to answer 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. The main symbols used in the chapter are reported in Tab. 3.1.

3.4.1 Model

We consider that sources generate new content and then sources 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 n^{th} 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 sources, which could be interested in its and therefore may want to publish its. A source may also be any user of an OSN. Each content has a potential level of popularity $Z_c(n) \in \mathbb{R}_+$. For any source c, the arrival rate of contents is an independent Poisson point process with intensity $\lambda_c \in \mathbb{R}_+$. Let $\xi(n) \in \mathscr{C}$ be the random variable that determines the source associated with the n^{th} content. The probability that the n^{th} opportunity concerns the source c is given by [82]:

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

For any source c, let $y_c(n)$ be the total number of posts of source c in the OSN during $[0,t_n]$. For each source c, let $x_c(n) := \frac{y_c(n)}{n}$ be the average number of posts from c. Note that, by definition, we have for all source 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 source 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 source has enough experience to predict the popularity of his content. The random variables $Z_c(n)$ are independent. The probability that the source c will posts his content is given by, for each n:

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

where the matrix $A^+ := \{a^+_{cc'}\}_{cc'}$ and $A^- := \{a^-_{cc'}\}_{cc'}$ describe, respectively, the positive (resp. the negative) impacts or externalities between sources. These two matrices are assumed to be substochastic. The difference $a^+_{cc} - a^-_{cc}$ gives the global influence of c' over c. For example if $a^+_{cc} - a^-_{cc}$ is positive then source c' has a positive influence over source c.

Note that for any source 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 [12]. Then the probability that the n^{th} content is posted by c is given by:

$$\begin{aligned} P(Z_c(n) \geq \theta_c, \xi(n) = c \mid \mathbf{x}(n)) &= P(\xi(n) = c) \times P(Z_c(n) \geq \theta_c \mid \xi(n) = c, \mathbf{x}(n)) \\ &= \frac{\lambda_c}{\sum_{c'} \lambda_{c'}} \left[\frac{1}{2}\sum_{c'} a^-_{cc'}(1 - x_{c'}(n)) + \frac{1}{2}\sum_{c'} a^+_{cc'} x_{c'}(n) \right]. \end{aligned}$$

The linearity property is similar to the one proposed in [12]. The authors consider a linear relation between the probability for user to get some content knowing 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 [12].

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 [83] or delay function in routing games [49]. In fact, in an economic framework, the inverse linear demand is a classical assumption [83] and [84]. It is also the case in network economic models [85].

Example from case study. In fig. 3.2 is depicted information flows in a News Business taking from the case study described in section 3.3. Different creators of contents generate contents, some of them are exclusive to sources and some of them

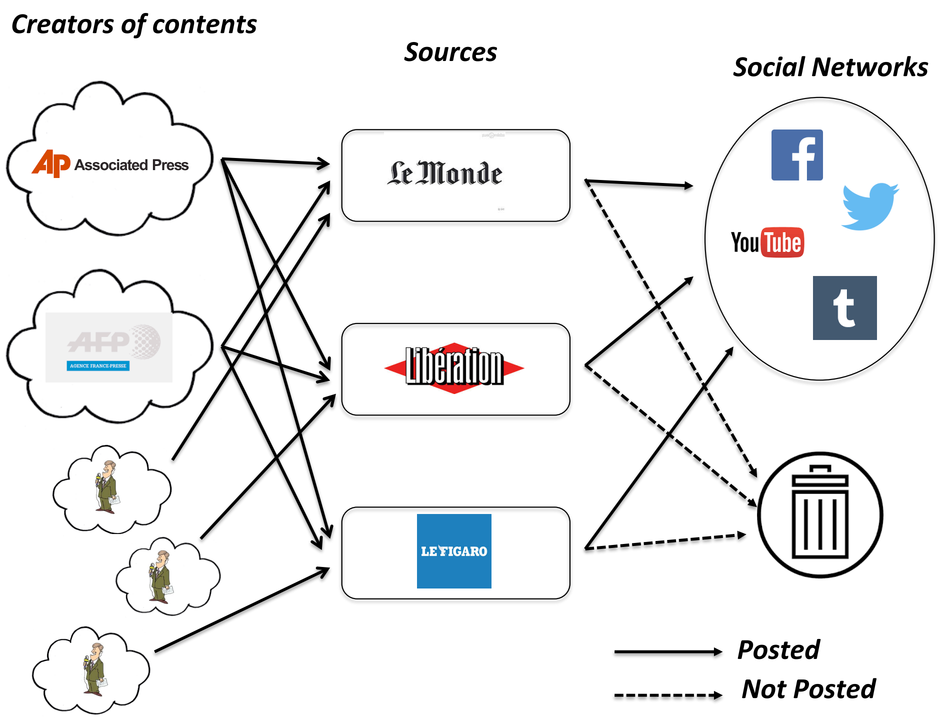


Figure 3.2: *Information flows in News business*

are common creators of contents like AP and AFP. Sources, which are famous News French companies (Le Monde, Liberation and Le Figaro) decide to post or not their incoming content on OSNs like Facebook, Tweeter, etc.

3.4.2 Rest point analysis

We first study the asymptotic behaviour 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 source 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:

$$\zeta_c(n) := \begin{cases} 1 & \text{w.p} \quad P(Z_c(n) \geq \theta_c, \xi(n) = c \mid \mathbf{x}(n)), \\ 0 & \text{w.p} \quad 1 - P(Z_c(n) \geq \theta_c, \xi(n) = c \mid \mathbf{x}(n)). \end{cases} \tag{3.2}$$

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

$$x_c(n+1) = x_c(n) + \frac{1}{n+1}\left(\zeta_c(n) - x_c(n)\right). \tag{3.3}$$

The posting rates vector $\mathbf{x}^*$ is called the rest point associated with (3.3). Intuitively, equation(3.3) can be seen as a finite difference Euler scheme of the following system of differential equations:

$$\dot{x}_c(t) = P(Z_c(t) \geq \theta_c, \xi(t) = c \mid \mathbf{x}(t)) - x_c(t),\ x_c(0) = x_c^0. \tag{3.4}$$

However, equation (3.3) is a stochastic difference equation, which is not the case for the classical Euler scheme. The theory of stochastic approximations [46] 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:

$$\forall c, c', \quad B_{cc'} = \frac{\lambda_c}{2\sum_{c''}\lambda_{c''}}(a^+_{cc'} - a^-_{cc'}) - 1_{c=c'}, \tag{3.5}$$

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

$$\mathbf{D} = \left[\frac{\lambda_1}{2\sum_{c'}\lambda_{c'}}\sum_{c'} a^-_{1c'}, \ldots, \frac{\lambda_C}{2\sum_{c'}\lambda_{c'}}\sum_{c'} a^-_{Cc'}\right]. \tag{3.6}$$

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

Assumption 1: The matrix $\dfrac{A^+ - A^-}{4} + \dfrac{(A^+ - A^-)^T}{4}$ is assumed to be negative definite. Under this assumption, we show in the next proposition the convergence of the dynamical process $\mathbf{x}(n)$ to the unique rest point $\mathbf{x}^*$.

Proposition 3.1. *We assume that assumption 1 holds. 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:

$$B\mathbf{x}^* + \mathbf{D} = 0. \tag{3.7}$$

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.

3.4.3 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 sources. Particularly, we show that the rest point $\mathbf{x}^*$ obtained in proposition 3.1 is a Nash equilibrium of a non-cooperative game in which each source 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 source has an interest to deviate from the equilibrium. Then, in other words, the rest point of the dynamical system described in 3.3 is strategically stable.

Each source c controls his posting rate x_c, which denotes the amount of contents posted by the source c per unit of time (we call it also the posting rate) in the OSN. We assume that the average popularity of a content posted by source c is a function of the posting rates vector $\mathbf{x}$ of all sources. sources 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 trough economic trades. The utility function of source c is given by:

$$U_c(x_c,\mathbf{x}_{-c}) = x_c\frac{\lambda_c}{\sum_{c'}\lambda_{c'}}\Big(\frac{1}{2}\sum_{c'\neq c} a^+_{cc'}x_{c'} + \frac{a^+_{cc}}{4}x_c + \frac{1}{2}\sum_{c'\neq c} a^-_{cc'}(1-x_{c'}) + \frac{a^-_{cc}}{4}(1-x_c)\Big) - \frac{1}{2}(x_c)^2,$$

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 3.2. *A vector $\mathbf{x}^{NE} \in \mathbb{R}^+$ is a Nash Equilibrium if for all source c,*

$$x_c^{NE} \in \underset{x_c\geq 0}{arg\ max}\, U_c(x_c,\mathbf{x}_{-c}^{NE}). \tag{3.8}$$

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 [49].

Asumption 2: It exists a vector of posting rates $\mathbf{x}^{NE}$ that satisfies equation (3.8) and such that for each source c, $x_c^{NE} > 0$.

Definition 3.3. *For each c, the noisy best response of source c against $\mathbf{x}_{-c}$ is given by:*

$$x_c^{NE} \in \underset{x_c\geq 0}{arg\ max}\, U_c(x_c,\mathbf{x}_{-c}^{NE}+\varepsilon_{-c}), \tag{3.9}$$

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:

$$
\begin{aligned}
0 &= \frac{\partial U_c}{\partial x_c}(x_c, x_{-c} + \varepsilon_{-c}) \\
&= \frac{\lambda_c}{\sum_{c'} \lambda_{c'}} \left(\sum_{c'} a^+_{cc'} x_{c'} + \sum_{c'} a^-_{cc'} (1 - x_{c'}) \right) - x_c + \frac{\lambda_c}{\sum_{c'} \lambda_{c'}} \sum_{c'} (a^+_{cc'} - a^-_{cc'}) \varepsilon_{c'} \\
&\Leftrightarrow \left(\frac{\sum_{c'} \lambda_{c'} + \lambda_c (a^-_{cc} - a^+_{cc})}{\sum_{c'} \lambda_{c'}} \right) x_c = \frac{\lambda_c}{\sum_{c'} \lambda_{c'}} \left(\sum_{c' \neq c} a^+_{cc'} x_{c'} + \sum_{c' \neq c} a^-_{cc'} (1 - x_{c'}) + a^-_{cc} \right) \\
&+ \frac{\lambda_c}{\sum_{c'} \lambda_{c'}} \sum_{c'} (a^+_{cc'} - a^-_{cc'}) \varepsilon_{c'} \\
&\Leftrightarrow x_c = \frac{\lambda_c}{(\sum_{c'} \lambda_{c'} + (a^+_{cc} - a^-_{cc}) \lambda_c)} \left(\sum_{c' \neq c} a^+_{cc'} x_{c'} + \sum_{c' \neq c} a^-_{cc'} (1 - x_{c'}) + a^-_{cc} \right) \\
&+ \frac{\lambda_c}{(\sum_{c'} \lambda_{c'} + (a^+_{cc} - a^-_{cc}) \lambda_c)} \sum_{c'} (a^+_{cc'} - a^-_{cc'}) \varepsilon_{c'}.
\end{aligned}
\tag{3.10}
$$

The previous relationship between x_c and $\mathbf{x_{-c}}$ provides an algorithmic way to compute the Nash equilibrium. Indeed, a decentralized algorithm that computes iteratively the best response of each source 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 sources at time slot t:

Best-response algorithm [86]
Initialization: $\mathbf{x}(1)$ is randomly chosen in $[0,1]^C$.
For each round $t = 1, 2, \ldots$

1. Each source 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 source updates his decision by computing his best response:

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

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

We illustrate the convergence of the best response algorithm in the numerical section 3.6. Finally, in the next proposition, we prove the equivalence between $\mathbf{x}^*$, the rest point of (3.4), and a vector $\mathbf{x}^{NE}$ which is an interior Nash equilibrium of the non-cooperative popularity game between sources.

Proposition 3.4. *We assume that Assumption 1 and Assumption 2 hold.* $\mathbf{x}^* \in [0,1]^C$ *is a rest point of* (3.4) *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, equivalently at the Nash Equilibrium point.

3.5 Content diversity

The second goal of this chapter is to increase the diversity of contents in an OSN. We are looking for contents that are posted by different sources. Note that even if the same content, provided by a common source, is posted by two different sources, their posted messages will not be necessary exactly identical and then perceived differently by the OSN consumers. In this section, we first define the diversity of contents in an OSN in terms of a relationship between the average numbers of posts of all the sources. 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 ensure this diversity property.

3.5.1 Diversity analysis

The diversity of contents published in an OSN is highly related to equal proportion of contents such that subscribers have access to all contents from different sources in an equal way. We then define the diversity property of source rate vector $\mathbf{x} \in \mathbb{R}^C$.

Definition 3.5. *Let* $\mathbf{x} \in \mathbb{R}^C$ *be a vector of the average number of posts associated with each source in an OSN. We say that* $\mathbf{x}$ *satisfies the diversity property if for all* $(c,c') \in \mathscr{C}$, *we have:*

$$x_c = x_{c'}. \tag{3.11}$$

Note that our definition of content diversity is related to the posting rates of the sources and not to the content itself. In fact, messages published by different sources may mention the same content, but these contents are coming from different sources or are exposed differently by the sources. We therefore talk about content diversity in order to specify that the messages posted on the OSN come from different sources. The next proposition provides a necessary and sufficient condition for the rest point $\mathbf{x}^*$ to satisfy the *diversity property* given by equation (3.11).

Proposition 3.6. *The rest point $\mathbf{x}^* \in [0,1]^C$ satisfies the diversity property if and only if: $\forall (c,c'') \in \mathscr{C}^2$*

$$\frac{\lambda_c \sum_{c'} a^-_{cc'}}{\lambda_c \sum_{c'} (a^-_{cc'} - a^+_{cc'}) + 2\sum_{c'} \lambda_{c'}} = \frac{\lambda_{c''} \sum_{c'} a^-_{c''c'}}{\lambda_{c''} \sum_{c'} (a^-_{c''c'} - a^+_{c''c'}) + 2\sum_{c'} \lambda_{c'}}.$$

Several measures of diversity are already proposed in different scientific fields. But the most well-known are the Gini index, defined in [87], and the Shannon entropy, defined in [88]. Usually, a diversity measure is defined as a function $D : \mathbb{R}^C \to \mathbb{R}$ satisfying the following property:

$$\mathbf{x}^E = \operatorname*{argmax}_x D(\mathbf{x}) \quad \text{iff} \quad x^E_c = x^E_{c'}, \ \forall (c,c') \in \mathscr{C}. \tag{3.12}$$

In this chapter, we look for a diversity property and not a measure of the level of diversity in an OSN. 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.

3.5.2 Increase diversity: a content active filtering approach

What kind of control can be used to increase diversity in an OSN? We answer this question by proposing a content active filtering (CAF) approach. For each news arrival n, if the message is posted, the OSN may accept or not the message. Therefore the OSN may refuse to relay some messages. We consider a source 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 source c is accepted, and therefore posted on the OSN. We assume that each source knows the OSN CAF policy $\mathbf{p}$. Considering a given CAF $\mathbf{p}$, for the source c the probability that a new arrival message is posted is then given by:

$$P(Z_c(n) \geq \theta_c \mid \xi(n) = c, p_c, \mathbf{x}(n)) = p_c\Big(\frac{1}{2}\sum_{c'} a^+_{cc'} x_{c'}(n) + \frac{1}{2}\sum_{c'} a^-_{cc'}(1 - x_{c'}(n))\Big).$$

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

$$x_c(n+1) = x_c(n) + \frac{1}{n+1}\left(\zeta^1_c(n) - x_c(n)\right), \tag{3.13}$$

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

$$\zeta_c^1(n) := \begin{cases} 1 & \text{w.p } \frac{\lambda_c}{\sum_{c'} \lambda_{c'}} P(Z_c(n) \geq \theta_c \mid \xi(n) = c, p_c, \mathbf{x}(n)), \\ 0 & \text{w.p } 1 - \frac{\lambda_c}{\sum_{c'} \lambda_{c'}} P(Z_c(n) \geq \theta_c \mid \xi(n) = c, p_c, \mathbf{x}(n)). \end{cases} \tag{3.14}$$

The rest point $\mathbf{x}^*$ of (3.13), according to theorem 3.1, is solution of the following system:

$$x_c^* = p_c \frac{\lambda_c}{\sum_{c'} \lambda_{c'}} \left[\frac{1}{2} \sum_{c'} a_{cc'}^+ x_{c'}^* + \frac{1}{2} \sum_{c'} a_{cc'}^- x_{c'}^* \right], \forall c \in \mathscr{C}. \tag{3.15}$$

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 OSN 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 (3.13) then $\mathbf{x} \in [0,1]^C$ (rest point of the dynamics) satisfies (3.11) (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 (3.13) converges to $\mathbf{x}^*$.

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

Proposition 3.7. *We assume that assumption 1 hold and we define:*

$$Y = \min_c \left(\frac{\lambda_c \sum_{c'} a_{cc'}^-}{\lambda_c \sum_{c'} (a_{cc'}^- - a_{cc'}^+) + 2 \sum_{c'} \lambda_{c'}} \right), \tag{3.16}$$

$$c^* = \arg\min_c \left(\frac{\lambda_c \sum_{c'} a_{cc'}^-}{\lambda_c \sum_{c'} (a_{cc'}^- - a_{cc'}^+) + 2 \sum_{c'} \lambda_{c'}} \right). \tag{3.17}$$

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

$$p_{c^*} = 1 \tag{3.18}$$

$$p_c = \frac{2Y \sum_{c'} \lambda_{c'}}{\lambda_c (\sum_{c'} a_{cc'}^- - Y(\sum_{c'} a_{cc'}^- - a_{cc'}^+))}, \quad \forall c \neq c^*, \tag{3.19}$$

satifies the following properties:

1. *the unique rest point $\mathbf{x}^* \in [0,1]^C$ of (3.13) satisfies the diversity property (3.11),*
2. *and the dynamical system given by equation (3.13) converges to $\mathbf{x}^*$.*

A CAF is a simple control mechanism that can be used by an OSN to control the contents posted by the sources. CAFs already exist in various forms in OSN, 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 user's news feed when a friend of him posts messages. A CAF is also used to filter notifications that are sent to members of groups on the group activity of other members. CAFs can be criticized for being non-democratic in the sense that an OSN 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 OSNs, and can be designed in an efficient way given our analysis.

3.6 Numerical illustrations and reverse engineering

3.6.1 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 3.7.

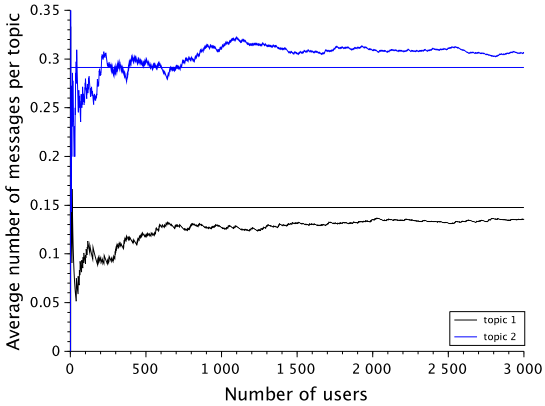
The contents posted in the OSN are restricted to $\mathscr{C} = 3$ sources, like the News business example with French News companies described in section 3.3. Estimating real influence matrices is out of the scope of the chapter. By the way, influence matrices estimators can be used as input in our framework. We consider for the sake of example, the following influence 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 sources have stronger influence with themselves than on another source. The intensity of the Poisson point process associated with contents arrival is assumed to be the same for each source c and equal to $\lambda_c = 10$. In fig. 3.3, we observe the

[3] https://www.facebook.com/notes/facebook/facebook-tips-whats-the-difference-betw
414305122130



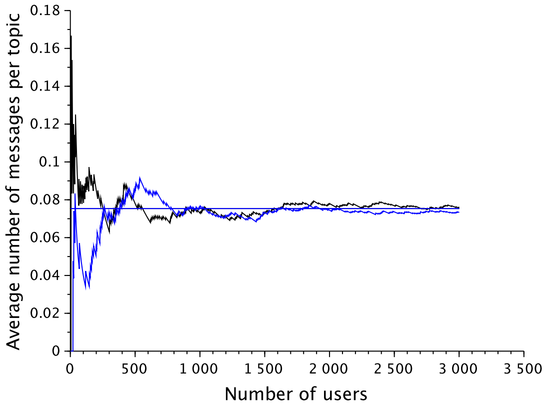


Figure 3.3: *Simulation of the arrival of users and messages associated with 3 contents without and with the CAF.*

convergence of the dynamics of the average number of posts for the sources 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. $x_1^* = x_2^* = x_3^*$.

3.6.2 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 influence matrices. Practically, we assume that the sources are competing for maximizing the popularity of their content as proposed in section 3.4.3. We assume that each day, denoted by t, sources are competing and they update their decision with the help of the best-response algorithm. Let $x_c^{NE}(t)$ be the posting rate of source c during day t. Thus according to (3.10), for each c,

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

We 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: $A^1_{cc'} := \frac{a^+_{cc'} - a^-_{cc'}}{a^+_{cc} - a^-_{cc}}$ (the associated matrix is A^1) and $D^1_c := \frac{\sum_{c'} a^-_{cc'}}{a^+_{cc} - a^-_{cc}}$ (where the associated vector is D^1). In the least square method, we solve the following optimization problem:

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

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 3.3, we obtain the following matrices:

$$A^1 = \begin{pmatrix} 0 & 0.49938 & 0.20559 \\ 0.30016 & 0 & -0.01863 \\ 0.34687 & 0.09159 & 0 \end{pmatrix},$$

and

$$D^1 = \begin{pmatrix} 4.69251 \\ 6.75663 \\ 13.66454 \end{pmatrix}.$$

Note that this method aims to estimate the total influence of each source on the others. In fact, we can assume that each source c knows his own influence, i.e. a^+_{cc} and a^-_{cc}. Given this local knowledge and estimating the coefficient $A^1_{cc'}$, each source c has an estimator of the total influence of each source 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^1, which can be confirmed by the fact that the correlation matrix (3.1) has also positive and negative coefficients. Moreover, it can be noticed that the source which has the most influence on the others is Le Monde, which makes sense because it is one of the most famous newspaper compared to the others[4].

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

3.6.3 Best response algorithm

We describe in this section the convergence of the best-response algorithm described in section 3.4.3. Sources 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 sources 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. 3.4 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 sources 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.

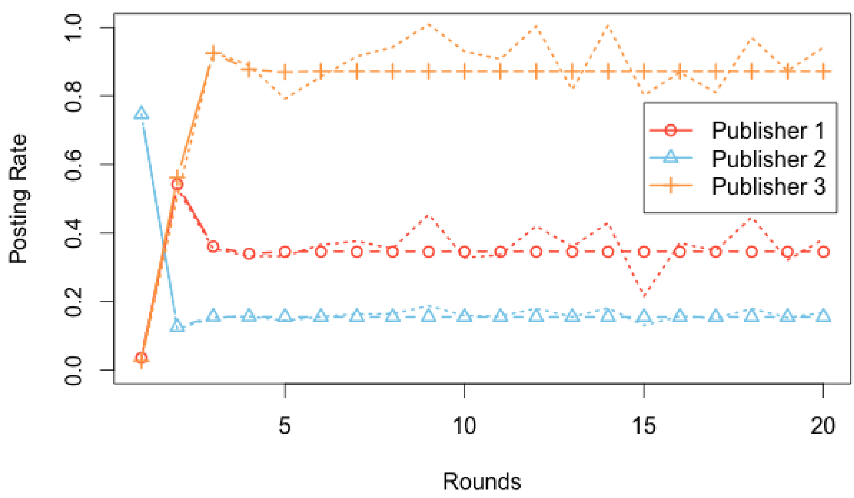


Figure 3.4: *Dynamic of the best response algorithm.*

3.7 Conclusion

In this chapter we first model the posting behaviour of sources in OSNs in several contents which have externalities impact one over the other. Second we propose a content active filtering in order to increase content diversity. We use a dynamical approach (based on stochastic approximation theory) to model the posting behaviour of sources taking into account these externalities. The convergence of the posting behaviour is proved and a game theoretical equivalent is proposed. Then we define a content active filtering control, with an explicit form, in order to improve content diversity in the OSN. Finally, all the theoretical results are illustrated through simulations and a data set extracted from a real OSN is used to assess our results.

APPENDIX

Proof of Proposition 3.1
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 (3.4). 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 [89] we prove that $\lim_{t\to\infty} \mathbf{x}(t) = \mathbf{x}^*$. Finally, by the using the theory of stochastic approximation [46] 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:

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

It can be noticed that, for time n and source c:

$$E\left[M_c(n+1) \mid \zeta_c(m),\, m \leq n\right] :=$$
$$E\left[\zeta_c(n+1) - P(Z_c(n+1) \geq \theta_c, \xi(n+1) = c \mid \mathbf{x}(n)) \mid \zeta_c(m),\, m \leq n\right]$$
$$= E[\zeta_c(n+1) \mid \zeta_c(m),\, m \leq n] - P(Z_c(n+1) \geq \theta_c, \xi(n+1) = c \mid \mathbf{x}(n))$$
$$= P(Z_c(n+1) \geq \theta_c, \xi(n+1) = c \mid \mathbf{x}(n)) \times 1$$
$$+ P(Z_c(n+1) < \theta_c, \xi(n+1) = c \mid \mathbf{x}(n)) \times 0$$
$$- P(Z_c(n+1) \geq \theta_c, \xi(n+1) = c \mid \mathbf{x}(n)) = 0.$$

Thus $M_c(n)$ is a martingale difference sequence of zero mean (see appendix of [46] 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:

$$\dot{x}_c(t) \quad = \quad P(Z_c(t) \geq \theta_c, \xi(t) = c \mid \mathbf{x}(t)) - x_c(t),\ \forall c. \tag{3.20}$$

Step 2: It can be easily deduced that the rest point $\mathbf{x}^*$ of (3.4) is solution of:

$$B\mathbf{x}^* + \mathbf{D} = 0. \tag{3.21}$$

It can be noticed that $B \in [0,1]^{C^2}$ is a strictly diagonally dominant matrix [90]. Indeed, for all c,

$$| b_{cc} | - \sum_{c' \neq c} | b_{cc'} | = 1 + \frac{\lambda_c}{2\sum_{c'} \lambda_{c'}}(a^+_{cc} - a^-_{cc}) - \frac{\lambda_c}{\sum_{c'} \lambda_{c'}} \sum_{c' \neq c} (a^+_{cc'} - a^-_{cc'}) \geq 0$$

Thus because B is strictly diagonally dominant, B is invertible and this implies that (3.4) 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 (3.20). We study the function:

$$V(\mathbf{x}(t)) := \sum_c (x_c(t) - x^*_c)^2, \tag{3.22}$$

where $\mathbf{x}^*$ is solution of (3.21). Then, for all $\mathbf{x}(t)$ we get

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

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 LaSalle's Invariance Principle [46], we can deduce the following:

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

Step 4: Finally, we apply theorem 2 p.15 of [46] to show that $\mathbf{x}(n)$ (whose components are given by (3.3)) converges almost surely to the compact invariant set of the o.d.e (3.4) given by the singleton $\mathbf{x}^*$.

Proof of Proposition 3.4
The proof is made in two steps. In the first step, we provide a characterization of the Nash Equilibrium. In the second step we prove the equivalence. According to definition 3.8, a vector $x^{NE} \in \mathbb{R}^C$ is a Nash Equilibrium if for each source $c \in \{1, \ldots, C\}$,

$$x^{NE}_c \in \arg\max_{x_c \geq 0} \left(x_c \frac{\lambda_c}{2\sum_{c'} \lambda_{c'}} (Y_c + \frac{a^+_{cc}}{2} x_c + \frac{a^-_{cc}}{2}(1 - x_c)) - \frac{1}{2}(x_c)^2 \right). \tag{3.23}$$

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

$$\frac{\partial U_c}{\partial x_c}(x^{NE}_c, \mathbf{x}^{NE}_{-c}) \;=\; 0, \text{ with } x^{NE}_c > 0, \tag{3.24}$$

$$\frac{\partial U_c}{\partial x_c}(x^{NE}_c, \mathbf{x}^{NE}_{-c}) \;\leq\; 0, \text{ with } x^{NE}_c \geq 0, \tag{3.25}$$

which are equivalent to the following:

$$\frac{\lambda_c}{2\sum_{c'} \lambda_{c'}}\left(\sum_{c'} a^-_{c,c'}(1 - x^{NE}_{c'}) + \sum_{c'} a^+_{c,c'} x^{NE}_{c'}\right) \;=\; x^{NE}_c, \text{ if } x^{NE}_c > 0,$$

$$\frac{\lambda_c}{2\sum_{c'} \lambda_{c'}}\left(\sum_{c'} a^-_{c,c'}(1 - x^{NE}_{c'}) + \sum_{c'} a^+_{c,c'} x^{NE}_{c'}\right) \;\leq\; x^{NE}_c, \text{ if } x^{NE}_c \geq 0.$$

Assuming the existence of an interior Nash equilibrium (Assumption 2), we then obtain that the first order conditions given by equation (3.26) are equivalent to the rest point solution obtained in Proposition 3.1. 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 (3.21).

Proof of Proposition 3.6

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

- We prove the first necessary condition, i.e. if the rest point $\mathbf{x}^* \in [0,1]^C$ satisfies (3.11) then $\forall (c, c'') \in \mathscr{C}^2$, we have:

  $$\frac{\lambda_c \sum_{c'} a^-_{cc'}}{\lambda_c \sum_{c'} (a^-_{cc'} - a^+_{cc'}) + 2\sum_{c'} \lambda_{c'}} = \frac{\lambda_{c''} \sum_{c'} a^-_{c''c'}}{\lambda_{c''} \sum_{c'} (a^-_{c''c'} - a^+_{c''c'}) + 2\sum_{c'} \lambda_{c'}}.$$

  According to Definition 3.5 and Theorem 3.1, if $\mathbf{x}^* := [x, \ldots, x]$ satisfies the diversity property then, for all content $c \in \mathscr{C}$, it is the unique solution of

  $$\frac{\lambda_c}{2\sum_{c'} \lambda_{c'}}\left(\sum_{c'} a^-_{c,c'}(1 - x^*) + \sum_{c'} a^+_{c,c'} x^*\right) = x^*, \tag{3.26}$$

  $$\Leftrightarrow x^* = \frac{\lambda_c \sum_{c'} a^-_{cc'}}{\lambda_c \sum_{c'} (a^-_{cc'} - a^+_{cc'}) + 2\sum_{c'} \lambda_{c'}}. \tag{3.27}$$

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

  $$\frac{\lambda_c \sum_{c'} a^-_{cc'}}{\lambda_c \sum_{c'} (a^-_{cc'} - a^+_{cc'}) + 2\sum_{c'} \lambda_{c'}} = \frac{\lambda_{c''} \sum_{c'} a^-_{c''c'}}{\lambda_{c''} \sum_{c'} (a^-_{c''c'} - a^+_{c''c'}) + 2\sum_{c'} \lambda_{c'}}.$$

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

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

then the rest point $\mathbf{x}^* \in [0,1]^C$ satisfies the diversity condition given by equation (3.11). We notice that for each content $c \in \mathscr{C}$, we have that

$$y := \frac{\lambda_c \sum_{c'} a^-_{cc'}}{\lambda_c \sum_{c'} (a^-_{cc'} - a^+_{cc'}) + 2\sum_{c'} \lambda_{c'}}$$

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

$$\frac{\lambda_c}{2\sum_{c'} \lambda_{c'}} (\sum_{c'} a^+_{cc'} y + \sum_{c'} a^-_{cc'} (1-y)) = y. \tag{3.28}$$

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

Proof of Proposition 3.7

According to proposition 3.6, having the equal proportion is equivalent to find $\mathbf{p} \in [0,1]^C$, $\forall c \in \mathscr{C} - \{C\}$, such that:

$$\frac{p_c \lambda_c \sum_{c'} a^-_{cc'}}{p_c \lambda_c \sum_{c'} (a^-_{cc'} - a^+_{cc'}) + 2\sum_{c'} \lambda_{c'}} = \frac{p_{c''} \lambda_{c''} \sum_{c'} a^-_{c''c'}}{p_{c''} \lambda_{c''} \sum_{c'} (a^-_{c''c'} - a^+_{c''c'}) + 2\sum_{c'} \lambda_{c'}}.$$

We can notice that we are in presence of a 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 (3.11) 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

$$Y := \frac{p_{c*} \lambda_{c*} \sum_{c'} a^-_{c*c'}}{p_{c*} \lambda_{c*} \sum_{c'} (a^-_{c*c'} - a^+_{c*c'}) + 2\sum_{c'} \lambda_{c'}}$$

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

$$Y = \frac{p_c \lambda_c \sum_{c'} a^-_{cc'}}{p_c \lambda_c \sum_{c'} (a^-_{cc'} - a^+_{cc'}) + 2\sum_{c'} \lambda_{c'}} \tag{3.29}$$

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

$$c^* = \text{ argmin }_c \{ \frac{\lambda_c \sum_{c'} a^-_{cc'}}{\lambda_c \sum_{c'} (a^-_{cc'} - a^+_{cc'}) + 2 \sum_{c'} \lambda_{c'}} \},$$

and $p_{c^*} = 1$. Finally, we just need to need to solve (3.29), and we get the solution. Moreover, the control $\mathbf{p}$ does not impact the diagonal dominance of matrix B. This is the reason why (3.13) still converges to $\mathbf{x}^*$.

Chapter 4

Explicit popularity Competition on Online OSNs

Contents

This chapter is based on the paper *Differential games of competition in online content diffusion*.

4.1 Introduction

Online content delivery represents an ever increasing fraction of Internet traffic. In the case of online videos, the support for content distribution is provided by commercial platforms such as Vimeo or YouTube. In many cases, such contents are also delivered by means of OSN platforms. One core feature of such systems is the delivery of user-generated content (UGC): platform users become often producers of the contents which populate those systems.

Reference figures for UGC platforms are indeed those of YouTube, with over 6 billion hours of videos watched each month, which averages as an hour for every person on Earth a month. New UGCs are continuously created and 100 hours of video uploaded every minute by the platform's users[1].

A relevant parameter for the UGC platforms owners is the *viewcount*, i.e., the number of times an item has been accessed. Viewcount in fact represents one of the possible metrics to measure content popularity. In turn, a popular video becomes a source of revenue because of click-through rates of linked advertisements: those are actually part of the YouTube's business model.

Among several research works in the field, many efforts have been spent to characterize the dynamics of popularity of online media contents [91; 92; 93; 94; 95; 96]. The ultimate target there would be indeed to provide models able to perform the early-stage prediction of a content's popularity [18].

Such studies have highlighted certain phenomena that are typical of UGC delivery. A key study in [18] shows that the dynamics of popularity of online contents experiences two phases. In the initial phase, a content gains popularity through advertisement and other marketing tools. Afterwards, UGC platform mechanisms induce users to access contents by re-ranking mechanisms. Those also appear to be main drivers of popularity.

Motivated by such findings, we model the behavior of those who create content – shortly sources in the rest of the chapter – as a dynamic game. Once they generate a content, in particular, they leverage on UGC platforms to diffuse it. We note that, by paying a fee for the advertisement service of the UGC platform, a content provider is able to receive a preferential treatment to her content in such a way that the rate of propagation is increased. Clearly, this leads to a competition among sources to capture the attention of potential viewers at faster rate than other contents.

[1] See: http://www.youtube.com/t/press_statistics/

To this respect, the notion of *acceleration* is a key concept. An example reported in Fig. 4.1 explains how a video can be accelerated by the UGC platform. In our example we performed a generic search *"Hakusai"* which produces a series of output results for matching contents. The one reported in the figure is one with viewcount 568507 that has been listed by the main YouTube.

In particular, Fig. 4.1 represents the viewer's screen: in the central part of the window stands the video. However, there exists a recommendation list on the right as provided by the platforms' search engine.

It is important to observe the two videos recommended on the top of the list. The first one is an advertisement of a known commercial activity. In order to appear in the top position of the list, that content has been paying a fee to the UGC platform owner. In the second position, a link appears to a video which is tagged *featured*. The meaning of the term featured is that the video linked there was placed high on the recommendation list either because it is a very popular video or because it is a partner video. A partner video, as in the case of advertisements, is from someone who pays a fee to rank higher in the recommendation list. The other videos in the recommendation list are ranked according to the default order, i.e., that induced by the recommendation system [97]. Another advertisement with the suggestion to buy a product is appearing at the bottom of the figure. In this chapter we focus on the acceleration of featured videos. In fact, in order to accelerate a video, customers perform a *promoted video campaign* on YouTube; to do so, content providers are required four steps: choose a video, attach promotional text, some keywords by which the promotion is performed, and set the *daily budget* amount allowed.

With respect to the acceleration cost, it is important to note that the so called pay-per-view model is applied. In other words, the YouTube pay-per-view policy for acceleration is meant to charge the source a fixed amount each time a viewer has accessed the content. Charging is triggered by a click-through on the icons of the promoted content which appear in the recommendation list. However, for the platform owner it is best that the customer's daily budget is attained. Then, the total cost paid in order to increment the number of views would increase linearly in time. A linear cumulative cost for acceleration is also one of our assumptions in the rest of the chapter.

Now, since the viewer' browser has finite size, only those who are able to appear in the higher end of the recommendation list are visible without scrolling. Thus, those are accessed with higher probability: the viewcount of a content is expected to grow faster, i.e., to be accelerated, whenever it is showed higher in the list.

In this work, we consider a competition between several contents. The promotion fees, i.e., the cost to accelerate the viewcount, will depend on the source and may depend on the content itself. Even the rate of propagation, i.e., the rate at which viewers access the content, may depend on the content. Finally, each source may decide whether or not to purchase priority to accelerate the popularity of the content for a certain period.

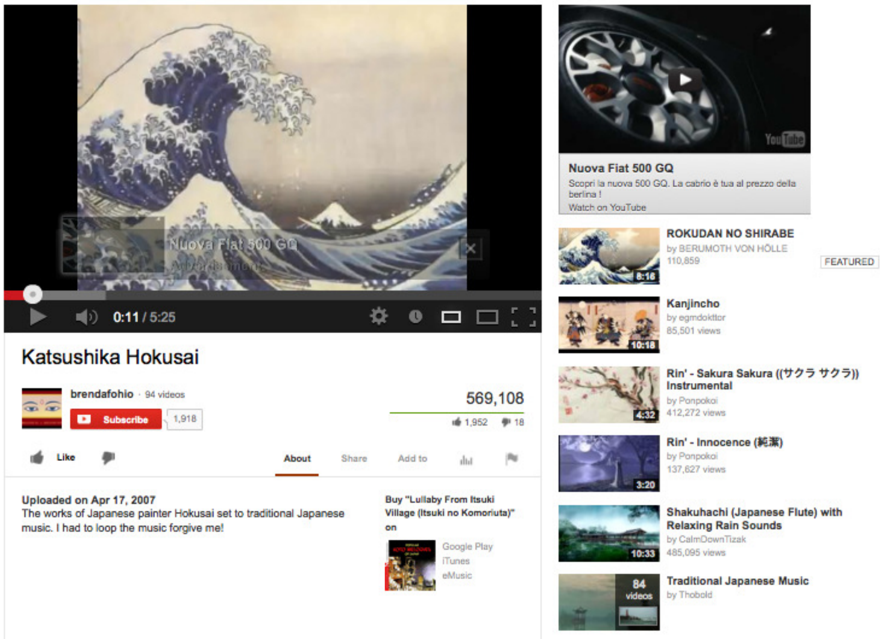


Figure 4.1: *A sample video: the icons on the right of the main window is the recommendation list. The upper entry seen there is a commercial advertisement, whereas the second entry is tagged "Featured". The remaining entries are sorted according to their viewcount.*

The objective of this chapter is to determine the best strategy for a source in order to accelerate a content and study the equilibria of the resulting system. To this aim, we propose a game theoretical framework rooted in differential games. The solution of the problem allows us to provide guidelines for the advertisement strategies of sources.

A brief outline of the chapter follows. In Sec. 4.2 we revise the main results of the literature for online content diffusion. In Sec. 4.3 we introduce the system model and the differential game subject of this chapter. In Sec. 4.4 we derive the analysis of best responses whereas symmetric Nash equilibria are characterized in Sec. 4.5. In Sec. 4.6 we tackle the limit case of small discounts. A section with conclusions, practical implications of our model and future research directions ends the chapter.

4.2 Related works

The dynamics of popularity of online contents has been attracting attention from the research community. [93] proposed an analysis of the YouTube system focusing on the characteristics of the traffic generated by that platform. [95] addressed the relation between metrics used to evaluate popularity, such as number of comments, ratings, or favorites. In this chapter our analysis is restricted to the viewcount.

In [94] the authors study the ranking change induced by UGC online platforms. Bursty acceleration in content's viewcount is found to depend on the way online platforms expose popular contents to users and on re-ranking of existing contents. Here, we model the competition that arises when several sources leverage such acceleration tools.

In the literature, competition in epidemic processes has been addressed with game theoretical tools. In [98], the authors focus on an economic game on graphs, where firms try to conquer the largest market share. They derive the complexity for the computation of the equilibria of the game; results for the price of anarchy in those games have been developed recently in [99].

The emergence of equilibria of the Wardrop type has been studied in [100] from viewer's perspective. Our objective here is to describe the source viewpoint in a dynamic game framework.

In [15] the authors consider information propagation through OSNs. The question there is how the finite budget of attention of individuals influences the rate at which contents can be pushed into the other players' network. In our work, we limit our focus to the case of online content diffusion in UGC platforms.

Novel contributions: In this chapter, we provide a complete framework for the analysis of dynamic games in UGC provision. Under a meanfield approximation for contents diffusion, differential games [43] provide the model for capturing the strategic behavior of competing sources. Our main findings are:

- The structure of the best response of sources and a method for calculating it;
- Conditions for the existence and uniqueness of symmetric Nash equilibria in threshold form;
- Approximated asymmetric Nash equilibria in the regime of small discounts.

To the best of the authors' knowledge, results on Nash equilibria for differential games in UGC provision have not been derived so far in the literature.

4.3 System Model

The main symbols used in the chapter are reported in Tab. 4.1.

In our system model we assume that N competing sources release a content each and viewers will access one of such contents at the earliest chance.

We assume a base of M potential viewers who can access each of the N contents. To this respect, we adopt a fluid approximation which is assumed to hold for large M, and we assume that content viewers access content i according to a point process with

Table 4.1: Main notations used throughout the chapter 4

Symbol	*Meaning*
N	number of players (sources)
λ_i	intensity of views per second for content i
τ	time horizon
$x_i(t)$	fraction of viewers having viewed content i at time t
$x(t)$	summation $\sum_i x_i(t)$
$y(t)$	$y(t) := 1 - x(t)$
z_i	$:=x_i(0)$ will be taken 0 unless otherwise stated.
$u_i(t)$	acceleration control (strategy) for player i; $\mathbf{u} = (u_1, u_2, \ldots, u_N)$
	$u_i(t) \in [u_{min}, u_{max}]$
	$1 \leq u_{min} < u_{max} < \infty$
$\mathbf{u}_{-i}(t)$	strategy profile for all players different from i
a	sum of the $\lambda_i u_i$
a_{-i}	sum of the $\lambda_j u_j$ for all players $j \neq i$
p_i	discount factor for player i, $p_i > 0$

intensity λ_i. In other words, content i is accessed by a randomly picked viewer every λ_i^{-1} seconds.

In general $\lambda_i \neq \lambda_j$: in fact, contents may experience diverse popularity, and therefore different intensities. Every source will participate to the content diffusion in some time frame $[0, \tau]$. Also, $\tau = \infty$ in the development of the infinite horizon formulation of the game.

Using the advertisement options of the platform, sources can pay a cost in order to accelerate the diffusion of their video: viewers will access the content according to the intensity $u_i\lambda_i$, where $u_i \geq 1$ is the acceleration control for player i. The maximum acceleration is bounded as u_{max}, and the minimum acceleration is u_{min}: $1 \leq u_{min} < u_{max}$.

Finally, there is a linear cost paid for the acceleration control: such a cost represents the ideal case, i.e., when a promoted content receives per day a certain number of new views per cent paid to the platform owner. Conversely, the case with no acceleration, namely, $u_i = u_{min}$, falls back to the default intensity $u_{min}\lambda_i$. Of course, this happens at zero cost. Also, we consider somehow an ideal platform where only the content owner who promotes the video can benefit from the acceleration.To do so, we assume that the platform owner acts in such a way to make the access exclusive, i.e., so that targeted viewers tend not acquire a competing content after accessing another one, e.g., by making recommendation sets disjoint [97].[2]

We introduce below the game model that we use to describe the competition among

[2]To confirm this rationale, we observe that featured videos never appear together in the recommendation list presented to users.

sources in order to accelerate the dynamics of the viewcount. The formulation of the problem is initially provided in general to cover both in the finite horizon and in the infinite horizon case. Within the scope of the chapter, our development is confined to the infinite horizon analysis.

4.3.1 Game model

The differential game model for online content diffusion is composed as follows.
Players: players are sources, who compete in order to diffuse their content over a base of potential viewers. Since all players share the same base, the formulation will result in a competitive differential game.
Strategies: the strategy of each player is the acceleration control. The control is thus dynamic, since each player should determine at each point in time the acceleration $u_i(t)$.
Utilities: the utility for player i is linear and has two terms. First, there is a cost paid for accelerating the content. Second, there is a revenue represented by the number of copies. The total utility is defined, as customary in differential games, as the integral of an instantaneous utility.

We denote by x_i the fraction of viewers who have accessed to the contents generated by the i-th source. The governing equation for the dynamics of the i-th content's viewcount is

$$\dot{x}_i = \lambda_i u_i (1-x), \quad i = 1,2,\ldots,N \tag{4.1}$$

where $x = \sum_{i=1}^{N} x_i$ is the total fraction of viewers who accessed some content; the initial condition is $x_i(0) = z_i$. Actually, (4.1) is a fluid approximation for the dynamics of the fraction of viewers of content i.

The fluid approximation which we use in this context can be justified formally with the derivation proposed by [101]. In particular, let $\widehat{\mathbf{X}}^{(M)}(t)$ be a N dimensional vector whose components are $\widehat{X}_i^{(M)}(t)$, for $i = 1,\ldots,N$. Here, $\widehat{X}_i^{(M)}(t)$ stands for the fraction of the potential viewers that watched the content at time t, when the basin of users has size M: it represents the branching process of the i-th content being watched. Thus, when we refer to fluid approximations that describe the dynamics of the fraction viewers watching the content, we are referring the meanfield approximation of such process. In particular, for a formal explanation of the convergence for large M to the fluid approximations of the type used hereafter, the reader can refer to [102].

The acceleration control u_i, namely the strategy of player i, belongs to the space of the piecewise continuous functions $\mathscr{U} = \{u \in \text{p.w.c. functions of } [u_{min}, u_{max}]^{[0,\tau]}\}$.

Hence, because the control is upper bounded, the above ODE system (4.1) is Lipschitz continuous, and because it is lower bounded, it is so uniformly in the control, so that a solution at large is guaranteed to exist unique for a given strategy profile $\mathbf{u} = (u_1, \dots, u_N)$ ([44], pp. 99).

The cost function for the i-th player is given by

$$J_i(x,u) = \int_0^\infty e^{-p_i s} \left(-\dot{x}_i(s) + \gamma_i(u_i(s) - u_{min})\right) ds$$
$$= \int_0^\infty e^{-p_i s} \ell(x_i, u_i) ds \qquad (4.2)$$

where $\gamma_i > 0$, $p_i \geq 0$ is a discount factor; here $\ell(x_i, u_i) = -\dot{x}_i + \gamma_i(u_i - 1)$.

A cautionary remark: in the infinite horizon case, the discount factor has the role of ensuring the existence of a finite cost. Besides that, looking at (4.2), we observe that a large value of $p_i > 0$ characterizes an "impatient" player who aims at fast dissemination of the content. Conversely, a "patient" player would use a small value of p_i.

In particular, we note that for $p = 0$ the cost function has a more familiar expression $J_i(x,u) = -x_i(\tau) + \gamma_i(\int_0^\tau (u_i(s) - u_{min}))ds$ where the dependence on the number of copies appears with no discount. In Sec. 4.6 we are studying an approximation of our differential game that provides closed form expression of the threshold type for the infinite horizon case for vanishing discounts. In that case, the first term of (4.2) can be approximated assuming very large values of τ so that

$$\int_0^\tau e^{-p_i s} \dot{x}_i(s) ds = \frac{e^{-p_i \tau} x_i(\tau) - z_i}{1 - p_i} \to x_i(\tau) - z_i \text{ as } p_i \downarrow 0$$

Finally, the problem we want to solve is thus to determine the optimal cost function, namely the *value function* $V_i(x)$.

Definition 4.1 (Best response). *For any strategy $\mathbf{u}_{-i}$ of the remaining players, determine the best response, i.e., the optimal control u_i^* of player i for which the value function is attained, i.e.,*

$$V_i(x) := \inf_{u_i \in \mathscr{U}} J_i(x, u_i). \qquad (4.3)$$

We will solve the problem using the discounted formulation in the infinite horizon: $p_i > 0$ for all players, and $\tau = \infty$.

4.4 Best response analysis

Best response strategies are determined using the Hamilton-Jacobi-Bellman equation (HJB) for the infinite horizon.

4.4.1 Infinite horizon with positive discount $p > 0$

The existence of the optimal cost function bounded and uniformly continuous is immediate from [44] Prop. 2.8, since the $\ell(\cdot,\cdot)$ is bounded and uniformly Lipschitz, it holds:

$$|\ell(x,u_i) - \ell(y,u_i)| \leq \lambda_i u_{max}|x-y|, \ell(x,u_i)| \leq \lambda_i u_{max}. \tag{4.4}$$

In particular, we can write the Hamiltonian for each one of the players with respect to the dynamical system (4.1) corresponding to the problem in (4.1). Before that, it is easy to see that the aggregated dynamics can be written as

$$\dot{x} = \sum \lambda_i u_i (1-x), \tag{4.5}$$

for which we will develop the optimal control for each one of the players having fixed the control of the competing ones. In particular, the optimal control needs to maximize the Hamiltonian

$$\begin{aligned} H_i(x,\zeta) \qquad &= \sup_{u\in\mathscr{U}} \{-\sum \lambda_i u_i(1-x)\zeta \\ &+ (u_i\lambda_i(1-x) - \gamma_i(u_i - u_{min}))\}. \end{aligned} \tag{4.6}$$

The maximization of (4.6) provides the closed loop solution of our problem. However, the optimal cost function V_i in turn is one solving the HJB equation

$$p_i V_i + H_i\left(x, \frac{\partial V}{\partial x}_i\right) = 0 \tag{4.7}$$

so that minimizing the cost function V_i is equivalent to maximize $H_i\left(x, \frac{\partial V}{\partial x}_i\right)$ [44], where $\frac{\partial V}{\partial x}_i = \frac{d}{dx}V_i(x)$. In turn, we can write in closed form based on (4.1) and (4.5).

$$\begin{aligned} H_i(x) \;&= u_{min}\gamma_i - \sum_{j\neq i} u_j\lambda_j(1-x)\frac{\partial V}{\partial x}_i \\ &+ \sup_{u_i\in\mathscr{U}}\left\{u_i[\lambda_i(1-x)(1-\frac{\partial V}{\partial x}_i) - \gamma_i]\right\}. \end{aligned} \tag{4.8}$$

Now, since we observe that the Hamiltonian is linear in u_i, if a maximum is attained at some control $u \in \mathscr{U}$, the intuition is that it may assume only extremal values as it is often seen in the case of open-loop type of solutions [43]. This is actually true: we can allow the control u to take values only on the vertices as the ones of the codomain polyhedron, i.e., $[u_{min}, u_{max}]^N$. The value function will be the same of the original problem where such restriction does not hold (see [44], pp.113). This is due to the fact that (4.1) is of the type $\dot{x} := f_1(x) + f_2(x)u$ and of the running cost function ℓ which is linear in the control.

Motivated by this observation, we are interested in a class of best responses, namely

Definition 4.2. *A strategy u_i is of the bang-bang type if it takes extremal values u_{min} and u_{max} only.*

Also, we denote $t_{ik}, k = 1,2,\ldots$ the switching times associated with best response u_i, and $[t_{ik}, t_{ik+1})$ represents the corresponding k-th switching period of player i. If we limit our analysis to the best responses of bang-bang type, then $u_i \in \{u_{min}, u_{max}\}$, $i = 1,\ldots,N$: best responses are in fact piecewise constants.

In particular, for the case of bang-bang strategies, the condition for optimality, i.e., a best response, writes from (4.8)

$$u_i(x) = \begin{cases} u_{min} & \text{if} \quad \lambda_i(1-x)(1-\frac{\partial V}{\partial x}_i) - \gamma_i < 0 \\ u_{max} & \text{if} \quad \lambda_i(1-x)(1-\frac{\partial V}{\partial x}_i) - \gamma_i > 0 \end{cases}. \tag{4.9}$$

We can denote by *switching interval* the interval of time between two consecutive switching instants: within such interval, the best responses are constant, i.e., $\mathbf{u}$ is constant and only assumes values in $\{u_{min}, u_{max}\}^N$. We can resume our findings above with the following theorem.

Theorem 4.3. *The value function V_i corresponding to the best response of the game can be attained by a strategy u_i of the bang-bang type.*

It is worth noting that in general the solution of HJB equations requires to search for a *viscosity solution* [44]. This is due to the fact that the classical solutions assuming the differentiability of the value function may not exist in general and thus require to solve a general notion of differentiation. Here, it is the structure of the system that spares us this step, since we know apriori that the best response is of the bang-bang type. The main difference with respect to the general case is that strategies u_is draw values in the finite set $u_i \in \{u_{min}, u_{max}\}$; again, compared to the general problem, this fundamental simplification is due to the linear structure of the source game.

4.4.2 Infinite horizon for $p_i > 0$

In order to decide on the sign of the above terms, we need to solve for the HJB equation in V_i. This is in fact possible, once we notice that (4.7) can be written as the following ODE

$$p_i V_i - a(1-x)DV_i + b_i(1-x) + c = 0, \tag{4.10}$$

where V_i is the value function that solves for the best response.

Here we simplify the notation by letting

$$\begin{aligned} c_i = & \quad \gamma_i(u_{min} - u_i), \\ b_i = & \quad \lambda_i u_i, \\ a = & \quad \sum \lambda_i u_i. \end{aligned} \tag{4.11}$$

Whenever convenient, for the sake of clarity, we will denote a by $a(u_i)$ to stress the fact that strategy profile a depends on the best response of player i, and we will also resort sometimes to the notation $a_{-i} := \sum_{j \neq i} \lambda_i u_i$ for the sum of piecewise constant controls played by all the remaining players. It is important to note that a and b_i are assumed constant during each switching interval.

Also, note that since we do not know the optimal control, the expression for solving (4.10) here depends on the specific switching interval. Hence, V_i will have a specific dependence on the control which we need to maximize aposteriori since we know that (4.3) holds.

The solution of the HJB equation above can be solved as (see App. 4.9)

$$V_i(x) = K \cdot (1-x)^{-\frac{p_i}{a}} - \frac{p_i b_i(1-x) + (p_i + a)c_i}{p_i(p_i + a)}, \tag{4.12}$$

where K is a real constant.

In the following considerations we need the closed form of the function that is maximized by the control u_i in (4.8)

$$T_i(u_i, a_{-i})(x) = u_i \left[(1 - \frac{b_i}{p_i + a})(1-x) - \frac{K p_i}{a}(1-x)^{-\frac{p_i}{a}} - \frac{\gamma_i}{\lambda_i} \right]. \tag{4.13}$$

As a first step, from (4.12) we can obtain information on the structure of the value function. In particular we resume some basic facts in the following.

Lemma 4.4. *i. The best response u_i^* has a finite number of switches.*
ii. There exists a threshold value of x_∞ such that $u_i^*(x) = u_{min}$ for all $x > x_\infty$ for every player i.

Furthermore, we can characterize immediately a class of problems where players have no incentive to accelerate anyway: in particular we see that the following sufficient condition holds:

Lemma 4.5 (Degenerate Nash Equilibrium)**.** *i. Consider $\lambda_i < \gamma_i$, then the best response for the player i is $u_i^* = u_{min}$ irrespective of the other players strategies. ii. If $\lambda_i < \gamma_i$ for $i = 1, \dots, N$, then $u_i^* = u_{min}$, $i = 1, \dots, n$ is the unique Nash equilibrium.*

From the above results, we can see that the general form of the best response for player i against the strategy profile $\mathbf{u}_{-i}$ for the remaining players can be determined by proceeding backwards from the latest switching value $x_{i,\infty}$ which is calculated first. Then, by continuity, the constant K appearing in (4.12) for the switching interval before the last one (where now player i would use u_{max}) can be determined by imposing the continuity of the value function. In fact, at the switching time, the expression (4.12) has different values of b_i and a_i in the two adjacent switching intervals. The procedure can be iterated backwards to determine all the threshold values of x when player i switches.

In the case of a symmetric game, i.e., when the sources have all the same parameters, the procedure described above can solve the game in closed form, as showed in the next section where symmetric equilibria are described.

4.5 Symmetric Nash Equilibrium

We consider now the symmetric case: $\lambda_i = \lambda$, $\gamma_i = \gamma$, $p_i = p$ for all $i = 1, \ldots, N$. The proofs of the statements hereafter are deferred to the Appendix. We consider tagged source i, and assume that all the remaining players use the same threshold type of strategy, i.e., of the type

$$u_j(x) = \begin{cases} u_{max} & \text{if } x < \hat{x} \\ u_{min} & \text{if } x > \hat{x} \end{cases} \tag{4.14}$$

for some $0 \leq \hat{x} \leq 1$. Denote x^* the last switch of player i: we are now ready to show that there exists a symmetric equilibrium where also player i will use $\hat{x} = x^*$, i.e., the threshold type strategy (4.14) is the best response to itself when all sources play it. Furthermore, it is the unique symmetric equilibrium of the game.

In particular such a Nash equilibrium is given by a threshold x^* which is derived by the form of the value function in the last switching interval (recall that $K = 0$ in that interval)

$$V_i(x) = -\frac{u_{min}\lambda(1-x)}{(p + N\lambda u_{min})}, \tag{4.15}$$

by imposing that $T(u_{min}, u_{min}(N-1))(x^*) = 0$ (switching condition). These results are detailed formally in the statements below.

Lemma 4.6. *Consider $x^* \geq \hat{x}$, then the following holds:*

i. Player i can switch at some $0 < x^* < 1$ iff $\lambda(1 - \frac{u_{min}\lambda}{(p + N\lambda u_{min})}) - \gamma > 0$; moreover,

$$x^* = 1 - \frac{\gamma(p + N\lambda u_{min})}{\lambda(p + (N-1)u_{min}\lambda)}. \tag{4.16}$$

ii. Consider all players switch at x^*: the constant K^* which ensures the continuity of the value function $V_i(x)$ at the switching threshold x^* is positive and the following relation holds

$$K^* = (u_{max} - u_{min})(1-x^*)^{\frac{p}{\lambda n u_{max}}} \left\{ \frac{(1-x^*)^{\frac{p}{\lambda u_{min}}}}{(N+\frac{p}{\lambda u_{max}})(N+\frac{p}{\lambda u_{min}})} + \frac{\gamma}{p} \right\}.$$

Theorem 4.7 (Symmetric Nash Equilibrium)**.** *If $\lambda(1-\dfrac{u_{min}\lambda}{(p+N\lambda u_{min})}) - \gamma > 0$, then the threshold type strategy (4.14) where $\hat{x} = x^*$ and x^* is as defined in Lemma 4.6 is the unique symmetric Nash equilibrium of the game.*

The existence of equilibria in the non symmetric case is the next question that we are answering. In particular, we obtain certain asymmetric equilibria which are ε-approximated Nash equilibria. In other words, the unilateral deviation from those strategy profiles may provide some improvement to the utility of a source. However, such improvement can be made arbitrarily small by choosing an appropriate value of the discount p_i.

4.6 Vanishing discount regime

Hereafter, we consider the cases of small discount factors. As described in Sec. 4.3.1, we can consider the case when p_i has a very small value. This means that player i does not pose much of a constraint on the time taken in order to make the content popular. In particular, we would consider the case of vanishing discounts sequences: $p_i(r) = o(1)$, $i = 1,\dots,N$ and consider the form of the best replies in the regime of vanishing discounts. This provides further insight into the structure of the equilibria for the sources game.

Corollary 4.8. *Consider $p_i(r) = o(1)$: there exists a best reply in threshold form for the i-th player that is arbitrarily close to the best reply of the game for a small enough discount factor.*

Because of the above result, we can always find a discount factor small enough so as to find a threshold type strategy which approximates the cost function of the best response within an arbitrarily small positive additive constant $\varepsilon > 0$. In turn, this also means that if a Nash equilibrium exists under the modified utility function $\tilde{T}(\cdot)$, then it is an ε-approximated Nash equilibrium in threshold policy for the original game. It is hence interesting to study the existence of a Nash equilibrium for the modified game.

Theorem 4.9 (Asymmetric ε- approximated Nash Equilibrium)**.** *If $\lambda_i = \lambda$ for $i = 1,\dots,N$ and $\lambda > \gamma_1 > \gamma_2 > \dots > \gamma_N$. Then, there exists an ε-approximated Nash equilibrium in threshold form for the original game.*

We note that the constructive proof of the ε-approximated Nash equilibrium states that if some player stops accelerating, it will not have incentives to accelerate later for larger values of the state x, because the increment in the state x is decreasing. Overall, the above result suggests that in the fully asymmetric case, the presence of diverse costs induces an equilibrium in threshold form where even if a source has an incentive in deviating from the given strategy profile, and therefore change strategy, the incentive that the player has in deviating can be made small at wish by choosing an appropriate value of the discount.

4.7 Finite horizon case

When there is a finite horizon $0 \le \tau < \infty$ under a nonnegative discount, the HJB equation becomes [44]

$$\dot{V}_i + pV_i + H(x, DV_i) = 0 \tag{4.17}$$

The natural initial condition $V_i(x,0) = 0$ for all $x \in \mathbb{R}$, because the terminal cost is null. Hence, the value function solves the following PDE

$$\dot{V}_i + pV_i - a(1-x)DV_i + b(1-x) + c = 0 \tag{4.18}$$

(4.18) is linear and the associated homogeneous PDE is

$$\dot{V}_i + pV_i - a(1-x)DV_i = 0 \tag{4.19}$$

whose solution is in the form $V_i^{om}(x,t) = \phi(a^{-1}\log(1-x) - t)(1-x)^{-\frac{p}{a}}$, where $\phi(v) : \mathbb{R} \to \mathbb{R}$ is a differentiable function.

Hence we just need a particular solution: we seek one such solution in the form $V_i(x,t) = V_{i,p}(x)$, so that it should solve

$$p_i V_i - a(1-x)DV_i + b(1-x) + c = 0$$

The solution is found to be:

$$V_i(x) = \begin{cases} (1-x)^{-\frac{p_i}{a}} - \dfrac{p_i b(1-x) + (p_i + a)c}{p_i(p_i + a)} & \text{if } p_i > 0 \\ \dfrac{b}{a}x - \dfrac{c}{a}\log(1-x) \quad \text{if } p_i = 0 \end{cases} \tag{4.20}$$

Finally, the solution of (4.18) is determined to be

$$V_i(x,t) = \phi(a^{-1}\log(1-x) - t)(1-x)^{-\frac{p_i}{a}} + V_{i,p}(x)$$

Since we are faced with a undetermined function ϕ, one per switching interval, we shortly describe how to calculate the best response. In the first switching interval $[0, t_1]$, the natural initial condition $V(x,0) = 0$ for all $x \in [0,1)$. This provides the closed form expression for $\phi(\cdot)$, which is found by

$$\phi(a^{-1}\log(1-x)-t)(1-x)^{-\frac{p}{a}} + V_p(x) = 0, \quad \forall x \in \mathbb{R}.$$

In particular, we obtain for the case $p > 0$:

$$\phi(v) = -1 + e^{pv}(\frac{b}{p+a}e^{av} + \frac{c}{p})$$

so that in the first switching interval we can state

$$V(x,t) = -(1-e^{-pt})[\frac{c}{p} + \frac{b}{p+a}(1-x)] \tag{4.21}$$

Now, once determined the best response for player i in the first switching interval by (4.8), we should impose the continuity condition on $V(x,t) = V(x,t_1)$. This provides the initial condition for the second interval. Proceeding to the subsequent intervals, the procedure can be iterated to determine the value function for the best response of player i. It is worth noting that in this case the switching thresholds will depend on time.

4.8 Numerical Results

We provide hereafted a numerical description of the results for the sources' best response. It is interesting to visualize the best reply of a certain source facing different strategies of the remaining ones. In Fig. 4.2a) and b) we reported on the best response in the vanishing discount regime as a function of a_{-i} in the homogeneous scenario, i.e., $\gamma_i = \gamma$ and $\lambda_i = \lambda_0$ for $i = 1,\dots,N$. Here, $u_{max} = 10$ and $u_{min} = 1$.

The graphs of Fig. 4.2a) and b) refer to two values $N = 10$ and $N = 30$, respectively. The value of λ_0 is settled to 100 views per day, whereas $\gamma = 0.7 \cdot \lambda$.

The best response is depicted using different markers for u_{min} and u_{max}: for a fixed value of a_{-i} the best response starts with u_{max} and switches to u_{min} above the threshold $x_{0,i}$. In both cases a) and b) the threshold value decreases with a_{-i}. This is competition's effect: larger a_{-i} values mean more players using u_{max}. Hence, the number of views that can be expected are small so that the cost for accelerating takes over. Hence the switch occurs at lower values of x.

Finally, we observe the floor at $1 - \frac{\gamma}{\lambda}$: this is the limit best response for a large number of players as from (4.23). For a very large number of competing contents, the best response of the single player should depend on γ and λ only. When this happens, in fact, the game becomes singular: a unique best response for every player exists which defines the only equilibrium of the system.

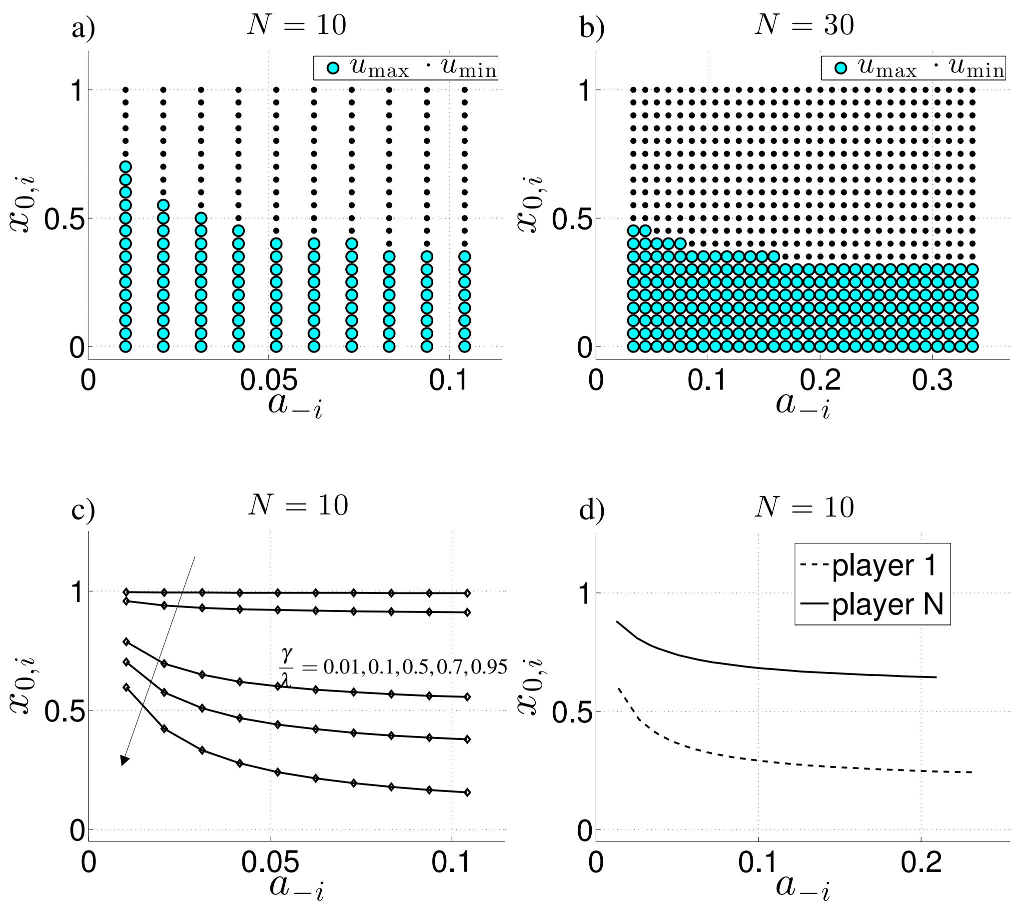


Figure 4.2: *a) and b): best reply for $\lambda_i = \lambda_0 = 100$ views/day, $u_{max} = 10$, $u_{min} = 1$, $p = \lambda$, and $\gamma_i = \gamma = 0.7\lambda$; a) $N = 10$ and b) $N = 30$; c): impact of cost on the best response; d) case for heterogeneous scenario: $\lambda_i = \lambda_0$ for $i = 1, \ldots, N/2$ and $\lambda_i = 2\lambda_0$ for $i = N/2+1, \ldots, N$, $N = 10$.*

In Fig. 4.2c) we observe the best response for $N = 10$ and for increasing values of γ/λ: clearly, larger values of the ratio makes players accelerate less because of increased cost. Again, we observe that the effect of competition is to reduce the acceleration for larger values of players using u_{max}, i.e., larger values of a_{-i}. However, for small values of the cost, not only the best reply is to use a large threshold, but the strategy of competing sources becomes less and less relevant so that the threshold becomes almost constant in a_{-i}.

In the last figure, i.e., Fig. 4.2d), we considered an heterogeneous scenario. In this case, contents $i = 1, \ldots, 5$ have $\lambda_i = \lambda_0$, whereas $i = 6, \ldots, 10$ have $\lambda_i = 2\lambda_0$. In this case we expect that a_{-1} and a_{-10} in the example may be different. Hence, we are interested in the relative behavior of best response of the two type of players.

As seen in the figure, the switching order $x_{0,1} < x_{0,10}$ is maintained for increasing values of a_{-i}: players with higher value of λ_i always switch before. Actually, even for a small number of players as in this example, the range of values taken by a_{-i} is basically the same for both types of players. As a consequence, the ratio γ_i/λ_i is the main parameter characterizing the relative behavior of the two classes of players, i.e., the switching order.

4.9 Conclusions

We have introduced models for advertisement in online content diffusion where sources compete to make contents popular. They leverage on acceleration tools of online platforms in order to increase the viewcount of their videos. By doing so, contents occupy positions that are more visible on the web pages of potential viewers. However, a fee is paid in order to profit from re-ranking mechanisms of online platforms.

Hence, each player, e.g., each source, needs to decide over time if it is worth to accelerate or not. And, this choice is dynamic depending on what other sources do. We introduced a model based on differential games: players' best responses are determined by solving an ODE involving the Hamiltonian of the system they observe.

We showed that in the infinite horizon case, the closed loop best replies of players are of the bang-bang type in the state x. We have identified a unique threshold-type Nash equilibrium in the symmetric case and we found that dual counterparts exist in the fully asymmetric case for small discounts.

Practical implications. We derived the existence of Nash equilibria in threshold form in the symmetric and (ε-approximate) in the asymmetric case. However, we conjecture that Nash equilibria in threshold form do exist also for any choice of the λ_is and the γ_is. Thus, sources would only pay when the total fraction of views is below a certain threshold and stop promoting above it. Now, we can observe (4.16) closer, and draw the following conclusions when the equilibrium is reached:
a) *all-or-nothing effect*: content i with low potential, i.e., very small λ_i, will not be accelerated at all when $\lambda_i < \gamma_i$. Hence, sources compete by promoting first contents likely to become most popular, e.g., with larger values of λ_i.
b) *best response and promotion*: the best response is of the threshold type, so that sources are able to maximize the number of views while minimizing the cost by using the maximum promotion budget per day until the threshold is reached.
c) *daily budget:* the daily budget γ determines the threshold x^* in such a way that the larger the cost which is paid per day, the lower x^*. From the platform owner perspective, for a given cumulative budget paid by a customer, the smaller the threshold, the lesser the promotion time will last. This is indeed convenient for the sake of system's resources: larger acceleration fares will lead to shorter promotion campaigns with smaller load for the platform's promotion mechanisms.

Future works. The results of our chapter indicate that these game models can lead to new tools for the pricing of online content advertisement and for the prediction of content popularity. There are several interesting research directions that are left for future work. First, the dynamic setting in the finite horizon case appears the most immediate extension. In future work, we plan to study the effect of the horizon duration onto the equilibria and the effect that time constraints have on sources' strategies. Another aspect relates to the number of competitors: in the case of large N, the strategy of single players does not change significantly other players' utility and the strategy profile $\mathbf{u}_{-i}$. To this respect, the dynamic game formulation could be reduced in the limit of large N to a static formulation which could be studied using Wardrop-like equilibria [100].

APPENDIX

Proof of lemma 4.4:

i. By contradiction: assume an infinite number of switches for player i and define constants K_r, $r = 1,\ldots,\infty$, for each such switching interval. By the continuity of x and the continuity of $V_i(x)$, together with (4.12), there exists an infinite sequence of K_rs which are non zero. Hence, since $x \uparrow 1$, by the continuity of x, we can find sequence $\{x_r\} \uparrow 1$ where x_r belongs to the r-th switching interval. Due again to (4.12), we found a subsequence of values of V_i which diverges. This is a contradiction since the value function is bounded.

ii. Denote $x_{i,\infty}$ the value of x above which the control switches to u_{min} for good: indeed, the constant appearing in (4.12) is zero. If it was not, again, V_i would grow unbounded as $x \to 1$. Hence, DV_i is bounded for $x > x_{i,\infty}$, so that by inspection of (4.6), the control needs to be u_{min} for values of x close to 1, and on the rest of the last switching interval as well since it is constant there. Finally, we can define $x_\infty = \max(x_{i,\infty}, 1 \le i \le N)$.

Proof of lemma 4.5:

i. This is a consequence of the statement in Lemma.4.4: in the last switching interval, since $K = 0$, from (4.13)

$$T_i(u_i, a_{-i})(x) \le u_i \left[\left(1 - \frac{\frac{b_i}{a(u_{min})}}{1 + \frac{p_i}{a(u_{min})}}\right) - \frac{\gamma_i}{\lambda_i} \right] < 0$$

The rightmost inequality is equivalent to

$$\frac{a_{-i} + p_i}{a(u_{min}) + p_i} < \frac{\gamma_i}{\lambda_i}$$

which can also be written as $(\lambda_i - \gamma_i)(a_{-i} + p_i) < \gamma_i \lambda_i u_{min}$. The statement follows once we observe that the previous result is independent of a_{-i}, i.e., the strategies played by all the other players.

ii. Follows immediately from i.

Proof of lemma 4.6:
(i) The proof is made in three steps. In the first step we prove that $T(u_{min},(N-1)u_{min})(x)$ is decreasing for $x > x^*$ (we use notation $T(x)$ when it does not generate confusion). In the second step we derive the sufficient condition in the assumptions. Finally, in the third step we compute x^* and the corresponding constant K^*.

Step 1. From (4.15), in the last switching interval we have

$$DV_i = \frac{u_{min}\lambda}{(p+N\lambda u_{min})}$$

If we plug in DV_i in $T(x)$ we finally have:

$$T(x) = (1-x)(1-\frac{u_{min}\lambda}{(p+N\lambda u_{min})}) - \frac{\gamma}{\lambda}$$

so that $DT(x) < 0$.

Step 2. Since $T(x)$ is decreasing in the last switching interval, a threshold when player i switches to u_{min} exists if and only if $T(0) > 0$, which is the assumption in the statement, namely $1 - \frac{u_{min}\lambda}{p+N\lambda u_{min}} > \frac{\gamma}{\lambda}$.

Step 3. The threshold x^* for player i is obtained by solving

$$T(x^*) = 0 \Leftrightarrow x^* = 1 - \frac{\gamma}{\lambda} \cdot \frac{p+\lambda N u_{min}}{p+\lambda(N-1)u_{min}}$$

Now we can assume for player i a switch occurs in x^*. Hence, we impose the continuity of the value function [44]. Because it is continuous on both sides of the threshold x^*, the limit values $V_i(x^-)$ for $x \uparrow x^*$ when $(u_i, a_{-i}) = (u_{max},(n-1)u_{max})$ and $V_i(x^+)$ for $x \downarrow x^*$ when $(u_i, a_{-i}) = (u_{min},(n-1)u_{min})$ need to be the same. This will determine constant K^*. The equation to be solved is thus

$$\begin{aligned} K^*(1-x^*)^{-\frac{p}{\lambda n u_{max}}} - \frac{\lambda u_{max}(1-x^*)}{p+\lambda N u_{max}} - \frac{\gamma}{p}(u_{max}-u_{min}) \\ = -\frac{u_{min}\lambda(1-x^*)}{p+N\lambda u_{min}} \end{aligned} \tag{4.22}$$

and the expression for K^* writes as in the statement. Finally, we observe from (4.22) that indeed, $K > 0$: in fact

$$\frac{u_{max}\lambda}{p+N\lambda u_{max}} > \frac{u_{min}\lambda}{p+N\lambda u_{min}}$$

which concludes the proof.

Proof of theorem 4.7:
We first need to ensure that when source i plays against (4.14) with $\hat{x} = x^*$ for all the remaining players, the switch for player i is unique. Indeed, for $x < x^*$ it holds

$$DT(u_i, a_{-i})(x) = \frac{-\lambda(1 - \frac{b_i}{a+p})(1-x)^{\frac{p}{a}+1} - \frac{\lambda^2 pK^*}{a^2}}{(1-x)^{\frac{p}{a}+1}}.$$

However, we note that we are in the assumptions of Lemma 4.6, so that $K^* > 0$. Thus, for $x < x^*$, indeed $DT(u_i, a_{-i})(x) < 0$ by inspection of the above equation. This ensures that there is not any other switch for player i, so that the strategy of player i is also threshold with $\hat{x} = x^*$. Indeed, threshold strategy (4.14) with $\hat{x} = x^*$ for all players is a best reply to itself for all players, so that it defines a Nash equilibrium for the game. The uniqueness of the equilibrium is obtained by the fact that (4.15) has a unique zero.

Proof of corollary 4.8:
If $\zeta_r = \dfrac{p_i(r)}{a}$, we can write (4.13) as

$$\begin{aligned}
T(u_i, a_{-i}) &= u_i \left[(1 - \frac{\frac{\lambda_i u_i}{\sum_k \lambda_k u_k}}{1+\zeta_n})y - K\zeta_r y^{-\zeta_r} - \frac{\gamma_i}{\lambda_i} \right] \\
&= u_i \left[(1 - \frac{\lambda_i u_i}{\sum_k \lambda_k u_k})(1 - \zeta_r + o(\zeta_r))y - \frac{\gamma_i}{\lambda_i} + K\zeta_r(yu_i - \zeta_n + o(\zeta_r)) \right] \\
&= u_i[(1 - \frac{\lambda_i u_i}{\sum_k \lambda_k u_k})y - \frac{\gamma_i}{\lambda_i}] + \zeta_r f(y) + o(\zeta_r)
\end{aligned}$$

where $y = 1 - x$ and $f(y) = u_i(y\dfrac{\lambda_i u_i}{\sum_k \lambda_k u_k} - 1/y^{\zeta_r})$. We already noticed that there exists x_∞ above which every player switches to u_{min}: we can hence restrict our discussion to the range $y \in [1 - x_\infty, 1]$. Indeed, $f(y)$ is bounded therein: denote $\tilde{T}(u_i, a_{-i}) = u_i(\dfrac{\lambda_i u_i}{\sum_k \lambda_k u_k} - \dfrac{\gamma_i}{\lambda_i})$.

Hence, we can fix $\varepsilon > 0$ and consider $r > r_o$ such that $|\tilde{T}(u_i) - T(u_i)| < \varepsilon$ uniformly in y. The best response of user i will at most produce a value function that differs by ε from the one which maximizes (4.8). Hence, we search for the solution of the maximization problem

$$\tilde{u}_i^* = \operatorname*{argmax}_{u_i \in \{u_{min}, u_{max}\}} \tilde{T}(u_i)$$

which corresponds to a modified game where the cost function is $\tilde{T}$. We hence need to state when $\tilde{T}(u_{min}, a_{-i})$ is larger or smaller than $\tilde{T}(u_{max}, a_{-i})$: this turns out to be equivalent to the condition for the state x to exceed or not the threshold

$$x_{0,i} := 1 - \frac{\gamma_i}{\lambda_i} \frac{1}{(1+\frac{\lambda_i u_{min}}{a_{-i}})(1+\frac{\lambda_i u_{max}}{a_{-i}})} \tag{4.23}$$

so that the final control law that governs the best reply of the i-th player. is

$$\tilde{u}_i^*(x) = \begin{cases} u_{max} & \text{if} \quad x \leq x_{0,i} \\ u_{min} & \text{if} \quad x > x_{0,i} \end{cases}, \tag{4.24}$$

which concludes the proof.

Proof of theorem 4.9:
We assume the regime of vanishing discounts in a such way that ε is defined in the sense of Thm. 4.8. The proof is based on the following observation: at time 0, $u_i^*(0) = u_{max}$ because of (4.23) and $\lambda > \gamma_i$ for all is. Clearly, since $\gamma_1 > \gamma_i$ for $i > 1$, then $x_{0,1} < x_{0,i}$ for $i > 1$, and t_1 corresponds to the switch of node 1 from u_{max} to u_{min}. Also, $a_{-i}(t_1^-) = u_{max}N > u_{max}(N-1) + u_{min} = a_{-i}(t_1^+)$ for all players $i > 1$, so that $x_{0,i}(t_1^-) < x_{0,i}(t_1^+)$. Thus, $u_i^*(t_1^-) = u_i^*(t_1^+) = u_{max}$. Finally, until $x < x_{0,2}(a_{-i}(t_1^+))$, all players different from i will use u_{max}.

By induction: assume that first $k-1$ players that switched from u_{max} to u_{min} did not switch back and prove that under the conditions in the assumptions even the k-th player will never switch from u_{min} to u_{max}.

In order to proceed further with the proof we need to precise some notations

- $x_{0,i}(k)$ is the threshold (4.23) for player i when k players already switched to u_{min};
- $a_{-i}(k)$ is the sum of the other players $\lambda_j u_j$ when k of them switched to u_{min}.

At the time when player $k+1$ switches, it holds $x(t_{k+1}) = x_{0,k}(k)$. Hence, in order for player $k+1$ not to switch back to u_{max}, it must hold that $x(t_{k+1}) > x_{0,k}(k+1)$. However, we know that the dynamics in the k-th switching period is

$$x(t_{k+1}) = 1 - (1 - x_{0,k}(k-1))e^{-a_k(t_{k+1}-t_k)}.$$

Also, by the inductive assumption, $x_{0,k}(k-1) = x_{0,k}(k)$ because no player switched back to u_{max}. Then, since a_k is constant in the k-th switching period, the dynamics in that interval are governed by condition

$$\frac{1 - x_{0,k}(k+1)}{1 - x_{0,k}(k)} > e^{-a_k \Delta t_{k+1}} \tag{4.25}$$

Moreover, we have a condition on $\Delta t_{k+1} = t_{k+1} - t_k$:

$$\begin{aligned} x_{0,k}(k) &= 1 - (1 - x_{0,k}(k))e^{-a_k(t_{k+1}-t_k)} \\ &\Rightarrow \Delta t_{k+1} = \log(\frac{1 - x_{0,k+1}(k)}{1 - x_{0,k}(k)}). \end{aligned} \tag{4.26}$$

combining (4.25) and (4.26) we obtain

$$\frac{1-x_{0,k}(k+1)}{1-x_{0,k}(k)} > \frac{1-x_{0,k+1}(k)}{1-x_{0,k}(k)} \Rightarrow x_{0,k}(k+1) < x_{0,k+1}(k) \tag{4.27}$$

We can now express the above condition by considering the explicit expression

$$x_{0,k}(k+1) = \frac{\gamma_k}{\lambda} \frac{1}{(1+\frac{u_{min}}{v_k})(1+\frac{u_{max}}{v_k})},$$

where $v_k := \dfrac{a_{-k}(k+1)}{\lambda_k}$. Also, in the same way,

$$x_{0,k+1}(k) = \frac{\gamma_k}{\lambda} \frac{1}{(1+\frac{u_{min}}{v_{k+1}})(1+\frac{u_{max}}{v_{k+1}})},$$

where $v_{k+1} := \dfrac{a_{-(k+1)}(k)}{\lambda_k}$. Finally, we observe that

$$v_k = (N-1-k)u_{max} + ku_{min} = v_{k+1},$$

so that the condition in (4.27) becomes $\gamma_k > \gamma_{k+1}$ which is true according to our assumptions. Hence the inductive step is complete and the statement is true.

ODE solution

The solution of (4.10) is equivalent to the solution of $DV_i - \dfrac{p_i}{a(1-x)}V_i = \dfrac{b_i}{a} + \dfrac{c}{a(1-x)}$, so that it is sufficient to observe that the integrating factor for this first order ODE is $(1-x)^{\frac{p_i}{a}}$. Hence the solution follows from

$$\begin{aligned} V_i(x) &= (1-x)^{-\frac{p_i}{a}} a^{-1} \int (1-x)^{\frac{p_i}{a}} \left(b_i + \frac{c_i}{1-x}\right) dx \\ &= K(1-x)^{-\frac{p_i}{a}} - \frac{b_i}{a+p_i}(1-x) - \frac{c}{p_i} \end{aligned} \tag{4.28}$$

where K is an arbitrary real constant.

Chapter 5

Optimization problems in Online Social Networks: maximization of popularity and minimization of gossip propagation

Contents

This chapter is based on two papers *Controlling the Katz-Bonacich Centrality in Social Network: Application to gossip in Online Social Networks* and *A Continuous Optimization Approach for Maximizing Content Popularity in Online Social Networks*. In this chapter, the first section is dedicated to the minimization of gossip and the second section is dedicated to the maximization of popularity.

5.1 Controlling the Katz-Bonacich Centrality in Social Network: Application to gossip in Online Social Networks

5.1.1 Introduction

Centrality measures are valuable keys to understanding Social Networks. The most famous ones are the degree centrality, the betweenness centrality, the closeness centrality and finally the eigenvector centrality [48]. The popularity of these measures comes from the fact that they are concerned with links between a given node and the overall network (a micro perspective) as opposed to the diameter of a graph, the small world property, etc. (macro perspective). In particular spectral centrality measures of a node depend on the whole topology of the network. This is one of the main differences between spectral centrality measures and the others. Among the huge number of applications of spectral centralities, we will mention two works as examples. In [78], the authors characterize the optimal targeted marketing strategies in a Social Network by a spectral centrality measure. The second recent example of application is [103], where authors use a spectral centrality measure to find the delinquent who, once sent to jail, decreases the maximum of the population of delinquency profiles.

The control of the spectral centrality score associated with each node, by modifying the graph topology, is a recent problem. To the best of our knowledge, the very first paper on this topic concerns the maximization of the PageRank, where a webmaster controls multiple pages and hyperlinks between them (see related work of [104]). In the framework of delinquency, the paper [103] proposes to model a network of delinquents who communicate between themselves and to design an algorithm to characterize the link to be removed to minimize the overall rate of delinquency. In [105] the authors try to find the minimum set so that the vector of score is fully controlled.

There are several spectral centrality measures: the eigenvector centrality, the alpha centrality, the Katz-Bonacich centrality, the PageRank[1] and the subgraph centrality [106]. We restrict ourselves to the Katz-Bonacich centrality. Indeed, in this case we are able to derive an efficient algorithm to control this centrality by choosing which nodes

[1] see [48] for definitions of the above spectral centrality measures

to remove. The score given by the Katz-Bonacich centrality to a node is based on a discounted sum of the walks that initially have started from it. There are applications in Social Network that use it. For instance, in [103], the authors use the Katz-Bonacich centrality to characterize the key delinquent. In other contexts, the Katz-Bonacich centrality has been used in a pricing context [107], to characterize the Nash equilibrium of a Cournot competition over a network [83] and in other economic and social network applications [108; 109; 110].

This section focuses on deriving an efficient algorithm to optimize the Katz-Bonacich centrality by removing nodes. We prove that it is equivalent to a linear programming formulation and that there is a polynomial solution algorithm. Thus algorithms from [111] can be used in the context of large graphs. Once the linear programming characterization is provided, our goal is to propose an application of the control of the Katz-Bonacich centrality based on the minimization of a gossip process over an online Social Network.

After a short related works section, we introduce, in section 5.1.3, the Katz-Bonacich centrality optimization problem. We recall the mathematical definition of Katz-Bonacich centrality. We then prove its equivalence with a linear programming problem. In section 5.1.4, we apply the Katz-Bonacich centrality minimization problem to the control of a gossip process over an OSN. Finally, we compute the solution of the Katz-Bonacich centrality minimization problem on real networks in section 5.1.5.

5.1.2 Related works

The control of the spectral centrality has been studied from different points of view:

Topology Control. Lately, the question of how to modify the topology of the network in order to control the centrality score of nodes became a subject of interest.
In [105], an algorithm is proposed to find the minimum controlling centrality subset of nodes in a complex network in the particular case of the eigenvector centrality. The authors only control interactions generated by nodes inside this subset.
In [104] the authors noticed the links between the optimization of the PageRank and the ergodic Markov decision problem. Based on it, they provide an efficient algorithm to optimize the PageRank.

Key node. Another question has also been investigated on how the centrality score of nodes evolves when a node is removed from the network.
In [103; 112], the authors highlight the equivalence between the delinquency effort level and the Katz-Bonacich centrality. Then they find which delinquent has to be in jail such that the global delinquent profile decreases to the maximum.
In [113; 114], the authors investigate the control of the eigenvalue centrality. Their study is based on the theory of control of linear system and the well-known Kalman's

Table 5.1: Main notations used throughout the section 5.1

Symbol	*Meaning*
$\mathscr{G}(\mathscr{I},E)$	social network
$\mathscr{I}$	set of nodes
E	adjacency matrix
ρ	discount rate
p_i	probability to keep node i
$\mathbf{x}^*(E,\rho)$	Katz-Bonacich centrality associated with $\mathscr{G}(\mathscr{I},E)$ and $\rho \in \mathbb{R}_+$
C	number of regions in the graph
ϕ_c	constraint over the Katz-Bonacich centrality in region c
$Y_i(n)$	cumulative interest of node i at time n; $x_i(n) = \dfrac{Y_i(n)}{n}$ $\mathbf{x}(n) := (x_1(n),\dots,x_I(n))$
λ_i	rate at which node i receives news
α_i	rate at which node i updates its interest

controllability rank condition [115]. They find which node can control the whole system based on results from [116].

5.1.3 Control of the Katz-Bonacich centrality

As previously mentioned, the focus of this study is to provide an efficient algorithm to solve the Katz-Bonacich centrality optimization problem. We first recall the definition of the Katz-Bonacich centrality measure. We then, formally, describe the Katz-Bonacich centrality optimization problem. The main symbols used in this section are reported in Tab. 5.1. We begin our analysis by recalling the definition of the Katz-Bonacich centrality. In one of his works [117], made in the 50s, Katz proposed to model the centrality or the prestige of a node in a network in the following manner: the score associated with a node is based on a discounted sum of the walks that initially started from it. Nowadays, this centrality measure is called Katz-Bonacich centrality because in [118] Bonacich proposed a similar spectral centrality measure. We next provide a formal definition of it.

The social network $\mathscr{G}(\mathscr{I},E)$ is composed of a set $\mathscr{I} := \{1,\dots,I\}$ of nodes and the interactions between them is described by a communication matrix E. The variable $e_{ij} \in [0,1]$ denotes the ij^{th} entry of E and represents the relative frequency of interaction between node $i \in \mathscr{I}$ and node $j \in \mathscr{I}$. Moreover, for each i, we assume that $\sum_j e_{ij} = 1$ except when a node does not communicate with any other nodes and therefore $\sum_j e_{ij} = 0$. The definition of the Katz-Bonacich centrality [117], is given by:

Definition 5.1. *Consider $\rho \in [0,1]$. The Katz-Bonacich centrality associated with $\mathscr{G}(\mathscr{I},E)$, denoted by $\mathbf{x}^*(E,\rho) \in \mathbb{R}_+^I$, is given by:*

$$\mathbf{x}^*(E,\rho) := \sum_{n=0}^{\infty} (\rho E)^n \mathbf{1}_I, \tag{5.1}$$

where $\mathbf{1}_I$ is the all ones vector of size I. Thus the Katz-Bonacich centrality is the limit, as t goes to infinity (whenever it exists), of the following deterministic process, where for each $t \in \mathbb{N}^*$:

$$x_i(t+1) = 1 + \sum_j \rho e_{ij} x_j(t), \forall i. \tag{5.2}$$

Moreover, if the Perron Frobenius eigenvalue (see [119]), $\lambda_{max}(\rho E)$, associated with the matrix ρE is smaller than one, then:

$$\mathbf{x}^*(E,\rho) = (Id_I - \rho E)^{-1} \mathbf{1}_I, \tag{5.3}$$

where Id_I is the identity matrix of size $I \times I$.

The problem we are interested in is the minimization (or the maximization) of the Katz-Bonacich centrality by removing nodes, in other words by controlling the communication matrix E. The variable $p_i \in [\underline{p}_i, \overline{p}_i]$ denotes the probability of keeping a node i and $(1-p_i)$ the probability of removing it. Thus the Katz-Bonacich centrality is now in expectation, for each i, the limit $\lim_{t\to\infty} \mathbf{E}[x_i(t)]$ (which exists if $\lambda_{max}(\rho E) < 1$) of the following dynamical system:

$$\begin{aligned} \mathbf{E}[x_i(t+1)] &= \mathbf{E}[1 + \sum_j \rho e_{ij} x_j(t)] && (5.4) \\ &= p_i(1 + \sum_j \rho e_{ij} \mathbf{E}[x_j(t)]). && (5.5) \end{aligned}$$

Moreover, we assume that the Katz-Bonacich cannot be lower than a certain level in some regions of the graph. Let $c \in \{1,\ldots,C\}$ be a particular region of the graph. Let $N(c) \subset \mathscr{I}$ be the subset of nodes that belongs to the region c. For each c and c', we assume that $N(c) \cap N(c') = \emptyset$. For each region c, the constraint over Katz-Bonacich centrality is:

$$\sum_{j \in N(c)} x_j^* \geq \phi_c, \tag{5.6}$$

where $\phi_c \geq 0$. These region constraints come from the fact it is not possible to remove all the nodes of the graph. The Katz-Bonacich centrality minimization problem is therefore defined as follows :

Definition 5.2. *For each i, let $p_i \in [\underline{p}_i, \overline{p}_i]$ be the probability of removing node i and $\mathbf{p} := (p_1,\ldots,p_I)$ the associated vector. Let $\mathbf{P} := \Pi_i [\underline{p}_i, \overline{p}_i]$ be the set of constraints. Consider*

$\rho \in [0,1]$. *Let E be a sub-stochastic matrix and $\lambda_{max}(\rho E) < 1$. Let $\phi := (\phi_1, \ldots, \phi_C)$ be the region constraints. The Katz-Bonacich centrality minimization problem associated with (E, ρ) is defined as:*

$$\min_{\mathbf{p}\in\mathbf{P}} \sum_i x_i^*(\mathbf{p}), \tag{5.7}$$

where for each i, x_i^* is the unique solution of:

$$x_i^*(\mathbf{p}) = p_i(1 + \sum_j \rho e_{ij} x_j^*(\mathbf{p})), \tag{5.8}$$

such that for each c,

$$\sum_{j\in N(c)} x_j^*(\mathbf{p}) \geq \phi_c. \tag{5.9}$$

Our main result is to provide an equivalence between the Katz-Bonacich centrality minimization problem and a linear programme. Our goal is to first compute, for a given centrality, a closed form of the control that allows us to reach it. Then given a particular matrix E and a scalar ρ we will describe the set of feasible centralities. Finally, we will prove the equivalence between the Katz-Bonacich centrality minimization with a linear programme. The region constraints will only appear in the formulation of the linear programme.

Proposition 5.3. *Consider $\mathbf{k} \in \mathbb{R}_+^I$, E and ρ. If for each i,*

$$\underline{p}_i \leq \frac{k_i}{1 + \sum_j \rho e_{ij} k_j} \leq \overline{p}_i, \tag{5.10}$$

and if for each i, x_i^* is the unique solution of:

$$x_i^*(\mathbf{p}^*) = p_i^*(1 + \sum_j \rho e_{ij} x_j^*(\mathbf{p}^*)), \tag{5.11}$$

where

$$p_i^* = \frac{k_i}{1 + \sum_j \rho e_{ij} k_j}, \forall i \tag{5.12}$$

then $x_i^*(\mathbf{p}^*) = k_i$ for all i.

Now we are interested in understanding the set of feasible centralities, in other words to characterize the following set:

$$\mathbf{F} := \left\{ \mathbf{k} \in \mathbf{R}_+^I \mid \exists \mathbf{p} \in \mathbf{P} \text{ such that } \mathbf{k} \text{ sol. of } (5.11) \right\}. \tag{5.13}$$

The next proposition characterizes this set:

Proposition 5.4. *Consider E and ρ.*

$$\mathbf{F} = \left\{ \mathbf{k} \mid \forall i,\ k_i \geq 0,\ \underline{p}_i \leq \frac{k_i}{1 + \sum_j \rho e_{ij} k_j} \leq \overline{p}_i, \right\}. \tag{5.14}$$

Now according to the two previous propositions we can deduce that the Katz-Bonacich centrality minimization problem is equivalent to:

$$\min_{\mathbf{k}} \sum_i k_i \tag{5.15}$$

subject to for each i and each c:

$$\underline{p}_i \left(1 + \sum_j \rho e_{ij} k_j \right) \leq k_i, \tag{5.16}$$

$$\overline{p}_i \left(1 + \sum_j \rho e_{ij} k_j \right) \geq k_i, \tag{5.17}$$

$$\sum_{j \in N(c)} k_j \geq \phi_c, \tag{5.18}$$

$$0 \leq k_i. \tag{5.19}$$

Once this linear programme is solved the associated control is the one proposed in proposition 5.3. Because it exists a linear programme equivalent to the Katz-Bonacich centrality minimization problem, we can use algorithm proposed in [111] for large graphs.

5.1.4 Which nodes to remove in a gossip process

The problem we are interested in is the minimization of the gossip propagation by removing nodes. We call this challenge the Gossip Minimization problem. As proposed in [109] the owner of the OSN (the controller) is allowed to block users. For instance, by sending a warning to the friends of a user, the controller disturbs communication between them. A different take of this section can be found in [120].

Let $\mathscr{I} := \{1, \ldots, I\}$ be the set of users and $i \in \mathscr{I}$ be a user index. A user i of a social network gets news, about the gossip according to a Poisson point process of intensity $\lambda_i \in \mathbb{R}_+$. Thus when a news arrival occurs, concerning the gossip, the probability that it is for user i is:

$$\frac{\lambda_i}{\sum_j \lambda_j}.$$

Let $t_n \in \mathbb{R}_+$ be the arrival rate of the n^{th} message. When a user i receives news, it increases his belief in the gossip. Thus the interest of an average user in the gossip at time t_n, is described by a random variable $Y_i(n) \in \mathbb{N}$. Let $x_i(n) := \dfrac{Y_i(n)}{n}$ be the time average interest of user i at time t_n. Thus, for each i the evolution of Y_i is described by:

$$Y_i(n+1) = Y_i(n) + \zeta_i(n),$$

where the interest update is modeled by:

$$\zeta_i(n) := \begin{cases} 1 & \text{w.p} \quad \dfrac{\lambda_i}{\sum_j \lambda_j}, \\ 0 & \text{w.p} \quad 1 - \dfrac{\lambda_i}{\sum_j \lambda_j}. \end{cases} \tag{5.20}$$

As an extension of the previous model, we consider that interest imitation can occur between users. Let $E \in [0,1]^{I \times I}$ be the imitation matrix, where the ij-entry of E, e_{ij}, is the probability that a user i imitates interest of user j. For each i, we assume that $\sum_j e_{ij} = 1$. The time instant when a user i decides to imitate one of his neighbours is modeled by a Poisson point process of intensity $\alpha_i \in \mathbb{R}_+$. An event is now the arrival of messages or the activation of a user who wants to imitate someone. When an event occurs, the probability that it is the imitation phase of user i is $\dfrac{\alpha_i}{\sum_j \lambda_j + \alpha_j}$. At event n, when user i imitates user j, he will add one to $Y_i(n)$ with probability $x_j(n)$. Thus the new version of $\zeta(n) := (\zeta_1(n), \dots, \zeta_I(n))$, associated with the evolution of $Y = (Y_1, \dots, Y_I)$, is described below:

$$\zeta_i(n) = \begin{cases} 1 & \text{w.p} \quad P_i(\mathbf{x}(n)) := \dfrac{\lambda_i}{\sum_j \lambda_j + \alpha_j} + \alpha_i \dfrac{\sum_j e_{ij} x_j(n)}{\sum_j \lambda_j + \alpha_j} \\ 0 & \text{w.p} \quad 1 - P_i(\mathbf{x}(n)). \end{cases} \tag{5.21}$$

Following the theory developed in [46], the next theorem provides a sufficient condition about the convergence of the sequence $\{\mathbf{x}(n)\}$.

Theorem 5.5. *[46] Let $\mathbf{x}^0 \in [0,1]^I$ be the initial conditions. If the matrix A defined as follows,*

$$a_{ij} := \frac{\alpha_i e_{ij}}{\sum_{i''=1}^{I} \lambda_{i''} + \alpha_{i''}} - 1_{i=j},$$

is irreducible then the sequence $\{\mathbf{x}(n)\}$ converges almost surely to an unique rest point $\mathbf{x}^* \in [0,1]^I$. Moreover,

$$\mathbf{x} = A^{-}1\Lambda, \tag{5.22}$$

where $\Lambda := (\dfrac{\lambda_1}{\sum_{i''=1}^{I} \lambda_{i''} + \alpha_{i''}}, \dots, \dfrac{\lambda_I}{\sum_{i''=1}^{I} \lambda_{i''} + \alpha_{i''}}).$

It is interesting to note that the rest point of the stochastic approximation (5.21) is the Katz Bonacich centrality of the graph defined by the following matrix C, where the ij-entry is given by

$$c_{ij} := \frac{\alpha_i e_{ij}}{\sum_{i'} \lambda_{i'} + \alpha_{i'}}.$$

Control description. In the Gossip minimization problem, the controller can decide which users to block. More precisely, he can reduce the impact of user i over his neighbours using a control $p_i \in [\underline{p}_i, \overline{p}_i]$, such that the interest of each user i is solution of the following linear system:

$$x_i^* = p_i(\frac{\lambda_i}{\sum_j \lambda_j + \alpha_j} + \alpha_i \frac{\sum_j e_{ij} x_j^*}{\sum_j \lambda_j + \alpha_j}). \tag{5.23}$$

Moreover, because the controller cannot fully control the interest of each user, the sum of the interests of each user cannot be lower than a certain level:

$$\sum_j x_j^* \geq \phi, \tag{5.24}$$

where $\phi > 0$.

Utility description. In the present section, we assume that the controller wishes to minimize a utility vector depending on the interest of each user. Then the Gossip minimization problem is:

$$\min_{\mathbf{p} \in \mathbf{P}} U(\mathbf{x}^*). \tag{5.25}$$

where for each i, x_i^* is the unique solution of:

$$x_i^* = p_i(\frac{\lambda_i}{\sum_j \lambda_j + \alpha_j} + \alpha_i \frac{\sum_j e_{ij} x_j^*}{\sum_j \lambda_j + \alpha_j}), \tag{5.26}$$

subject to

$$\sum_j x_j^* \geq \phi. \tag{5.27}$$

We shall be interested in reducing the overall interest, in other words:

$$U(\mathbf{x}) := \sum_i x_i^*. \tag{5.28}$$

Finally, according to the previous section 5.1.3, the Gossip minimization problem is equivalent to the following linear programming:

$$\min_{\mathbf{k}} \sum_i k_i \tag{5.29}$$

Network	I	$\overline{d}$	$\sum_i x_i^*$	$\sum_i x_i^*(\mathbf{p})$
Zachary's karate club [122]	34	9.2	37.8	20
Miserables [123]	77	13.2	85.5	20
Network of American college football games [124]	115	21.3	127.8	20
Dolphin social network [125]	62	10.3	68.9	20

Table 5.2: *Summary table*

subject to for each i:

$$\underline{p}_i \left(\frac{\lambda_i}{\sum_j \lambda_j + \alpha_j} + \alpha_i \frac{\sum_j e_{ij} k_j}{\sum_j \lambda_j + \alpha_j} \right) \leq k_i, \tag{5.30}$$

$$\overline{p}_i \left(\frac{\lambda_i}{\sum_j \lambda_j + \alpha_j} + \alpha_i \frac{\sum_j e_{ij} k_j}{\sum_j \lambda_j + \alpha_j} \right) \geq k_i, \tag{5.31}$$

$$\sum_j k_j \geq \phi, \tag{5.32}$$

$$0 \leq k_i. \tag{5.33}$$

It is easy to generalize to a convex utility function. Indeed, as $U(\mathbf{x_i}^*)$ is convex in $\mathbf{x_i}^*$, we can use classical convex optimization algorithm [121] to solve this problem.

5.1.5 Simulations

In this section we compute over different networks the solution of the Katz-Bonacich centrality minimization problem. We study it over 4 different topologies depicted in fig. 5.1. The summary of experimental results are described in table 5.2. The first column provides the number of nodes, the second column the average degree ($\overline{d}$). We study each network with its normalized adjacency matrix E, where the ij-entry is given by: $e_{ij} := \frac{a_{ij}}{\sum_j a_{ij}}$, where A is the real adjacency matrix. For each network, the Katz-Bonacich centrality is computed with $\rho = 0.1$. The centrality average value of each network is given in the third column. We applied the Katz-Bonacich centrality minimization problem with, $C = 1$, $\phi_C = 20$ and for each i, $\underline{p}_i = 0.1$ and $\overline{p}_i = 1$. The fourth column describes the utility of the Katz-Bonacich centrality minimization problem. The first observation is, that for each network, the constraint $\sum_i x_i^* \geq 20$, is saturated. It follows from the fact that, for each network, the aggregated Katz-Bonacich centrality is bigger than 20. We expect that if we increase $\underline{p}_i$ this constraint will not be saturated anymore. When we look at fig. 5.1, where the size of the node is proportional to the solution of the Katz-Bonacich centrality minimization problem (x_i^* for each i), we remark that we do not have a trivial solution such as only one node being active or each node having the same centrality.

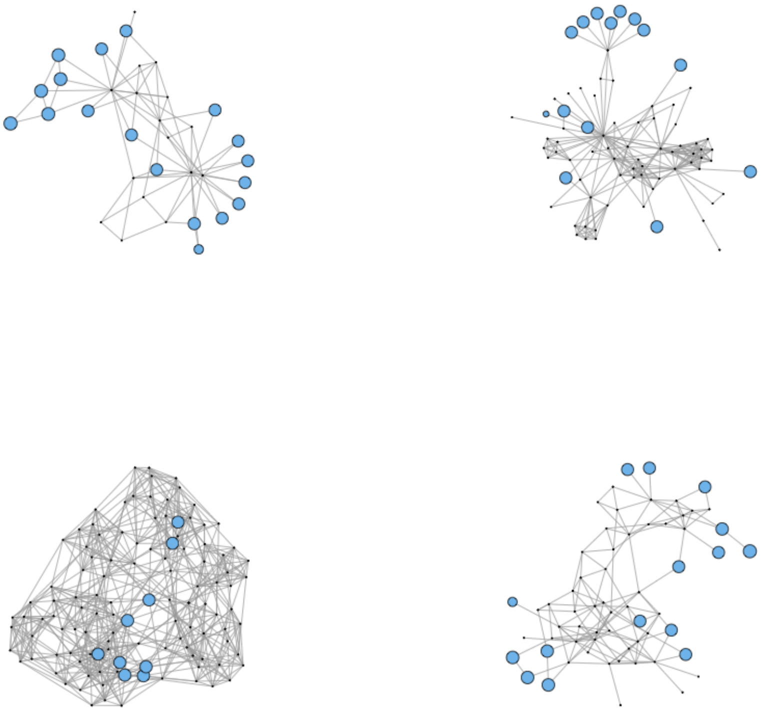


Figure 5.1: *Visualization of the solution of the Katz-Bonacich centrality minimization problem*

5.1.6 Conclusion and Perspectives

The main result of this section is the equivalence between a linear programming problem and the Katz-Bonacich centrality minimization problem. Once this equivalence is proved, we also describe an application to the control of gossip on OSN.

There is one major follow-up of this work. The extension is to study the Katz-Bonacich centrality minimization problem in an online context, in other words when the structure of the social network changes over time. The formulation of the problem could be the following: The evolution of the interaction occurs each unit time $t \in \mathbb{N}$. Let $E(t)$ be the interaction matrix at time t. For each t we assume that $E(t)$ a substochastic matrix. The index $p_i(t) \in [0,1]$ denotes the node i that the controller will perturbate at time t. The online version of the Katz-Bonacich centrality minimization problem is defined as follows:

Known parameters:. Set of nodes $\mathscr{I}$.
For each round $t = 1, 2, \ldots$

1. The controller removes nodes by using a control $\mathbf{p}(t) := (p_1(t), \ldots, p_I(t))$,
2. simultaneously, the adversary select a matrix $E(t)$ of interaction, $\rho(t)$ and $\alpha(t)$,
3. The controller observes $E(t)$, and receives

   $$\mathbf{1}_I^T (Id_I - \rho(t)E(t))^{-1} \mathbf{1}_I \alpha(t),$$

   as an instantaneous payoff vector.

In order to solve this problem we can use the theory of online convex optimization techniques [126].

5.2 A Continuous Optimization Approach for Maximizing Content Popularity in Online Social Networks

5.2.1 Introduction

In this section, we propose a novel approach for maximizing content popularity, where a source controls the flow of posted contents and the topics associated with each content. We develop an algorithm to find the optimal solution to this problem: In order to provide an answer to this challenge of popularity in OSNs, our methodology pursues the following steps.

- We start by understanding how content popularity behaves on OSNs. A vast amount of literature about this question exists (for instance [18; 13] and references therein). We follow the bottom line of the work proposed in [21] in which the authors consider that the popularity of content is highly related to its position in a News Feed. In Facebook for example, a *News Feed* is a feed where the user's friends and subscribed pages' posts are published. These posts are displayed in a top-down anti-chronological order from the newest to the oldest. More explanations about the News Feed dynamics and models are provided in [21].
- Then, we assume that the maximization of content popularity is similar to the oil producer problem [127]. This problem assumes that an oil producer would pump oil from a single well and after some time, when the return on the well decreases, he would abandon that source and start exploiting another one. In the case of OSNs, we assume that a well pumping is equivalent to a content staying in the first position of the News Feed. At some point, the popularity return will decrease and it is recommended to send a new message, meaning to exploit a new

source. Our model of popularity is expressed as a system of discrete differential equations.

- Finally, we study the maximization of content popularity. We assume that a company (or source) that tries to be popular will focus on two types of control: the number of posted contents and the topics of its messages. A continuous formulation of the popularity maximization problem is proposed and we prove its equivalence with a "generalized" fractional programme.

Organization of the section. In section 5.2.3, we propose a mathematical model of the popularity. The section 5.2.4 is about the Popularity maximization problem. First, we provide its definition, and then we prove its equivalence with a pseudoconcave optimization problem. Finally, an algorithm that converges to the optimal solution is provided.

5.2.2 Data experiment: importance of the first position

In this section, we explain how the popularity of a content is related to the position in the News Feed, specifically the duration the content spend in the first position in the News Feed. Our dataset is composed of messages posted on a Facebook page associated with a french political personality (Anne Hidalgo[2]). We have extracted 892 messages messages and 39083 associated comments using the Netvizz app [19]. The dataset was collected from June 15 2013 to December 01 2014. Messages and comments are associated with timestamps. The histogram in fig. 5.2 shows that more than 70% of the posted contents reached 80% of their total popularity while the content is in the first position in the News Feed. Therefore it makes sense to consider a popularity model, described in detail in the next section, by assuming that almost all the popularity of a content is obtained while the content is in the first position in the News Feed.

5.2.3 Popularity model

The main symbols used in this section are reported in Tab. 5.3. The level of popularity associated with a topic is defined as the total number of comments, given by OSN subscribers, of all contents related to this topic. We model the dynamic of the popularity of each topic through a discrete time equation. Let $c \in \mathscr{C} := \{1, \dots C\}$ be a topic index. The random state variable $X_c(t) \in \mathbb{N}_+$ is the total number of comments that gets all contents related to topic c, at a given time $t \in \mathbb{R}_+$. For each topic $c \in \mathscr{C}$ the arrival process of a type c content in the News Feed is a Poisson Point Process with intensity $0 < \lambda_c < \infty$. All arrival processes are independent. We also consider that there exists a flow, with intensity σ of posts which are not related to any topics. Let $t_m \in \mathbb{R}_+$ be the

[2] https://www.facebook.com/HidalgoAnne

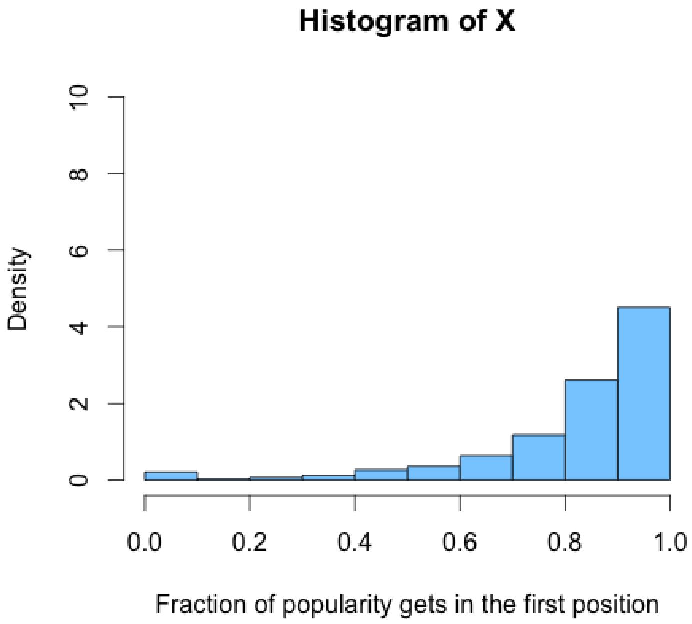


Figure 5.2: *Histogram of first position fraction of popularity*

random variable denoting the arrival instant of the m^{th} content, with $m \in \mathbb{N}^*_+$. As the arrival processes are independent Poisson Point Processes, the inter-arrival time denoted by Δ_m between consecutive messages arrived at t_m and t_{m-1}, follows an exponential distribution with rate $\Lambda := \sum_{d=1}^{C} \lambda_d + \sigma$. Therefore, the probability associated with the inter arrival time Δ_m of two consecutive messages is given by:

$$\forall m,\ \delta \in \mathbb{R}_+,\ P(\Delta_m < \delta) = \int_0^{\delta} \Lambda e^{-\Lambda t} dt. \tag{5.34}$$

A content can also be called a post on an OSNs. We define by $\eta(m) \in \mathscr{C}$ the topic of the m^{th} content. According to [82], for each message m, $\eta(m)$ is a random variable with the following distribution:

$$\forall m,\ \ P(\eta(m) = c) = \frac{\lambda_c}{\Lambda}. \tag{5.35}$$

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

$$X_c(t_m) = X_c(t_{m-1}) + 1_{\{\eta(m-1)=c\}} Y_c(\Delta_m), \tag{5.36}$$
$$X_c(t_1) = 0, \tag{5.37}$$

Table 5.3: Main notations used throughout the section 5.2

Symbol	*Meaning*
$\mathscr{C}$	set of topics
$X_c(t)$	cumulative popularity associated with topic c, at time t
λ_c	contents arrival rate concerning topic c
t_m	arrival instant of the m^{th} content
Δ_m	inter-arrival time between message m and $m-1$
$\eta(m)$	topic of the m^{th} content
$Y_c(\delta)$	cumulative popularity obtained by a content of type c that stays δ units of time on the first position
γ_c	linear cost associated with λ_c

where the random variable $Y_c(\delta) \in \mathbb{N}_+$, with $\delta \in \mathbb{R}_+$, describes the cumulative popularity obtained by a content of type c that stays δ units of time on the first position of the News Feed and $1_{\{\eta(m-1)=c\}} = 1$ if the $(m-1)^{th}$ content is of type c, it is equal to 0 otherwise. It has been recently demonstrated in [128] that the exponential model is a good approximation of the cumulative popularity. Thus we assume that for each c, the random variable $Y_c(\delta)$ is described by a non-homogeneous Poisson process of intensity $E[Y_c(\delta)] := \int_0^\delta e^{-\alpha_c \tau} d\tau$ where $\alpha_c \in \mathbb{R}^+$ is called the coefficient of innovation. A closed form expression of the expected cumulative popularity for each topic c is given in the following proposition.

Proposition 5.6. *The expected cumulative popularity when the m^{th} content is posted is given by:*

$$\forall c \in \mathscr{C}, t_m \in \mathbb{R}_+^*, \quad E[X_c(t_m)] = m\frac{\lambda_c}{\alpha_c}\left(\frac{1}{\Lambda+\sigma} - \frac{1}{\Lambda+\sigma+\alpha_c}\right). \tag{5.38}$$

5.2.4 Popularity optimization

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

$$\forall c \in \mathscr{C}, \quad X_c^\infty := \lim_{m\to\infty} \frac{1}{m} E[X_c(t_m)] \quad = \quad \frac{\lambda_c}{\alpha_c}\left(\frac{1}{\Lambda+\sigma} - \frac{1}{\Lambda+\alpha_c+\sigma}\right). \tag{5.39}$$

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

$$U(\lambda_1, \ldots, \lambda_C) := \sum_{c=1}^{C} \frac{\lambda_c}{\alpha_c} \left(\frac{1}{\Lambda + \sigma} - \frac{1}{\Lambda + \alpha_c + \sigma} \right) - \sum_{c=1}^{C} \gamma_c \lambda_c. \tag{5.40}$$

The *Popularity optimization problem* is the optimization problem defined below:

$$\max_{\lambda_c \in [0, \overline{\lambda}],\, \forall c} \quad U(\lambda_1, \ldots, \lambda_C) \tag{5.41}$$

$$(\lambda_1^{OA}, \ldots, \lambda_C^{OA}) = \text{argmax}_{\lambda_c \in [0, \overline{\lambda}],\, \forall c} U(\lambda_1, \ldots, \lambda_C) \tag{5.42}$$

The purpose of the rest of this section is to find an equivalent form of the Popularity optimization problem that can be solved efficiently. Indeed, this optimization problem is not a concave optimization problem as it can be observed in fig. 5.3. However, we

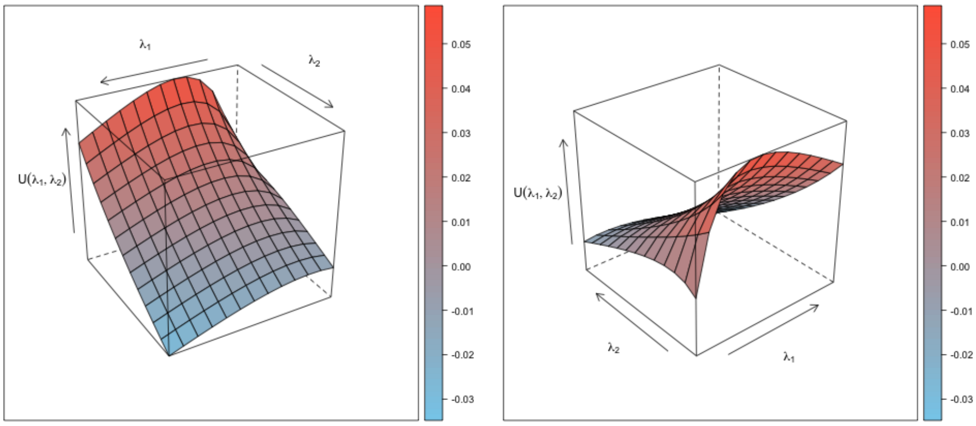


Figure 5.3: *Plot of* $U(\lambda_1, \lambda_2)$ *with* $\sigma = 1$, $\alpha_1 = 2$, $\alpha_2 = 5$ *and* $\gamma_1 = \gamma_2 = 0.01$.

can see that the global optimum point is such that $\lambda_2 = 0$. Thus, in this particular case, it is optimal for the source to send contents only about a single and unique topic. We next generalize and prove this observation. The next theorem proves that the optimal vector rate for the source is to send content only related to a single and unique topic. Therefore, the Popularity optimization problem defined in (5.41) can be reduced to a simpler optimization problem.

Proposition 5.7. *If for each c and c' and for all $x \in [0, C\overline{\lambda}]$*

$$\frac{1}{(x+\sigma)(x+\sigma+\alpha_c)} - \gamma_c \neq \frac{1}{(x+\sigma)(x+\sigma+\alpha_{c'})} - \gamma_{c'}, \tag{5.43}$$

then $\lambda_c^{OA} := (0,\ldots,0,\lambda_{c^*},0,\ldots,0)$ *where* c^* *and* λ_c^* *are solutions of the following optimization problem:*

$$\max_c \left[\max_{\lambda\in[0,\overline{\lambda}]} \left\{ \frac{\lambda}{\alpha_c}\left(\frac{1}{\lambda+\sigma} - \frac{1}{\lambda+\sigma+\alpha_c}\right)\right\} - \gamma_c\lambda \right]. \tag{5.44}$$

The hypothesis made in the previous proposition is satisfied when $\gamma_c = \gamma_{c'}$ and $\alpha_c \neq \alpha_{c'}$ for all c, c'. According to the previous proposition 5.7, the Popularity optimization problem is now equivalent to the following generalized fractional problem:

$$\max_c \left[\max_{\lambda\in[0,\overline{\lambda}]} \left\{ \frac{\lambda}{\alpha_c}\left(\frac{1}{\lambda+\sigma} - \frac{1}{\lambda+\sigma+\alpha_c}\right)\right\} - \gamma_c\lambda \right], \tag{5.45}$$

A generalized fractional problem [61] is described by the following optimization problem: $\max_{\mathbf{x}} \min_i \{\frac{f_i(\mathbf{x})}{g_i(\mathbf{x})}\}$. We define for each topic $c \in \mathscr{C}$ the function $U_c(\lambda) = \frac{\lambda}{\alpha_c}\left(\frac{1}{\lambda+\sigma} - \frac{1}{\lambda+\sigma+\alpha_c}\right) - \gamma_c\lambda$. In the next proposition we prove the pseudo-concavity of the function $U_c(\lambda)$ for any topic $c \in \mathscr{C}$.

Proposition 5.8. *For each* $c \in \mathscr{C}$*, a projected gradient algorithm, where the update is proposed in* (5.46)*, will converge to the optimal solution of* $U_c(\lambda)$.

We have formulated the Popularity optimization problem as a generalized fractional problem and we prove that the sub-optimization problem is pseudo-concave. Considering such gradient algorithm, we finally propose a global algorithm that first looks for the optimal rate for each topic c and then determines, by comparing all the outputs of the C gradient algorithms, on which topic the source has to send his contents. We describe the updating rule in the equation below.

Algorithm

1. *Initialization:* For each $c \in \{1,\ldots,C\}$, $\lambda_c(1) \in [0,\overline{\lambda}]$ randomly. The horizon T of gradient algorithms is chosen.

2. *Updating rule:* For each $c \in \{1,\ldots,C\}$ and for each round $t = 1,2,\ldots,T$, the source updates the vector $\lambda_c(t)$:

$$\lambda_c(t+1) = \Gamma_2\left[\lambda_c(t) + \alpha\frac{\partial U_c}{\partial \lambda_c}(\lambda_c(t))\right], \tag{5.46}$$

where $\Gamma_2[\cdot]$ is the projection over interval $[0,\overline{\lambda}]$ and $\alpha(t)$ is chosen according to the rule describes in [60].

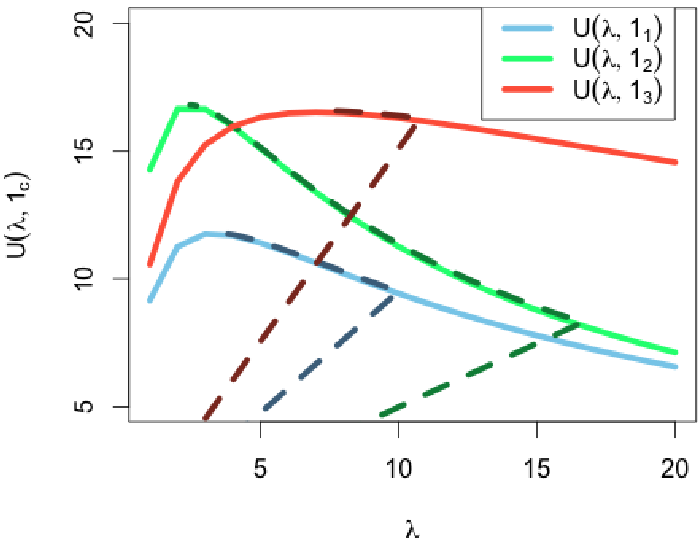


Figure 5.4: *Numerical illustration of the algorithm.*

3. *Solve:*

$$c^* = \arg \max_{c \in \{1,\dots,C\}} \{U_c(\lambda_c(T))\}.$$

We illustrate our algorithmic solution on a numerical example where the parameters are $C = 3$, $(\alpha_1, \alpha_2, \alpha_3) = (10, 5, 50)$. In fig. 5.4, each curve draws the utility in the case when the source decides to send content related to a unique topic. Each dashed line shows the iteration of the gradient algorithm for each type c. We observe the convergence of the gradient in each case as stated previously. The last iteration of the algorithm is not shown in fig. 5.4 , but it is easy to deduce that the source has to create contents about the second topic, in order to become popular.

5.2.5 Discussion and Conclusion

In this section, we have defined a popularity optimization problem in OSNs. The model is based on the observation that most of the popularity of a content is obtained during the time a content is in the first position in a News Feed. We define an optimization problem where a source decides the rate of his content and also the topics. An equivalent fractional problem is obtained and an efficient algorithm that converges to the optimal solution is proposed.

APPENDIX

Proof of proposition 5.3:
If

$$\underline{p}_i \leq \frac{k_i}{1+\sum_j \rho e_{ij} k_j} \leq \overline{p}_i, \tag{5.47}$$

then $p_i^* \in [\underline{p}_i, \overline{p}_i]$. This is the reason why p_i^* exists. Moreover, for each i

$$\begin{aligned} x_i^* &= p_j^*(1+\sum_j \rho e_{ij} x_j^*) &(5.48)\\ x_i^* &= \frac{k_i}{1+\sum_j \rho e_{ij} k_j}(1+\sum_j \rho e_{ij} x_j^*). &(5.49)\end{aligned}$$

This linear system admit a unique solution. And it is easy to check that the unique solution is $x_i^* = k_i$:

$$\begin{aligned} k_i &= \frac{k_i}{1+\sum_j \rho e_{ij} k_j}(1+\sum_j \rho e_{ij} k_j) &(5.50)\\ k_i &= k_i. &(5.51)\end{aligned}$$

Proof of proposition 5.4:
According to the proposition 5.3, it exists $\mathbf{p}$ such that $x_i^* = k_i$ if for each i:

$$\underline{p}_i \leq \frac{k_i}{1+\sum_j \rho e_{ij} k_j} \leq \overline{p}_i. \tag{5.52}$$

Proof of proposition 5.6:
For each topic $c \in \mathscr{C}$ and for each time t_m, we have:

$$\begin{aligned} E[X_c(t_m)] &= E[X_c(t_{m-1})] + E[1_{\eta(m)=c} Y_c(\Delta_m)],\\ &= E[X_c(t_{m-1})] + \frac{\lambda_c}{\Lambda+\sigma}\int_0^\infty (\Lambda+\sigma)e^{-(\Lambda+\sigma)\tau}\int_0^\tau e^{-\alpha_c t'}dt'd\tau,\\ &= E[X_c(t_{m-1})] + \frac{\lambda_c}{\alpha_c}\left(\frac{1}{\Lambda+\sigma} - \frac{1}{\Lambda+\sigma+\alpha_c}\right),\end{aligned}$$

as the inter-arrival time between two consecutive contents follows an exponential distribution with rate $\Lambda+\sigma$ and this duration is independent of the type of content. Therefore, we repeat this evaluation of the expectation and we get:

$$\forall c \in \mathscr{C}, t_m \in \mathbb{R}_+^*, \quad E[X_c(t_m)] = m\frac{\lambda_c}{\alpha_c}\left(\frac{1}{\Lambda+\sigma} - \frac{1}{\Lambda+\sigma+\alpha_c}\right).$$

Proof of proposition 5.7:
We assume that there exits $(c,c') \in \mathscr{C}^2$ such that $\lambda_c^{OA} > 0$ and $\lambda_{c'}^{OA} > 0$. This assumption implies that

$$\begin{aligned} \frac{\partial U}{\partial \lambda_c}(\lambda_1^{OA},\ldots,\lambda_C^{OA}) &= 0 && (5.53)\\ \frac{\partial U}{\partial \lambda_{c'}}(\lambda_1^{OA},\ldots,\lambda_C^{OA}) &= 0 && (5.54)\end{aligned}$$

We compute the partial derivative which is equal to

$$\begin{aligned} \frac{\partial U}{\partial \lambda_c}(\lambda_1^{OA},\ldots,\lambda_C^{OA}) &= \frac{1}{(\sum_d \lambda_d^{OA}+\sigma)(\sum_d \lambda_d^{OA}+\sigma+\alpha_c)} \\ &\quad -\sum_{d'} \frac{\lambda_{d'}}{(\sum_d \lambda_d^{OA}+\sigma)^2(\sum_d \lambda_d^s sOA+\sigma+\alpha_{d'})} \\ &\quad -\sum_{d'} \frac{\lambda_{d'}}{(\sum_d \lambda_d^{OA}+\sigma)(\sum_d \lambda_d^{OA}+\sigma+\alpha_{d'})^2} - \gamma_c. \end{aligned}$$

We deduce from the previous equations and from the stationary condition (5.53), (5.54) that, for each c and c'

$$\frac{1}{(\sum_d \lambda_d^{OA}+\sigma)(\sum_d \lambda_d^{OA}+\sigma+\alpha_c)} - \gamma_c = \frac{1}{(\sum_d \lambda_d^{OA}+\sigma)(\sum_d \lambda_d^{OA}+\sigma+\alpha_{c'})} - \gamma_{c'}. \qquad (5.55)$$

These equations contradict the assumption (5.43) made in the theorem. Thus there is only stationary point of the form : $\lambda_c^{OA} = \mathbf{1_{c=c'}}\lambda^{OA}$ for each $c' \in \mathscr{C}$. The rest of the theorem follows from this conclusion.

Proof of proposition 5.8:
We first prove that, for each c, $U_c(\cdot)$ is pseudo concave. Then the projected gradient algorithm described in [60] converges to the global optimum. It remains to show the pseudo concavity of $U_c(\dot{})$ in λ. Consider $c \in \mathscr{C}$ and $\lambda \in \{x \in [0,\overline{\lambda}] \mid U(x,\mathbf{p}) \geq 0\}$. $\lambda_c^{OA} \in \{x \in [0,\overline{\lambda}] \mid U(x,\mathbf{p}) \geq 0\}$ because if λ_c^{OA} is such that $U(\lambda_c^{OA},\mathbf{p}) < 0$, λ_c^{OA} can be equal to 0 to get a better payoff. We can rewrite $U_c(\lambda)$ as a fractional problem:

$$\begin{aligned} U_c(\lambda) &= \frac{1}{\alpha_c}\lambda\left(\frac{1}{\lambda+\sigma} - \frac{1}{\lambda+\sigma+\alpha_c}\right) - \gamma_c\lambda && (5.56)\\ &= \frac{\lambda}{(\lambda+\sigma)(\lambda+\sigma+\alpha_c)} - \gamma_c\lambda && (5.57)\\ &= \frac{\lambda - \gamma_c\lambda(\lambda+\sigma)(\lambda+\sigma+\alpha_c)}{(\lambda+\sigma)(\lambda+\sigma+\alpha_c)}. && (5.58)\end{aligned}$$

Consider $h_c(\lambda) := \lambda - \gamma_c\lambda(\lambda+\sigma)(\lambda+\sigma+\alpha_c)$ and $g(\lambda) := (\lambda+\sigma)(\lambda+\sigma+\alpha_c)$. According to theorem 5.17 p.162 of [61] if $h_c(\lambda)$ is nonnegative, concave and differentiable and $g(\lambda)$ is positive, convex and differentiable, then $\dfrac{h_c(\lambda)}{g(\lambda)}$ is pseudoconcave. It is easy to note $g(\lambda)$ is a convex function differentiable and positive. As a polynomial, $h_c(\lambda)$ is differentiable, $h_c(\lambda)$ is positive because of $\lambda \in \{x \in [0,\overline{\lambda}] \mid U(x,\mathbf{p}) \geq 0\}$. About the concavity of $h_c(\lambda)$, the second derivative of $h_c(\lambda)$ is equal to:

$$\frac{\partial^2 h_c}{\partial \lambda^2}(\lambda) = -2\gamma_c\lambda - 2\gamma_c(\lambda+\sigma+\alpha_c) - 2\gamma_c(\lambda+\sigma), \tag{5.59}$$

and we deduce that $\dfrac{\partial^2 f}{\partial \lambda^2}(\lambda) \leq 0$ for all $\lambda \in \{x \in [0,\overline{\lambda}] \mid U(x,\mathbf{p}) \geq 0\}$. This establishes the peusoconcavity of U_c and concludes the proof.

Chapter 6

Conclusions and Perspectives

Contents

6.1 Conclusions

In this PhD thesis, we model complex behaviour of OSNs users by using Game Theory and stochastic processes. These approaches allow us to discover paradoxes and inefficiency phenomena that cannot be observed by using data analysis. We believe that, with the theoretical results obtained during this PhD thesis, it is possible to help the different actors of OSNs (sources, readers and OSN owner) to use it more effectively. In the rest of the section we summarize the result obtain during this PhD thesis.

In chapter 2 we have defined a routing game where sources try to maximize their visibility in News Feeds. They can decide how many messages they have to post, the topic of their messages and in which News Feed to post. We prove that the Nash equilibrium is unique and we provide a mathematical closed form of it. We develop a decentralized algorithm that converges to the NE. Finally, we model the situation where there are multiple news feeds and we compute the symmetric Nash equilibrium in this case. The study of the Nash equilibrium leads us to the following observations. Firstly, when a source maximizes his visibility, it is optimal for him to send messages about a unique topic, which is negative in terms of content diversity in OSNs. Secondly the flow sent by the sources at the unique Nash equilibrium is concave quadratic in the number of players. Therefore, when the number of players increases, the equilibrium flow will increase up to a threshold, then it will decrease. Thirdly, we extend our work to multiple news feeds, where the sharing phenomenon is modeled. In chapter 2 the model proposed is not a dynamical one, which is a strong assumption. In chapter 3 we relax this hypothesis.

In chapter 3 the posting behaviour of sources in OSNs is modeled by using stochastic approximation. The sources have influence one over another. We develop a content active filtering in order to increase content diversity. Sufficient conditions, over the externalities matrix, are given in order to ensure the convergence of the posting behaviour. Moreover, we develop an equivalent game theoretical model. Finally, all the theoretical results are illustrated through simulations and a data set extracted from a real OSN is used to assess our results. This chapter is more practical because once we extract the influence matrix we can control the posting behaviour by using the CAF. In chapter 2 and 3 we assume that the control of sources is the number of posted messages. We introduce a new control, the payment strategy.

In chapter 4, we introduce a differential game to model advertisement in online content diffusion, where sources compete to make contents popular. We demonstrate that in discounted infinite horizon case the closed loop best Nash equilibrium is bang-bang type in the view count. Moreover, we found that dual counterparts exist in the fully asymmetric case for small discounts. The practical implications of this chapter is the following: From the sources point of view, the optimal strategy that maximize the Nash equilibrium is bang-bang, and so by estimating parameters of the model it is easy to implement in a real scenario.

Chapter 5 is dedicated to two optimization problems found in OSNs. The first one is the minimization of gossip propagation, which we simulate on real network situations. We prove that it is equivalent to a convex programming problem and therefore, simple to implement. The second optimization problem of this chapter is concerned with popularity optimization in OSN. We propose a more realistic model than the one developed in the previous chapter, because we are not in a game theoretical framework. This problem is demonstrated to be equivalent to a "generalized" fractional program and we provide an algorithm that converges to the optimal solution.

6.2 Perspectives

6.2.1 Welfare in OSNs: a game theory approach

According to the American customer satisfaction index[1], OSNs "are among the worst-performing companies"[2] in term of customer satisfaction. This is the reason why understanding the needs of users in these particular platforms is becoming a major issue. In chapter 2 we have studied the behaviour and needs of publishers and we have provided few insights. However, several questions remain unanswered: How to model reader's satisfaction? Which paradoxes can appear when the topology changes over time? Is it beneficial for a reader to reveal its interests?

These challenges can lead to several possible answers. We present some suggestions for future research:
- *Readers incorporation* : We study the incorporation of readers in the model proposed in chapter 2. There is very little literature about game theoretical model with both sources and readers in OSNs. To the best of our knowledge, there is only one paper proposing a model [15] where both sources and readers are introduced. Once a model is proposed, indexes that measure users' and sources' satisfaction need to be developed because the existing ones (for instance, the price of anarchy), are not well adapted. Moreover, we expect that these indexes will reveal some paradoxes, similar to the Braess paradox, which can help the OSNs to improve the design of their platform.
- *Temporal network* : Usually, in a game theoretical framework, a static description (not a dynamical one), of the relationship between users in OSNs is proposed. However,

[1] http://www.theacsi.org/
[2] http://goo.gl/zAvSXl

we expect that the dynamical topology can have intriguing impacts over the Nash equilibrium. There are many scenarios that can be modeled where the influence between sources changes over time. For example, when a source posts many messages about a particular topic, readers are more likely to unsubscribe to this source. Consequently, the influence of the source will drastically decrease. In this possible framework, it is of interest to understand how sources will behave, due to the risk of losing connections, and which possible new paradoxes will emerge.
- *Incomplete information games* : In OSNs one fact is clear: Except for the OSNs owner himself, a random user does not have full information about other users nor about how they receive contents. Moreover, because there are several users that try to maximize their own utilities, it is not clear that for particular users, having more information is a better choice. Indeed, in a seminal paper [129] the authors prove that in a game theoretical framework, having too much information can be harmful for a player. So it is of key interest to adapt the framework of incomplete information game to the OSNs and observe whether or not the value of information is negative or positive for users.

6.2.2 A differential game approach to model popularity competition in OSNs

ONS platforms have allowed their subscribers to use them for content large scale dissemination. However, ONSs gives preferential treatment to the publishers who pay for advertisement. When a user pays for the improvement of his popularity, his strategy is naturally dynamic. For instance, an intuitive strategy would be to invest as much as possible in one content until his popularity return becomes low. However, is this simple strategy still optimal when adopted by many users ? What is the optimal strategy when a publisher wants to invest in several contents at the same time? Are the popularity of contents correlated between themselves? If so, what are the benefits resulting from these correlations?

We propose future lines of research that can lead to interesting answers to the previous questions:
- *Threshold strategy* : We would like to see whether or not a threshold strategy (or bang-bang strategy) is a Nash equilibrium. This implies some theoretical challenges because the necessary and sufficient conditions of optimality (Pontryaguin conditions and Hamilton Jacobi Bellman conditions) are not well adapted in this context. Assuming that we are able to prove the optimality of a bang bang strategy for both open loop and closed loop cases, it could be interesting to compare these two strategy. Indeed, the amount of information required for the computation of both strategies is not the same and it is not clear which one would be the best in this case.
- *Influence effect* : Recent papers that model popularity competition by using differential games do not take into account complex interaction between contents. However, we

think that a reasonable model would be the one proposed in [130]. This model is based on an ode system that belongs to the susceptible-infected-susceptible epidemiological models class. The first step in this study would be to validate the model over real data. The second step is the computation of a Nash equilibrium and to observe the impact of the influence between contents over it. It could help the OSNs platforms to design a better pricing model for advertising.

6.2.3 Optimal control of information diffusion

In times of crisis, it is of main interest for the government to control the information that propagate on OSNs. One way of doing so is to find the key node, which needs to be removed from this OSNs by the government so that it will minimize the diffusion of misinformation. However, how to do so when the network changes over time and we do not observe these change? Can we design a more realistic model that can lead to better performance?

- *Stochastic approximation scheme* : Assuming that the government wants to know which node to remove in order to minimize the propagation of gossip without knowing the topology. We first use the model proposed in chapter 5 to model the propagation of misinformation. If we are able to prove that removing the key node is equivalent to a convex optimization problem, then we can use a stochastic gradient approximation scheme [46] to solve the minimization problem without any knowledge of the diffusion topology. Finally, we can consider that the topology evolves over time and the online convex optimization theory [131] seems to be adapted to tackle this problem.
- *Consensus model over signed graph* : The model proposed in chapter 5 can be seen as the basic model of the misinformation propagation which is equivalent to the Degroot consensus model [47]. A simple extension could be to consider that users can trust or distrust their neighbours and so this model would be consensus over a signed graph. Now it could be interesting to find whether or not the problem of finding the key node has an equivalent convex form. Then we can try to solve this extension when the graph is not known.

List of Figures

List of Tables

Bibliography

[1] Nicole Ellison and Danna Boyd. Social network sites: Definition, history, and scholarship. *Journal of Computer-Mediated Communication*, 13(1):210–230, 2007.

[2] Caroline Haythornthwaite. Social network methods and measures for examining e-learning. *Social Networks*, 2005:1–22, 2005.

[3] Julia Heidemann, Mathias Klier, and Florian Probst. Online social networks: A survey of a global phenomenon. *Computer Networks*, 56(18):3866–3878, 2012.

[4] Carolyn Heller Baird and Gautam Parasnis. From social media to social customer relationship management. *Strategy & Leadership*, 39(5):30–37, 2011.

[5] Kate Starbird, Jim Maddock, Mania Orand, Peg Achterman, and Robert M Mason. Rumors, false flags, and digital vigilantes: Misinformation on twitter after the 2013 boston marathon bombing. *iConference 2014 Proceedings*, 2014.

[6] 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.

[7] Hanna Krasnova, Thomas Hildebrand, and Oliver Guenther. Investigating the value of privacy in online social networks: conjoint analysis. *ICIS 2009 Proceedings*, page 173, 2009.

[8] Eytan Bakshy, Itamar Rosenn, Cameron Marlow, and Lada Adamic. The role of social networks in information diffusion. In *Proceedings of the 21st international conference on World Wide Web*, pages 519–528. ACM, 2012.

[9] Adrien Guille, Hakim Hacid, Cécile Favre, and Djamel A Zighed. Information diffusion in online social networks: A survey. *ACM SIGMOD Record*, 42(2):17–28, 2013.

[10] Kristina Lerman and Rumi Ghosh. Information contagion: An empirical study of the spread of news on digg and twitter social networks. *ICWSM*, 10:90–97, 2010.

[11] Fei Xiong, Yun Liu, Zhen-jiang Zhang, Jiang Zhu, and Ying Zhang. An information diffusion model based on retweeting mechanism for online social media. *Physics Letters A*, 376(30):2103–2108, 2012.

[12] Seth A Myers and Jure Leskovec. Clash of the contagions: Cooperation and competition in information diffusion. In *Data Mining (ICDM), 2012 IEEE 12th International Conference on*, pages 539–548. IEEE, 2012.

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

[14] Henry Jenkins. Confronting the challenges of participatory culture: Media education for the 21st century. an occasional paper on digital media and learning. *John D. and Catherine T. MacArthur Foundation*, 2006.

[15] Bo Jiang, Nidhi Hegde, Laurent Massoulie, and Don Towsley. How to optimally allocate your budget of attention in social networks. In *Proc. of INFOCOM*, Turin, Italy, 17-19 April 2013.

[16] Taina Bucher. Want to be on the top? algorithmic power and the threat of invisibility on facebook. *New Media & Society*, 14(7):1164–1180, 2012.

[17] Solomon Messing and Sean J Westwood. How social media introduces biases in selecting and processing news content. Technical report, Working paper, available at http://www. stanford. edu/seanjw/papers/SMH. pdf, 2012.

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

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

[20] Oliver J Rutz and Michael Trusov. Zooming in on paid search ads-a consumer-level model calibrated on aggregated data. *Marketing Science*, 30(5):789–800, 2011.

[21] Eitan Altman, Parmod Kumar, Srinivasan Venkatramanan, and Anurag Kumar. Competition over timeline in social networks. In *Advances in Social Networks Analysis and Mining (ASONAM), 2013 IEEE/ACM International Conference on*, pages 1352–1357. IEEE, 2013.

[22] Roja Bandari, Sitaram Asur, and Bernardo A Huberman. The pulse of news in social media: Forecasting popularity. In *ICWSM*, pages 26–33, 2012.

[23] Panpan Xu, Yingcai Wu, Enxun Wei, Tai-Quan Peng, Shixia Liu, Jonathan JH Zhu, and Huamin Qu. Visual analysis of topic competition on social media. *Visualization and Computer Graphics, IEEE Transactions on*, 19(12):2012–2021, 2013.

[24] Anjana Susarla, Jeong-Ha Oh, and Yong Tan. Social networks and the diffusion of user-generated content: Evidence from youtube. *Information Systems Research*, 23(1):23–41, 2012.

[25] Benjamin Edelman and Michael Ostrovsky. Strategic bidder behavior in sponsored search auctions. *Decision support systems*, 43(1):192–198, 2007.

[26] Nemanja Spasojevic, Zhisheng Li, Adithya Rao, and Prantik Bhattacharyya. When-to-post on social networks. In *Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pages 2127–2136. ACM, 2015.

[27] Eitan Altman and Nahum Shimkin. The ordered timeline game: Strategic posting times over a temporally ordered shared medium. *Dynamic Games and Applications*, pages 1–27, 2015.

[28] Justin Cheng, Lada Adamic, P Alex Dow, Jon Michael Kleinberg, and Jure Leskovec. Can cascades be predicted? In *Proceedings of the 23rd international conference on World wide web*, pages 925–936. ACM, 2014.

[29] 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.

[30] Ananya Sen and Pinar Yildirim. Clicks and editorial decisions: How does popularity shape online news coverage? *Available at SSRN 2619440*, 2015.

[31] John F Nash. Equilibrium points in n-person games. *Proc. Nat. Acad. Sci. USA*, 36(1):48–49, 1950.

[32] Refael Hassin and Moshe Haviv. *To queue or not to queue: Equilibrium behavior in queueing systems*, volume 59. Springer Science & Business Media, 2003.

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

[34] Erin E Buckels, Paul D Trapnell, and Delroy L Paulhus. Trolls just want to have fun. *Personality and individual Differences*, 67:97–102, 2014.

[35] Eva Buechel and Jonah Berger. Facebook therapy? why do people share self-relevant content online? In *presentation at Association for Consumer Research Conference, Vancouver, BC*, 2012.

[36] 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.

[37] Devavrat Shah and Tauhid Zaman. Rumors in a network: Who's the culprit? *Information Theory, IEEE Transactions on*, 57(8):5163–5181, 2011.

[38] Kalpesh Gandhi, Rahul Panditrao, and Vibha B Lahane. A survey on osn message filtering. *International Journal of Computer Applications*, 113(17), 2015.

[39] Moira Burke, Cameron Marlow, and Thomas Lento. Feed me: motivating newcomer contribution in social network sites. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 945–954. ACM, 2009.

[40] Hongyu Gao, Yan Chen, Kathy Lee, Diana Palsetia, and Alok N Choudhary. Towards online spam filtering in social networks. In *NDSS*, 2012.

[41] Huajing Li, Yuan Tian, Wang-Chien Lee, C Lee Giles, and Meng-Chang Chen. Personalized feed recommendation service for social networks. In *Social Computing (SocialCom), 2010 IEEE Second International Conference on*, pages 96–103. IEEE, 2010.

[42] Cheng Wan. *Contributions à la théorie des jeux d'évolution et de congestion*. PhD thesis, Université Pierre et Marie Curie-Paris VI, 2012.

[43] T. Basar and G. J. Olsder. *Dynamic Noncooperative Game Theory*. SIAM, Philadelphia, PA, 2nd edition, 1999.

[44] M. Bardi and I. Capuzzo Dolcetta. *Optimal Control and Viscosity Solutions of Hamilton-Jacobi-Bellman Equations*. Birkhauser, 2008.

[45] Francois Baccelli and Pierre Brémaud. *Elements of queueing theory: Palm Martingale calculus and stochastic recurrences*, volume 26. Springer Science & Business Media, 2013.

[46] Vivek S Borkar. Stochastic approximation. *Cambridge Books*, 2008.

[47] Morris H DeGroot. Reaching a consensus. *Journal of the American Statistical Association*, 69(345):118–121, 1974.

[48] Matthew O Jackson. *Social and economic networks*, volume 3. Princeton University Press Princeton, 2008.

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

[50] Alexandre Reiffers-Masson, Yezekael Hayel, and Eitan Altman. Game theory approach for modeling competition over visibility on social networks. In *Communication Systems and Networks (COMSNETS), 2014 Sixth International Conference on*, pages 1–6. IEEE, 2014.

[51] Gustav Feichtinger, Richard F Hartl, and Suresh P Sethi. Dynamic optimal control models in advertising: recent developments. *Management Science*, 40(2):195–226, 1994.

[52] Steffen Jørgensen and Georges Zaccour. *Differential games in marketing*, volume 15. Springer Science & Business Media, 2004.

[53] 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.

[54] Giovanni Neglia, Xiuhui Ye, Maksym Gabielkov, and Arnaud Legout. How to network in online social networks. In *Computer Communications Workshops (INFOCOM WKSHPS), 2014 IEEE Conference on*, pages 819–824. IEEE, 2014.

[55] Allan Borodin, Yuval Filmus, and Joel Oren. Threshold models for competitive influence in social networks. In *Internet and network economics*, pages 539–550. Springer, 2010.

[56] Mangesh Gupte, MohammadTaghi Hajiaghayi, Lu Han, Liviu Iftode, Pravin Shankar, and Raluca M Ursu. News posting by strategic users in a social network. In *Internet and Network Economics*, pages 632–639. Springer, 2009.

[57] Thanh Nguyen and Milan Vojnovic. Weighted proportional allocation. In *Proceedings of the ACM SIGMETRICS joint international conference on Measurement and modeling of computer systems*, pages 173–184. ACM, 2011.

[58] Koji Okuguchi and Takeshi Yamazaki. Global stability of unique nash equilibrium in cournot oligopoly and rent-seeking game. *Journal of Economic Dynamics and Control*, 32(4):1204–1211, 2008.

[59] Burkhard C Schipper. Submodularity and the evolution of walrasian behavior. *International Journal of Game Theory*, 32(4):471–477, 2004.

[60] Changyu Wang and Naihua Xiu. Convergence of the gradient projection method for generalized convex minimization. *Computational Optimization and Applications*, 16(2):111–120, 2000.

[61] Mordecai Avriel, Walter E Diewert, Siegfried Schaible, and Israel Zang. *Generalized concavity*, volume 63. SIAM, 1988.

[62] Eitan Altman, Yezekael Hayel, and Hisao Kameda. Evolutionary dynamics and potential games in non-cooperative routing. In *Modeling and Optimization in Mobile, Ad Hoc and Wireless Networks and Workshops, 2007. WiOpt 2007. 5th International Symposium on*, pages 1–5. IEEE, 2007.

[63] John N Tsitsiklis, Dimitri P Bertsekas, and Michael Athans. Distributed asynchronous deterministic and stochastic gradient optimization algorithms. *IEEE transactions on automatic control*, 31(9):803–812, 1986.

[64] Noam Nisan, Tim Roughgarden, Eva Tardos, and Vijay V Vazirani. *Algorithmic game theory*, volume 1. Cambridge University Press Cambridge, 2007.

[65] J Ben Rosen. Existence and uniqueness of equilibrium points for concave n-person games. *Econometrica: Journal of the Econometric Society*, pages 520–534, 1965.

[66] Dov Monderer and Lloyd S Shapley. Potential games. *Games and economic behavior*, 14(1):124–143, 1996.

[67] Eitan Altman and Zwi Altman. S-modular games and power control in wireless networks. *Automatic Control, IEEE Transactions on*, 48(5):839–842, 2003.

[68] 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.

[69] 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. *Academy of Science and Engineering (ASE), USA, ASE 2014*, 2014.

[70] 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.

[71] 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.

[72] 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.

[73] 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.

[74] 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, and Marielena Lara. Health information on the internet: accessibility, quality, and readability in english and spanish. *Jama*, 285(20):2612–2621, 2001.

[75] 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.

[76] 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.

[77] 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.

[78] Kostas Bimpikis, Asuman Ozdaglar, and Ercan Yildiz. Competing over networks. *working paper*, 2013.

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

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

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

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

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

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

[85] 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.

[86] Drew Fudenberg and David K Levine. *The theory of learning in games*, volume 2. MIT press, 1998.

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

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

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

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

[91] Meeyoung Cha, Haewoon Kwak, Pablo Rodriguez, Yong-Yeol Ahn, and Sue Moon. I tube, you tube, everybody tubes: analyzing the world's largest user generated content video system. In *Proc. of ACM IMC*, San Diego, California, USA, October 24-26 2007.

[92] R. Crane and D. Sornette. Viral, quality, and junk videos on YouTube: Separating content from noise in an information-rich environment. In *Proc. of AAAI symposium on Social Information Processing*, Menlo Park, California, CA, March 26-28 2008.

[93] Phillipa Gill, Martin Arlitt, Zongpeng Li, and Anirban Mahanti. YouTube traffic characterization: A view from the edge. In *Proc. of ACM IMC*, San Diego, California, USA, October 24-26 2007.

[94] J. Ratkiewicz, F. Menczer, S. Fortunato, A. Flammini, and A. Vespignani. Traffic in Social Media II: Modeling Bursty popularity. In *Proc. of IEEE SocialCom*, Minneapolis, August 20-22 2010.

[95] G. Chatzopoulou, Cheng Sheng, and Michaelis Faloutsos. A First Step Towards Understanding Popularity in YouTube. In *Proc. of IEEE INFOCOM*, pages 1 –6, San Diego, California, USA, March 15-19 2010.

[96] Meeyoung Cha, Haewoon Kwak, P. Rodriguez, Yong-Yeol Ahn, and Sue Moon. Analyzing the video popularity characteristics of large-scale user generated content systems. *IEEE/ACM Transactions on Networking*, 17(5):1357 – 1370, 2009.

[97] J. Davidson, B. Liebald, J. Liu, P. Nandy, and T. Van Vleet. The YouTube video recommendation system. In *Proc. of ACM RecSys*, pages 293–296, Sept. 26-30 2010.

[98] Noga Alon, Michal Feldman, Ariel D. Procaccia, and Moshe Tennenholtz. A note on competitive diffusion through social networks. *Elsevier Information Processing Letters*, 2010.

[99] V. Tzoumas, C. Amanatidis, and E. Markaki. A game-theoretic analysis of a competitive diffusion process over social networks. In *Proc. of 8th Workshop on Internet and Network Economics (WINE)*, 2012.

[100] Eitan Altman, Francesco De Pellegrini, Rachid El Azouzi, Daniele Miorandi, and Tania Jiménez. Emergence of Equilibria from Individual Strategies in Online Content Diffusion. In *Proc. of IEEE NetSciCom*, Turin, Italy, April 19 2013.

[101] M. Benaïm and J.-Y. Le Boudec. A class of mean field interaction models for computer and communication systems. *Performance Evaluation*, 65(11-12):823–838, 2008.

[102] Eitan Altman, Lucile Sassatelli, and Francesco De Pellegrini. Dynamic control of coding for progressive packet arrivals in dtns. *IEEE Transactions on Wireless Communications*, 12(2):725–735, 2013.

[103] Coralio Ballester, Yves Zenou, and Antoni Calvó-Armengol. Delinquent networks. *Journal of the European Economic Association*, 8(1):34–61, 2010.

[104] Olivier Fercoq, Marianne Akian, Mustapha Bouhtou, and Stéphane Gaubert. Ergodic control and polyhedral approaches to pagerank optimization. *Automatic Control, IEEE Transactions on*, 58(1):134–148, 2013.

[105] Vincenzo Nicosia, Regino Criado, Miguel Romance, Giovanni Russo, and Vito Latora. Controlling centrality in complex networks. *Scientific reports*, 2, 2012.

[106] Ernesto Estrada and Juan A Rodriguez-Velazquez. Subgraph centrality in complex networks. *Physical Review E*, 71(5):056103, 2005.

[107] Francis Bloch and Nicolas Quérou. Pricing with local network externalities. Technical report, mimeo, 2009.

[108] Abhijit Banerjee, Arun G Chandrasekhar, Esther Duflo, and Matthew O Jackson. The diffusion of microfinance. *Science*, 341(6144):1236498, 2013.

[109] Abhijit Banerjee, Arun G Chandrasekhar, Esther Duflo, and Matthew O Jackson. Gossip: Identifying central individuals in a social network. Technical report, National Bureau of Economic Research, 2014.

[110] Umed Temurshoev. Who's who in networks-wanted: The key group. *Available at SSRN 1285752*, 2008.

[111] James R Bunch and Donald J Rose. *Sparse matrix computations*. Academic Press, 2014.

[112] Coralio Ballester, Antoni Calvó-Armengol, and Yves Zenou. Who's who in crime network. wanted the key player. Technical report, IUI Working Paper, 2004.

[113] Yang-Yu Liu, Jean-Jacques Slotine, and Albert-László Barabási. Control centrality and hierarchical structure in complex networks. *Plos one*, 7(9):e44459, 2012.

[114] Yujian Pan and Xiang Li. Power of individuals–controlling centrality of temporal networks. *arXiv preprint arXiv:1401.3056*, 2014.

[115] Rudolf Emil Kalman. Contributions to the theory of optimal control. *Bol. Soc. Mat. Mexicana*, 5(2):102–119, 1960.

[116] Shigeyuki Hosoe and Kojyun Matsumoto. On the irreducibility condition in the structural controllability theorem. *Automatic Control, IEEE Transactions on*, 24(6):963–966, 1979.

[117] Leo Katz. A new status index derived from sociometric analysis. *Psychometrika*, 18(1):39–43, 1953.

[118] Phillip Bonacich. Power and centrality: A family of measures. *American journal of sociology*, pages 1170–1182, 1987.

[119] Carl D Meyer. *Matrix analysis and applied linear algebra*. Siam, 2000.

[120] Vivek S Borkar, Aditya Karnik, Jisha M Nair, and Sanketh Nalli. Manufacturing consent. *Automatic Control, IEEE Transactions on*, 60(1):104–117, 2015.

[121] Stephen Boyd and Lieven Vandenberghe. *Convex optimization*. Cambridge university press, 2004.

[122] Wayne W Zachary. An information flow model for conflict and fission in small groups. *Journal of anthropological research*, pages 452–473, 1977.

[123] Donald Ervin Knuth, Donald Ervin Knuth, and Donald Ervin Knuth. *The Stanford GraphBase: a platform for combinatorial computing*, volume 37. Addison-Wesley Reading, 1993.

[124] Michelle Girvan and Mark EJ Newman. Community structure in social and biological networks. *Proceedings of the national academy of sciences*, 99(12):7821–7826, 2002.

[125] David Lusseau, Karsten Schneider, Oliver J Boisseau, Patti Haase, Elisabeth Slooten, and Steve M Dawson. The bottlenose dolphin community of doubtful sound features a large proportion of long-lasting associations. *Behavioral Ecology and Sociobiology*, 54(4):396–405, 2003.

[126] Sébastien Bubeck and Nicolo Cesa-Bianchi. Regret analysis of stochastic and nonstochastic multi-armed bandit problems. *arXiv preprint arXiv:1204.5721*, 2012.

[127] Suresh Sethi and Gerald Thompson. *Optimal Control Theory–Applications to Management Science and Economics*. Kluwer, 2000.

[128] Tongqing Qiu, Zihui Ge, Seungjoon Lee, Jia Wang, Qi Zhao, and Jun Xu. Modeling channel popularity dynamics in a large iptv system. In *ACM SIGMETRICS Performance Evaluation Review*, volume 37, pages 275–286. ACM, 2009.

[129] Morton I Kamien, Yair Tauman, and Shmuel Zamir. On the value of information in a strategic conflict. *Games and Economic Behavior*, 2(2):129–153, 1990.

[130] Ali Khanafer and Tamer Basar. An optimal control problem over infected networks. In *Proceedings of the International Conference of Control, Dynamic Systems, and Robotics, Ottawa, Ontario, Canada*, 2014.

[131] John Duchi, Elad Hazan, and Yoram Singer. Adaptive subgradient methods for online learning and stochastic optimization. *The Journal of Machine Learning Research*, 12:2121–2159, 2011.

Personal Bibliography

Reiffers-Masson, A., Hayel, Y. Altman, E. (2013). Collusions entre fournisseurs de services et de contenus dans les réseaux. In AlgoTel-15èmes Rencontres Francophones sur les Aspects Algorithmiques des Télécommunications (pp. 1-4).

Reiffers-Masson, A., Hayel, Y., Altman, E. (2014, January). Game theory approach for modeling competition over visibility on social networks. In Communication Systems and Networks (COMSNETS), 2014 Sixth International Conference on (pp. 1-6). IEEE.

De Pellegrini, F., Reiffers, A., Altman, E. (2014, June). Differential games of competition in online content diffusion. In Networking Conference, 2014 IFIP (pp. 1-9). IEEE.

Reiffers-Masson, A., Hayel, Y., Altman, E. (2014, January). 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).

Maggi, L., De Pellegrini, F., Reiffers-Masson, A., Herings, J. J., Altman, E. (2014, October). Coordination of epidemic control policies: A game theoretic perspective. In International Conference on NETwork Games COntrol and oPtimization 2014 (NetGCoop'14).

Reiffers-Masson, A., Marrel, G., (2015, January). Les posts du maire de Paris en temps ordinaires. Des stratégies de visibilisation du travail politique sur Facebook ? In 6ème Congrès des associations francophones de Science Politique.

Reiffers-Masson, A., Altman, E., Hayel, Y. (2015, August). Posting behavior in Social Networks and Content Active Filtering. In 1st International Workshop on Dynamics in Networks (DyNo2015), in conjunction with IEEE/ACM ASONAM 2015.

Reiffers-Masson, A., Altman, E., Hayel, Y. (2015, December). Controlling the Katz-Bonacich Centrality in Social Network: Application to gossip in Online Social Networks. In 3rd International workshop on Big Data and Social Networking Management and Security (BDSN-2015), in conjunction with the 8th IEEE/ACM Conference On Utility and Cloud Computing (UCC-2015),

Portilla, Y., Reiffers-Masson, A., Altman, E., EL-Azouzi R. (2015, December). A Study of YouTube recommendation graph based on measurements and stochastic tools. In 3rd International workshop on Big Data and Social Networking Management and Security (BDSN-2015), in conjunction with the 8th IEEE/ACM Conference On Utility and Cloud Computing (UCC-2015),

Reiffers-Masson, A., Altman, E., Hayel, Y. (2015). Posting behaviour Dynamics and Active Filtering for Content Diversity in Social Networks. Submitted for publication.

Reiffers-Masson, A., Hayel, Y., Altman, E., and Marrel, G. (2015). A Continuous Optimization Approach for Maximizing Content Popularity in Online Social Networks. Submitted for publication.

Reiffers-Masson, A., Hayel, Y., Altman, E. (2015). Visibility Competition on Online Social Networks. Submitted for publication.